A general method of generating and classifying Clifford algebras

Deming Li,[a)] Charles P. Poole, Jr., and Horacio A. Farach
Department of Physics, University of South Carolina, Columbia, South Carolina 29208

(Received 5 March 1985; accepted for publication 6 November 1985)

A general construction program is given for the generation of higher-order Clifford algebras of various signatures together with their faithful representations in terms of Pauli-type operators. The analysis is based upon several isomorphism theorems that provide a simpler understanding of the standard classification scheme of these algebras, and an improved signature index notation is suggested that identifies each of the five types of Clifford algebras

I. INTRODUCTION

The algebraic settings of physics have been proven to be an indication of its level of development throughout its history from vector algebra to tensor analysis, to Lie algebras, and more recently to Clifford algebras and the related exterior forms. We are able to present the ever-expanding scope of physics in more and more compact formalisms. However, it takes time for physicists to adapt themselves to each new mathematical setting. In the case of Clifford algebra, it was half a century after its discovery[1] that Pauli and Dirac initiated its first applications in quantum physics, and only after about 30 more years did physicists begin to recognize the profound significance of Clifford algebras for physics, mainly through the pioneering works of Hestenes.[2] Compared with other mathematical methods used in physics, this is a rather long course. In the case of group theory, the first works on continuous groups[3] appeared in 1888–1893. However, the physically important representation theory of Lie groups appeared as late as 1924, and four years later Weyl published his monumental work[4] on group theory applied to quantum mechanics. In the case of Clifford algebra the reason for this time lag in application is partly in the mathematics itself; it had not been put in a form convenient for use. On the one hand, we need a suitable formulation directly related to physical space-time, and on the other hand we need a representation in block matrix form as in the Dirac equation. In recent years, a great deal of important work has been done for the foundations,[5] for the representations,[6–9] and for such applications as relativistic quantum theory.[10,11] Clifford algebras of higher order are now being extensively used, especially in particle physics.[12]

For mathematicians the classifications are of central importance, but the representations of Clifford algebras seem of little interest to them. This is because, unlike a group, a Clifford algebra has essentially only one faithful inequivalent irreducible representation. For physicists a representation is not only indispensable for numerical calculations, it also helps with the understanding. This is why there are so many different but equivalent representations associated with the Dirac algebra. The construction of a representation itself can provide us with a better overall understanding of the classifications, as will be shown in this paper.

It is no wonder that the first representations of even-order Clifford algebras were given by the physicists Jordan and Wigner.[13] Later Brauer and Weyl[14] (cf. also Ref. 15) obtained a complete result for any order, but it was "unnecessarily" sophisticated and has so far rarely been used. On the other hand, Atiyah *et al.*[16] have given a complete classification for universal Clifford algebras without providing any realization to be used for algebraic manipulations.

It was Salingaros who, in a series of important papers,[17–21] provided an explicit realization of the universal Clifford algebras in terms of differential forms, duality, and an associative product, which lead to a natural classification of the universal Clifford algebras based on their associated group structure. Salingaros' articles show clearly how profitable a study of the representations of Clifford algebras can be.

In this paper we concentrate on the generation and the classification of Clifford algebras, starting directly from the Clifford basis elements. In analyzing the possible combinations of their direct products we found two universal procedures for the construction of the higher-order basis elements. Following these procedures, starting from Pauli-type operators, we are able to construct all universal Clifford algebras in terms of their basis elements. This basis element analysis also made it simpler to establish a complete set of equivalence theorems among different-signatured Clifford algebras. Combined with the two construction procedures the complete classification of universal Clifford algebras is given in a rather simpler way.

The whole work is divided into three main parts. In the first part, the general theory and equivalence relations are presented, together with an elucidation of the properties and importance of the canonical element. In the second part generating methods are formulated and a particular one is developed that generates the hierarchies of even and odd order algebras. It is as easy to apply as a rule of thumb, yet it provides us with a clear understanding of the classification scheme of Clifford algebras without involving any abstract propositions. The third part discusses the various types of Clifford algebras and their interrelationships. The essential differences between even- and odd-order algebras are clarified with the aid of the canonical element.

II. GENERAL PROPERTIES OF EVEN- AND ODD-ORDER CLIFFORD ALGEBRAS

In this section we will present proofs of the two fundamental theorems that provfide the overall classification

[a)] Permanent address: Department of Physics, Shanxi University, Taiyuan, People's Republic of China.

 0022-2488/86/051173-08$02.50

scheme of the Clifford algebras. We begin by defining the notation.

A universal Clifford algebra is an n-dimensional real linear space with n anticommuting basis elements

$$A_n = \{\sigma_1, \sigma_2, \ldots, \sigma_n\} \tag{1}$$

endowed with the multiplication rule

$$\sigma_i \sigma_j + \sigma_j \sigma_i = 2g^{ij} e , \tag{2}$$

where e is the unit element of the algebra and $g^{ij} = \pm \delta_{ij}$ is the metric tensor. Metrics other than the Euclidean type are important in physical applications, and we will give a systematic discussion of them. For a particular Clifford algebra, p of the basis elements have a positive square $\sigma_j^2 = +1$ and q of them have a negative square $(\sigma_j')^2 = -1$. The notation σ_j' will be employed for a basis element of a square of -1. For each order n there are $n+1$ Clifford algebras, which we denote by A_n^s, where

$$A_n^s = \{\sigma_1, \sigma_2, \ldots, \sigma_p; \sigma'_{p+1}, \ldots, \sigma'_{p+q=n}\} , \tag{3}$$

$$s = p - q , \tag{4}$$

and, of course,

$$-n \leqslant s \leqslant n, \quad n = p + q . \tag{5}$$

The pair of integers p, q is called the signature and the difference $s = p - q$ is referred to as the signature index. We should note that s is even when n is even and it is odd when n is odd.

Two Clifford algebras of the same order but with different signature indices can be algebraically equivalent to each other. The following two theorems and their corollaries exhaust all of the general equivalence relations.

Theorem I: Two Clifford algebras are equivalent when their indices are the same modulo 8

$$A_n^s \cong A_n^{s \bmod 8} , \tag{6}$$

where the symbol $\cong$ denotes algebraic equivalence.

Proof: Assuming $q > 4$, starting from (3), first let us define

$$\Gamma \equiv \sigma'_{p+1} \sigma'_{p+2} \sigma'_{p+3} \sigma'_{p+4} , \tag{7}$$

where

$$\Gamma^2 = \sigma'_{p+1} \sigma'_{p+2} \sigma'_{p+3} \sigma'_{p+4} \sigma'_{p+1} \sigma'_{p+2} \sigma'_{p+3} \sigma'_{p+4} .$$

We can easily show that the element $\Gamma \sigma'_{p+j}$,

$$\Gamma \sigma'_{p+j} = \sigma'_{p+1} \sigma'_{p+2} \sigma'_{p+3} \sigma'_{p+4} \sigma'_{p+j} , \tag{8}$$

where $j = 1, 2, 3, 4$ has a positive square

$$(\Gamma \sigma'_{p+j})^2 = \Gamma \sigma'_{p+j} \Gamma \sigma'_{p+j} = 1 .$$

It is easy to verify that the set of elements

$$\{\sigma_1, \sigma_2, \ldots, \sigma_p, \Gamma\sigma'_{p+1}, \Gamma\sigma'_{p+2}, \Gamma\sigma'_{p+3}, \Gamma\sigma'_{p+4}, \sigma'_{p+5}, \ldots, \sigma'_n\} , \tag{9}$$

in which the four elements σ'_{p+j}, with $j = 1, 2, 3, 4$ of the basis (3), are replaced by the four elements and $\Gamma \sigma'_{p+j}$ constitutes the basis set of the algebra A_n^{s+8}, which has the same overall set of elements as the algebra A_n^s. This means that the two algebras are equivalent to each other. Conversely, we can get back from (9) to (3) by

$$\Gamma' \equiv \Gamma\sigma'_{p+1} \Gamma\sigma'_{p+2} \Gamma\sigma'_{p+3} \Gamma\sigma'_{p+4} .$$

By carrying out the product operations we find that

$$\Gamma' = \Gamma . \tag{10}$$

This process can be repeated as long as $q > 4$ negative square elements remain.

Theorem II: Two Clifford algebras are equivalent when their signature indices are reflected about the value $s = +1$, which means that

$$A_n^s \cong A_n^{2-s} . \tag{11}$$

Proof: Starting from the basis elements of A_n^s, with the signature p,q,

$$\{\sigma_1, \sigma_2, \ldots, \sigma_p; \sigma'_{p+1}, \ldots, \sigma'_n\} , \tag{12}$$

we can form a new set of basis elements for A_n^{2-s} with the signature $q+1, p-1$,

$$\{\sigma_1, \sigma_1\sigma'_{p+1}, \sigma_1\sigma'_{p+2}, \ldots, \sigma_1\sigma'_n; \sigma_1\sigma_2, \sigma_1\sigma_3, \ldots, \sigma_1\sigma_p\} . \tag{13}$$

Let us check the various products:

$$\sigma_1\sigma_s \cdot \sigma_1\sigma_t = -\sigma_s\sigma_t = \sigma_t\sigma_s = \sigma_1^2\sigma_t\sigma_s = -\sigma_1\sigma_t\sigma_1\sigma_s ,$$

$$\sigma_1(\sigma_1\sigma_r) = -(\sigma_1\sigma_r)\sigma_1 ,$$

$$(\sigma_1\sigma_i')^2 = \sigma_1\sigma_i'\sigma_1\sigma_i' = -\sigma_1^2\sigma_i'^2 = e ,$$

$$(\sigma_1\sigma_j)^2 = \sigma_1\sigma_j\sigma_1\sigma_j = -\sigma_1^2\sigma_j^2 = -e .$$

So they really do form the algebra A_n^{2-s}. Conversely, from (13) we can get back to (12) by the same procedure. Hence (12) and (13) are equivalent.

Corollary I:

For $s = p - q = 0 \pmod 4$, $A_n^s \cong A_n^{-s}$. (14)

Proof: Let $p = q + 4l$, then by Theorem I,

$$A_n^s \cong A_n^{s-8l} = A_n^{p-4l-(q+4l)} = A_n^{q-p} = A_n^{-s} .$$

Corollary II:

For $s = p - q = 0 \pmod 4$, $A_n^s \cong A_n^{s+2}$. (15)

Proof: Using Theorem II and Corollary I, we have

$$A_n^s \cong A_n^{-s} \cong A_n^{2-(-s)} = A_n^{s+2} .$$

Corollary III:

For $s = -1 \pmod 4$, $A_n^s \cong A_n^{s+4l}$. (16)

Proof: $s = -1 \pmod 4$ means $p = q - 1 + 4(t-1)$, t being some arbitrary integer. Then by Theorem II and Theorem I we get

$$A_n^s = A_n^{-1+4(t-1)} \cong A_n^{2-[-1+4(t-1)]} = A_n^{3-4(t-1)}$$

$$\cong A_n^{3+4(t-1)} = A_n^{-1+4t} = A_n^{s+4} . \tag{17}$$

Repeating this process l times and we have finally

$$A_n^s \cong A_n^{s+4l} . \tag{18}$$

These two theorems and their three corollaries exhaust all the equivalence relations among Clifford algebras with different indices. The Clifford algebras with $s = -3$ and 1 have no additional equivalence relations besides the general period of 8 and the reflection about $s = 1$.

III. CANONICAL ELEMENT

The properties of Clifford algebras are critically related to the properties of the "canonical element" or "volume element in n dimensions,"[17,22] which is defined as the product of all the basis elements:

$$\sigma(n) \equiv \sigma_1\sigma_2\cdots\sigma_n . \tag{19}$$

It is easily demonstrated that

$$\sigma_k\sigma(n) = \pm\sigma(n)\sigma_k , \tag{20}$$

where the positive sign appears when the order n is odd and the negative sign appears when n is even. Thus the canonical element anticommutes with all of the basis elements when n is even and commutes with them when n is odd. The square of $\sigma(n)$ is

$$[\sigma(n)]^2 = \sigma_1\sigma_2\cdots\sigma_n\sigma_1\cdots\sigma_n = (-1)^{n(n-1)/2}\sigma_1^2\sigma_2^2\cdots\sigma_n^2 = (-1)^{n(n-1)/2+q} , \tag{21}$$

and recalling Eq. (6) we have

$$[\sigma(n)]^2 = (-1)^{(1/2)(n^2-s)} = \begin{cases} (-1)^{(1/2)s}, & n = \text{even} , \\ (-1)^{(1/2)(s-1)}, & n = \text{odd} . \end{cases} \tag{22}$$

Hence there are four possibilities:

$$\text{(i)} \quad s = 0 \pmod 4 \quad (n \text{ even}) \quad [\sigma(n)]^2 = e , \tag{23}$$

$$\text{(ii)} \quad s = 2 \pmod 4 \quad (n \text{ even}) \quad [\sigma(n)]^2 = -e , \tag{24}$$

$$\text{(iii)} \quad s = -1 \pmod 4 \quad (n \text{ odd}) \quad [\sigma(n)]^2 = -e , \tag{25}$$

$$\text{(iv)} \quad s = 1 \pmod 4 \quad (n \text{ odd}) \quad [\sigma(n)]^2 = e . \tag{26}$$

Equation (22) has profound consequences on the properties of even- and odd-order Clifford algebras. Because of the minus sign in (22), in an even-order Clifford algebra only the identity element commutes with all other elements and there is only one irreducible representation (cf. Appendix A). For n odd, $\sigma(n)$ commutes with the basis elements, so we have a larger center consisting of the two elements e and $\sigma(n)$ that commute with all elements of the algebra. This suggests, in the $p - q = s = 1 \pmod 4$ case, the possibility of introducing the projectors $P_n^{\pm}$ used by Basri and Barut,[12,20]

$$P_n^{\pm} \equiv \tfrac{1}{2}[e \pm \sigma(n)] , \tag{27}$$

which, because of (26), are idempotent, and satisfy

$$(P_n^{\pm})^2 = P_n^{\pm} , \quad P_n^+P_n^- = P_n^-P_n^+ = 0 . \tag{28}$$

Let A_n^E denote the subalgebra of elements that are formed from the set of all even products of nonidentical basis elements:

$$A_n^E : \{e, \sigma_i\sigma_j, \sigma_i\sigma_j\sigma_k\sigma_l, \ldots\} . \tag{29}$$

Then using (28) we get

$$A_n = A_{n-1}^+ + A_{n-1}^- , \tag{30}$$

with

$$A_{n-1}^{\pm} \equiv P_n^{\pm}A_n^E . \tag{31}$$

Equation (30) means that when $s = 1 \pmod 4$, A_n is composed of two even-order disjoint subalgebras and the faithful representation M_F of A_n is in block diagonal form

$$M_F = \begin{pmatrix} M^+ & 0 \\ 0 & M^- \end{pmatrix} ,$$

where the $M^{\pm}$ are the representation matrices of $A_{n-1}^{\pm}$. A similar decomposition was given by Brauer and Weyl.[14]

IV. THE GENERAL GENERATING METHOD

It is well known that the direct product gives rise to higher-order Clifford algebras:

$$A_m \times A_n \cong A_{m+n} , \tag{32}$$

where either m or n must be even. In this section we will examine a systematic way to generate such higher-order algebras. To accomplish this we must analyze the implications of this expression (32); more explicitly we need the detailed relationships between the basis elements of A_{m+n} and those of A_m and A_n.

We will proceed to show how the direct product of an even-order algebra A_m^s of signature (p,q), called the generating algebra with an algebra A_n^t of signature (u,v) called the starting algebra (which is either of even or odd order) may be employed to generate a final algebra A_{m+n}^w of order $m+n$ and signature (P,Q) (which is even or odd, respectively) through the use of Eq. (32). To accomplish this we will take direct products of the basis elements:

$$A_m^s : \{\sigma_1, \ldots, \sigma_p; \sigma'_{p+1}, \ldots, \sigma'_m\} , \tag{33}$$

$$A_n^t : \{\tau_1, \ldots, \tau_u; \tau'_{u+1}, \ldots, \tau'_n\} , \tag{34}$$

where $s = 2p - m$ and $t = 2u - n$.

Note that for the even-order generating algebra the canonical element $\sigma(m)$ anticommutes with the basis and it could be treated on the same footing as other generating elements, so let us agree to use Σ_k, $k = 1,2,\ldots,m+1$, to represent all of the σ_i and σ'_i of Eq. (33) plus $\sigma(m)$.

The problem is to choose a set of suitable compound basis elements $\Sigma_k \times \tau_i$, $\Sigma_l \times \tau_j$ such that the anticommunication relation

$$(\Sigma_k \times \tau_i)(\Sigma_l \times \tau_j) + (\Sigma_l \times \tau_j)(\Sigma_k \times \tau_i) = 2g^{(ki),(lj)}e \tag{35}$$

is satisfied. Here the compound metric is dependent on that of the component metrics and will be given later.

Basically there are two cases.

(1) Part of the compound basis set is formed from m of the $m+1$ possible choices of Σ_k,

$$\Sigma_k \times e_\tau , \quad k = 1,2,\ldots,m, k \neq k_0 ,$$

where e_τ is the unit element of the starting algebra A_n^t, so the other part must be of the form

$$\Sigma_{k_0} \times \tau_j , \quad j = 1,2,\ldots,n ,$$

where Σ_{k_0} is the remaining Σ_k element that was omitted above. Thus we obtain the complete basis

$$A_m \times A_n: \quad \{\Sigma_{k_0}\times\tau_j;\Sigma_k\times e_\tau\}\to A_{m+n}: \quad P[\Sigma_{k_0}x], \tag{36}$$

where k_0 is fixed for a particular basis set and $k\neq k_0$ runs through all of other indices of Σ_k. The symbol $P[\Sigma_{k_0}x]$ is used to designate this generating program.

(2) Part of the compound basis is formed from $n-1$ of the n possible choices of τ_j as follows:

$$e_\sigma\times\tau_j, \quad j=1,\ldots,n; \quad j\neq j_0.$$

Here e_σ is the unit element of the generating algebra A^s_m, so the other part must be of the form

$$\Sigma_k\times\tau_{j_0}, \quad k=1,\ldots,m+1,$$

with fixed j_0 and $j\neq j_0$ runs through all other indices of τ:

$$A_m\times A_n: \{\Sigma_k\times\tau_{j_0}, e_\sigma\times\tau_j\}\to A_{m+n}: \; P[\times\tau_{j_0}]. \tag{37}$$

The notations $P[\Sigma_{k_0}\times]$ and $P[\times\tau_{j_0}]$ will be explained below. Each of these two classes of basis elements can be easily shown to make up a complete basis for the $(m+n)$-order compound final algebra A^w_{m+n}.

In order to carry out this direct product method whereby an even-order m generating algebra A^s_m operates on a starting algebra A^t_n to produce the higher-order final algebra A^w_{m+n},

$$A^s_m\times A^t_n = A^w_{m+n}, \tag{38}$$

there are three choices to make: (1) the generating algebra's signature index s can be selected as either 0 mod 4 or as 2 mod 4, (2) either the program $\Sigma_k\times e_\tau$ or $e_\sigma\times\tau_j$ may be chosen, and (3) the characteristic element may have a positive square (Σ_{k_0} or τ_{j_0}) or a negative square (Σ'_{k_0} or τ'_{k_0}). These three binary choices provide us with the eight different generating procedures that are listed in Table I. Each program is designated by the symbol P^s (proc) in which the argument "proc" gives the type ($\Sigma_k\times e_\tau$ or $e_\tau\times\tau_j$) and the sign of the square (σ_{k_0} or τ_{j_0} for positive and σ'_{k_0} or τ'_{j_0} for negative) of the characteristic element. The table gives the signature index w of the generated algebra in terms of the indices s and t of the two lower-order algebras. It also gives the same result in terms of the signatures themselves, which are contained in the following equivalent expression for the algebra generating procedure:

$$C_m(p,q)\times C_n(u,v) = C_{m+n}(P,Q), \tag{39}$$

where (P,Q) is the signature of the final algebra and, of course, $w=P-Q$. Coquereaux[22] gives a few constructions similar to ours.

V. ALGEBRA GENERATION PROCEDURES

Now it is a simple matter to use the algebras A^s_2 as the generating algebras for constructing the complete set of Clifford algebras. The $P^s[\times T_{j_0}]$ procedures are more convenient for this purpose, and because $A^2_2\cong A^0_2$ by Theorem II we only use A^0_2 and A^{-2}_2 as the generating algebras. This causes Table I to reduce to Table II.

The operations R, M, and L defined in Table II may be looked upon as generators of higher-order algebras with the properties

$$RA^t_n = A^{t-4}_{n+2}, \tag{40}$$

$$MA^t_n = A^t_{n+2}, \tag{41}$$

$$LA^t_n = A^{t+4}_{n+2}, \tag{42}$$

where Eqs. (40)–(42) are special cases of Eq. (38) in a new notation. The repeated use of these operations R, M, and L generate the hierarchies of algebras presented in Figs. 1 and 2. The R operation generates an algebra below and two to the right, M generates one immediately below, and L generates one below and two to the left, as mentioned in the captions. The algebras A^t_2 with $t=0$, ± 2 serve as the starting algebras for generating the even hierarchy given in Fig. 1 and A^t_1 with $t=\pm 1$ are the starting algebras for the odd hierarchy presented on Fig. 2. Thus, A^0_2 can be generated from A^0_0 by this method, but the other two algebras A^2_2 and A^{-2}_2 cannot.

Brauer and Weyl[4] also expressed higher-order algebras in terms of direct products involving A^s_2 matrices but they did not present a systematic hierarchy of generated algebras of the type shown in Figs. 1 and 2.

VI. INEQUIVALENCE THEOREMS

We showed that the equivalence theorems limit the number of different types of Clifford algebras to no more than 5, and mathematicians have already worked out the classification of the five types of algebras according to the representation spaces[5] or to the direct product construction.[23] We have not found any inequivalence theorems that

TABLE I. Characteristics of the eight procedures for generating higher-order Clifford algebras by the formation of direct products.

$s=p-q$	Program	Final signature (P,Q)	Final signature index w	Equivalence
	$P^0[\Sigma_{k_0}\times]$,	$(p+u,q+v)$	$s+t$	···
0(mod 4)	$P^0[\Sigma'_{k_0}\times]$,	$(p+v+1,q+u-1)$	$s-t+2$	$-s+t$
	$P^0[\times\tau_{j_0}]$,	$(p+u,q+v)$	$s+t$	···
	$P^0[\times\tau'_{j_0}]$,	$(q+u,p+v)$	$-s+t$	$s-t+2$
	$P^2[\Sigma_{k_0}\times]$,	$(p+u-1,q+v+1)$	$s+t-2$	···
2(mod 4)	$P^2[\Sigma'_{k_0}\times]$,	$(p+v,q+u)$	$s-t$	$-s+t+2$
	$P^2[\times\tau_{j_0}]$,	$(p+u-1,q+v+1)$	$s+t-2$	···
	$P^2[\times\tau'_{j_0}]$,	$(q+u+1,p+v-1)$	$-s+t+2$	$s-t$

TABLE II. Procedures for forming higher-order algebras using the second-order generating algebra A_2^s. The signatures (P,Q) and of the final generated algebra A_{m+n}^w are given in terms of those (u,v) of the starting algebra A_n^t.

Generating algebra	Program	Final signature (P,Q)	Final signature index w	Symbol
A_2^0	$P^0[\times\tau'_{j_0}]$	$(u+1,v+1)$	t	M
A_2^0	$P^0[\times\tau'_{j_0}]$	$(u+1,v+1)$	t	M
A_2^{-2}	$P^2[\times\tau_{j_0}]$	$(u-1,v+3)$	$t-4$	R
A_2^{-2}	$P^2[\times\tau'_{j_0}]$	$(u+3,v-1)$	$t+4$	L

prove that these five algebra types are definitely inequivalent with each other. For a long time some physicists have assumed that $C_4(3,1)\equiv A_4^2$ is equivalent to $C_4(1,3)\equiv A_4^{-2}$, and it was Salingaros who first made it clear that they are not by the use of a group-structure approach.[19] This might have important implications in various branches of theoretical physics. In this section we will give a more straightforward discussion of this problem based on direct product procedures. The inequivalence theorems that we prove also follow from a group theoretic construction.[24–25]

Algebras constructed by the programs given above differ not only by the generating and starting algebras but also by the procedures used. For example, there is no general rule as to whether or not

$$A_m^s\times A_n^t \tilde{\neq} A_m^s\times A_n^{t'},$$

if $A_n^t\tilde{\neq} A_n^{t'}$. Nevertheless we do have the following theorem.

Theorem III: Given $A_n^t\tilde{\neq} A_n^{t-2}$, then

$$A_{n+2}^t\tilde{\neq} A_{n+2}^{t-2}. \quad (43)$$

Proof: Consider the two algebras A_n^t and A_n^{t-2} with the following sets of basis elements:

$$A_n^t:\{\tau_1,\tau_2,\ldots,\tau_u;\ \tau'_{u+1},\ldots,\tau'_{u+v}\} \quad (44)$$

and

$$A_n^{t-2}:\{\tau_1,\ldots,\tau_{u-1};\ \tau'_u,\tau'_{u+1},\ldots,\tau'_{u+v}\}. \quad (45)$$

These algebras differ only by the replacement of τ_u by τ'_u as a basis element. Using A_2^0 through program $P[\times\tau_{j_0}]$ (or $P[\times\tau'_{j_0}]$ (M step)) we get the two compound algebras

$$A_2^0\times A_n^t=A_{n+2}^t:\{\sigma_1\times\tau_{j_0},\sigma_1\sigma'_2\times\tau_{j_0};\sigma'_2\times\tau_{j_0};e_\sigma\times\tau_1,\ldots,$$
$$e_\sigma\times\tau_{j_0-1},e_\sigma\times\tau_{j_0+1},\ldots,e_\sigma\times\tau_u;$$
$$e_\sigma\times\tau'_{u+1},\ldots,e_\sigma\times\tau'_{u+v}\},$$

$$A_2^0\times A_n^{t-2}=A_{n+2}^{t-2}:\{\sigma_1\times\tau_{j_0},\sigma_1\sigma'_2\times\tau_{j_0};\sigma'_2\times\tau_{j_0};$$
$$e_\sigma\times\tau_1,\ldots,e_\sigma\times\tau_{j_0-1},\ldots,e_\sigma\times\tau_{u-1};$$
$$e_\sigma\times\tau'_u,\ldots,e_\sigma\times\tau'_{u+v}\}.$$

All but two of the compound basis elements from these two sets are the same, since only $e_\sigma\times\tau_u$ in A_{n+2}^t and $e_\sigma\times\tau'_u$ in A_{n+2}^{t-2} differ from each other. But according to Appendix B, these two algebras can be equivalent only when $t=2$ (mod 4), otherwise $A_{n+2}^t\tilde{\neq} A_{n+2}^{t-2}$. On the other hand, when $t=2$ (mod 4), according to Corollary II we will have $A_n^t\cong A_n^{t-2}$, which is against our assumption, so this case is eliminated. Hence the theorem.

Theorem IV:

$$A_1^1\tilde{\neq} A_1^{-1}, \quad (46)$$

$$A_2^{t\neq-2}\tilde{\neq} A_2^{-2}, \quad (47)$$

$$A_3^{t\neq-3}\tilde{\neq} A_3^{-3}. \quad (48)$$

Proof: Suppose that the opposite is true. Then it would be possible to show, for example, that $(\sigma_1,\sigma_2,\sigma'_3)\cong(\sigma'_1,\sigma'_2,\sigma'_3)$, which means there would be at least one element of A_3^{-3} that behaves like σ_1, anticommutes with some other elements, and has a positive square.

We will show that this is impossible in the case of (48). The most general form of an element in A_3^{-3} is

$$w=ae+\sum_i b_i\sigma'_i+\sum_{j,k}c_{j,k}\sigma'_j\sigma'_k+d\sigma(3),$$

where $i,j,k=1,2,3,j\neq k$, and the coefficients a, b_i, c_i, and d are real. The commuting part $ae+d\sigma(3)$ should be dropped if w is to have the same algebraic properties as a basis element. In the A_3^{-3} case, w can also be put in the form

A_0^0

A_2^2 A_2^0 A_2^{-2}

A_4^4 A_4^2 A_4^0 A_4^{-2} A_4^{-4}

A_6^6 A_6^4 A_6^2 A_6^0 A_6^{-2} A_6^{-4} A_6^{-6}

FIG. 1. Hierarchy of even-order algebras generated by the procedures of Table II. For example, L generates A_6^6, M generates A_6^2, and R generates A_6^{-2} from A_4^2.

A_1^1 A_1^{-1}

A_3^3 A_3^1 A_3^{-1} A_3^{-3}

A_5^5 A_5^3 A_5^1 A_5^{-1} A_5^{-3} A_5^{-5}

A_7^7 A_7^5 A_7^3 A_7^1 A_7^{-1} A_7^{-3} A_7^{-5} A_7^{-7}

FIG. 2. Hierarchy of odd-order algebras generated by the procedures of Table II. For example, L generates A_7^1, M generates A_7^{-3}, and R generates A_7^{-7}, from A_5^{-3}.

$$w=\sum_i b_i\sigma_i' + \sum_j c_j\sigma(3)\sigma_j', \tag{49}$$

since, for example, $\sigma(3)\ \sigma_3' = \sigma_1'\ \sigma_2'\ \sigma_3'\ \sigma_3' = -\ \sigma_1'\ \sigma_2'$. Then

$$w^2=\Big(\sum_i b_i\sigma_i'\Big)^2+\Big(\sum_j c_j\sigma(3)\sigma_j'\Big)^2,$$
$$+\sum_{i,j} b_ic_j(\sigma_i'\sigma'(3)\sigma_j'+\sigma(3)\sigma_j'\sigma_i')$$
$$=-\sum_i(b_i^2+c_i^2)-2\sigma(3)\sum_i b_ic_i, \tag{50}$$

where $\sigma(3)\equiv\sigma_1', \sigma_2', \sigma_3'$, is algebraically independent of e even though it commutes with the whole algebra A_3^{-3}. There is no way to choose the real coefficients b_i and c_i to make $w^2=e$, with a positive square, and hence $A_3^{-3}\not\cong A_3^{s\neq -3}$. Similar but much easier proofs can be applied to the other two cases (46) and (47).

VII. TYPES OF CLIFFORD ALGEBRAS

We are now in a position to write down each distinct algebra type for all of the signature indices in the principal range from -3 to $+4$:

$$\begin{array}{lcccccccccc} \text{signature index } s = & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ \text{algebra type} \quad = & E & D & C & B & A & B & C & D \end{array} \tag{51}$$

Theorem II was used to identify the algebra types that are the same due to the symmetry about the value $s=+1$, and Theorem I ensures that this pattern repeats modulo 8 for signature indices s outside the principal range. Theorem II ensures that $A^0\cong A^2$, Corollary II entails $A^4\cong A^{-2}$, and from Corollary III, $A^3\cong A^{-1}$. Then Theorems IV and III ensure that A, B, C, D, and E are all inequivalent with each other, so there are really only five different types of Clifford algebras. They differ by their "starting algebras": $A_3^{-3}, A_2^{-2}, A_1^{-1}, A_0^0, A_1^1$. (We choose A_0^0 to characterize the A_n^0 type rather than A_2^0.) Three of these five algebras have the following carrier fields:

$$\begin{aligned} &A_0^0: \text{ real number field: } R,\\ &A_1^{-1}: \text{ complex number field: } C,\\ &A_2^{-2}: \text{ Hamilton's well-known quaternion field: } H. \end{aligned}\tag{52}$$

What about A_3^{-3} and A_1^1? For A_1^1, the two elements $\{e,\sigma\}$ can be written as the linear combination $\{(e+\sigma)/2, (e-\sigma)/2\}\equiv\{\alpha,\beta\}$ and the whole algebra is composed of two disjoint parts:

$$\alpha^2=\alpha,\quad \beta^2=\beta,\quad \alpha\beta=\beta\alpha=0.$$

So we have $A_1^1=R+R$, or following Porteus,[5] $A_1^1={}^2R$. From Fig. 2 and Table II,

$$A_3^{-3}=A_2^{-2}\times A_1^1$$
$$=H\times{}^2R=H\times(R+R)=H+H={}^2H, \tag{53}$$

which is also a direct sum. Both A_1^1 and A_3^{-3} are called double field algebras.

Thus we have five different types of Clifford algebras:

$$\begin{array}{lccccc} s= & -3, & -2, & -1, & 0, & 1,\\ \text{algebra}= & {}^2H, & H, & C, & R, & {}^2R, \end{array}\tag{54}$$

The order of the algebra also turns out to be related to the dimension of the representation space. Because each M step has a 2×2 representation matrix it can easily be shown that[5]

$$\begin{aligned} &A_n^1\cong{}^2R(2^{(n-1)/2}),\quad A_n^0\cong R(2^{n/2}),\\ &A_n^{-1}\cong C(2^{(n-1)/2}),\quad A_n^{-2}\cong H(2^{n/2-1}),\\ &A_n^{-3}\cong{}^2H(2^{(n-3)/2}), \end{aligned}\tag{55}$$

where, for example, $H(2^{n/2-1})$ means that this is a linear space of dimension $2^{n/2-1}$ with the carrier field H. Table III summarizes the notation employed for several schemes of classification.

VIII. RELATIONS BETWEEN GENERATING AND GENERATED ALGEBRAS

In the previous section we gave the signature indices for the five types of Clifford algebras. The generator M increases the order n of the algebra by the amount 2 without changing the signature index s so the algebra type remains the same. On the other hand the generators L and R decrease or increase s, respectively, by 4, and hence they can change the type. In the case of even-order algebras these two generators form R-type algebras from H types and H types from R algebras as follows:

$$\begin{aligned} &\mathrm{L}R_n^s=H_{n+2}^{s-4},\quad \mathrm{L}H_n^s=R_{n+2}^{s-4},\\ &\mathrm{R}R_n^s=H_{n+2}^{s+4},\quad \mathrm{R}H_n^s=R_{n+2}^{s+4}. \end{aligned}\tag{56}$$

In these expressions one should be careful not to confuse the symbol R used for the generator and R denoting the real algebra type. A similar relationship exists in the odd-order case in the sense that the generators L and R form ${}^2H$-type double field algebras from ${}^2R$ double field ones, and they also form ${}^2R$ from ${}^2H$. When these two generators operate on a C-type algebra they form another C type,

$$\mathrm{L}C_n^s=C_{n+2}^{s-4},\quad \mathrm{R}C_n^s=C_{n+2}^{s+4}. \tag{57}$$

In other words, in the odd-order case there is no mixing between the double field and the C algebras. This is important because the faithful inequivalent irreducible representation of a double field algebra has twice the dimensionality as that of the C-type algebra of the same order. On the other hand, R- and H-type algebras of the same order have representations of the same dimensionality, as is the case with double field ${}^2R$ and ${}^2H$ algebras of the same order.

Another way to see the relationship between the various algebras is to carry out the generation of an odd-order algebra using the fundamental complex number algebra A_1^{-1}, which has only one basis element $i=[-1]^{1/2}$ as the starting algebra. We generate as follows:

$$A_m^s\times A_1^{-1}=A_{m+1}^w, \tag{58}$$

where m is even using the procedure $P^m[\Sigma_{k_0}\times]$ of Eq. (36). For this case there is only one τ_j element, namely $\tau_j=i$. We select τ_{k_0} as the canonical element $\sigma(m)$ of the A_m^s algebra, and Eq. (36) gives for the new basis

$$\Sigma_k\times e_\tau=\sigma_k,\quad k=1,\ldots,m,\qquad \Sigma_{k_0}\times\tau_j=i\sigma(m), \tag{59}$$

TABLE III. Comparison of various prevailing nonmenclatures for Clifford algebras.

Order	Signature index	Present work	Porteus[a]	Salingaros[b]	Porteus[a]	Salingaros[b]	Coquereaux[c]
n	s	A_n^s	...	...	$R_{p,q}$	$A^{p,q}$	(p,q)
0	0	$A_0^0=R_0^0$	$R(1)=R$	$N_0=R$	$R_{0,0}$	$A^{0,0}$	(0,0)
1	−1	$A_1^{-1}=C_1^{-1}$	$C(1)=C$	$S_0=C$	$R_{0,1}$	$A^{0,1}$	(0,1)
1	1	$A_1^1={}^2R_1^1$	${}^2R(1)={}^2R$	$N_0\times N_0=R+R$	$R_{1,0}$	$A^{1,0}$	(1,0)
2	−2	$A_2^{-2}=H_2^{-2}$	$H(1)=H$	$N_2=H$	$R_{0,2}$	$A^{0,2}$	(0,2)
2	0	$A_2^0=R_2^0$	$R(2)$	N_1	$R_{1,1}$	$A^{1,1}$	(1,1)
2	2	$A_2^2=R_2^2$	$R(2)$	N_1	$R_{2,0}$	$A^{2,0}$	(2,0)
3	−3	$A_3^{-3}={}^2H_3^{-3}$	${}^2H(1)={}^2H$	$N_2\times N_0=H+H$	$R_{0,3}$	$A^{0,3}$	(0,3)
3	−1	$A_3^{-1}=C_3^{-1}$	$C(2)$	$S_1=S$	$R_{1,2}$	$A^{1,2}$	(1,2)
3	1	$A_3^1={}^2R_3^1$	${}^2R(2)$	$N_1\times N_0=N_1+N_1$	$R_{2,1}$	$A^{2,1}$	(2,1)
3	3	$A_3^3=C_3^3$	$C(2)$	$S_1=S$	$R_{3,0}$	$A^{3,0}$	(3,0)
4	−4	$A_4^{-4}=H_4^{-4}$	$H(2)$	N_4	$R_{0,4}$	$A^{0,4}$	(0,4)
4	−2	$A_4^{-2}=H_4^{-2}$	$H(2)$	N_4	$R_{1,3}$	$A^{1,3}$	(1,3)
4	0	$A_4^0=R_4^0$	$R(4)$	$N_3=M$	$R_{2,2}$	$A^{2,2}$	(2,2)
4	2	$A_4^2=R_4^2$	$R(4)$	$N_3=M$	$R_{3,1}$	$A^{3,1}$	(3,1)
4	4	$A_4^4=H_4^4$	$H(2)$	N_4	$R_{4,0}$	$A^{4,0}$	(4,0)

[a] Reference 5.
[b] Reference 21.
[c] Reference 22.

where the elements σ_k are the basis of the algebra A_m^s. We know from Eqs. (23) and (24) that for an even-order algebra the canonical element $\sigma(m)$ has a positive or negative square depending upon whether $s=0 \bmod 4$ or $s=2 \bmod 4$, respectively. Therefore for these two cases the generated algebras will be

$$A_m^s\times A_1^{-1}=\begin{cases}A_{m+1}^{s-1}, & s=0 \bmod 4,\\ A_{m+1}^{s+1}, & s=2 \bmod 4.\end{cases} \tag{60}$$

Thus the final algebra that is generated has the signature index $w=-1 \bmod 4$. This means that irrespective of whether or not we start with an R or an H algebra we always generate a C algebra. There is no mixing with the double field algebras.

If we start the generation process with the ${}^2R$ double field algebra A_1^1 instead of A_1^{-1} and carry out the same procedure with the Pauli matrix σ_z selected as σ_{j_0} then the final signature indices will be reversed,

$$A_m^s\times A_1^1=\begin{cases}A_{m+1}^{s+1}, & s=0 \bmod 4,\\ A_{m+1}^{s-1}, & s=2 \bmod 4,\end{cases} \tag{61}$$

and only the double field algebras ${}^2R$ and ${}^2H$ with $w=1 \bmod 4$ and $w=-3 \bmod 4$, respectively, will be generated. The resulting basis set Σ_k, $k=1,\dots,m+1$ is given by

$$\Sigma_k=\sigma_k\times I,\quad k=1,\dots,m,\quad \Sigma_{k+1}=\sigma(m)\times\sigma_z, \tag{62}$$

where I is the 2×2 unit matrix. This representation has twice the dimensionality of the corresponding C-algebra one.

IX. DISCUSSION

In this article we have analyzed the relationships between the various types of Clifford algebras, and we have proposed a systematic way of generating higher-order algebras from lower-order ones. The generation method itself and the steps that were followed in arriving at it provide some important insights into the interconnectiveness of the various Clifford algebras of different orders and signatures. In addition we believe that the approach followed in this article and the notations and classification scheme that we have adopted provide a formalism for these algebras that is particularly useful for physical applications.

APPENDIX A: PROOF THAT EACH EVEN-ORDER CLIFFORD ALGEBRA HAS ONLY ONE INEQUIVALENT IRREDUCIBLE REPRESENTATION

In this appendix we present a proof that each even-order Clifford algebra A_n has only one inequivalent irreducible representation of dimension $2^{n/2}$.

First, let Σ_w be any elements of A_n like σ_1, σ_i, σ_j,..., $\{\pm\Sigma_w\}$, which form a group of order $2\cdot 2^n\equiv N$. For any $\Sigma_w\neq\pm 1$, we can show that the trace vanishes, $\mathrm{Tr}\,\Sigma_w=0$. For this end, it is enough to show that there always exists an element $\Sigma_{\bar w}$ such that

$$\Sigma_w\Sigma_{\bar w}+\Sigma_{\bar w}\Sigma_w=0 \quad\text{or}\quad \Sigma_{\bar w}\Sigma_w\Sigma_{\bar w}=-(\Sigma_{\bar w})^2\Sigma_w, \tag{A1}$$

where $(\Sigma_{\bar w})^2=\pm 1$ depending on the nature of $\Sigma_{\bar w}$. Then we have

$$\mathrm{Tr}\,\Sigma_{\bar w}\Sigma_w\Sigma_{\bar w}=-(\Sigma_{\bar w})^2\,\mathrm{Tr}\,\Sigma_w, \tag{A2}$$

but

$$\mathrm{Tr}\,\Sigma_{\bar w}\Sigma_w\Sigma_{\bar w}=\mathrm{Tr}\,\Sigma_w(\Sigma_{\bar w})^2=+(\Sigma_{\bar w})^2\,\mathrm{Tr}\,\Sigma_w, \tag{A3}$$

hence $\mathrm{Tr}\,\Sigma_w=0$. For Σ_w of odd rank, $(\sigma_i,\sigma_i,\sigma_j,\sigma_k,\dots)$ we can always choose $\Sigma_{\bar w}$ to be a basis element, Σ_k which is not a factor of Σ_w. For Σ_w even, we choose Σ_k^- to be a factor of Σ_w. [Note, for n odd, this procedure fails in the case of the canonical element $\sigma(n)=\sigma_1\sigma_2\cdots\sigma_n$.] For both cases it is clear that

$$\sigma_k \Sigma_w = - \Sigma_w \sigma_k. \tag{A4}$$

From group theory, we have

$$\sum_i \chi(a_i)\chi(a_i^{-1}) = N, \tag{A5}$$

where $\chi(a_i)$ is the character of any irreducible representation matrix of element a_i. From what has been shown above we have only one $+1$ and one -1 element contributing to the sum

$$2 \cdot 2^{2m} = \chi(1)\chi(1) + \chi(-1)\chi(-1) = 2 \cdot \mathrm{Dim}^2(\chi), \tag{A6}$$

where Dim denotes the dimension of the representation χ. Therefore, we know that the irreducible representation of a Clifford algebra of order $n = 2m$ must be of dimension 2^m, as given by Rasheviski's method.[6]

There is another theorem in group theory:

$$\sum_\nu n_\nu^2 = N, \tag{A7}$$

where n_ν is the degree or dimension of the irreducible representation D_ν, and ν runs through all possible different inequivalent irreducible representation of a finite group. In the case of a Clifford algebra, $N = 2 \cdot 2^{2m}$ and every irreducible representation of the Clifford algebra is at the same time an irreducible representation of the associated group. But any irreducible representation that is at the same time an irreducible representation of the Clifford algebra and the associated group must be of 2^m dimension. If there were more than one such kind of representation, then we would have

$$N = 2 \cdot 2^{2m} = 2^{2m} + 2^{2m} + \text{other representations}, \tag{A8}$$

and one possibility is to have two inequivalent irreducible representations of order $2m$ with no others. However, this leaves no room for other representations of the associated group such as the trivial identity representation that maps each element on to $+1$. Therefore we conclude that there is only one inequivalent irreducible representation of a Clifford algebra of order $n = 2m$.

APPENDIX B: EQUIVALENCE OF TWO CLIFFORD ALGEBRAS A_n^t AND A_n^{t-2}

Two Clifford algebras A_n^t and A_n^{t-2}, which differ only by one basis element with a different square, can only be equivalent if the orders of the two algebras are even and the signature index is

$$t = 2 (\mathrm{mod}\ 4).$$

Proof: Given the two algebras with the basis elements

$$A_n^t: \ \{\sigma_1, \sigma_2, ..., \sigma_n\}, \quad A_n^{t-2}: \ \{\sigma_1', \sigma_2, ..., \sigma_n\}, \tag{B1}$$

which differ only by the replacement of σ_1 by σ_1', where $(\sigma_1')^2 = -\sigma_1^2$. If $A_n^t \cong A_n^{t-2}$, it must be possible to construct within A_n^t an element w that behaves like σ_1' in A_n^{t-2}, namely, we have the following.

(1) w anticommutes with all basis elements σ_2 through σ_n.

(2) w cannot be formed exclusively from σ_2 through σ_n, because in the case A_n^{t-2} would be A_{n-1}. In others words, σ_1' is not algebraically independent.

(3) $w^2 = -\sigma_1^2 = -e$.

Consider a term τ of w that contains σ_j but does not contain σ_{k_0} as a factor. Then if

$$\sigma_{k_0}\tau = -\tau\sigma_{k_0}, \tag{B2}$$

we would have

$$\sigma_j \tau = \tau \sigma_j. \tag{B3}$$

So, to maintain (1), w must be either

$$\prod_{j=2}^{n} \sigma_j \quad \text{or} \quad \prod_{j=1}^{n} \sigma_j \equiv \sigma(n). \tag{B4}$$

The first possibility is excluded by (2), so

$$w = \sigma(n), \tag{B5}$$

and when n is odd, (1) cannot be true. Finally (3) and Eq. (24) entail that $t = 2$ (mod 4). Otherwise, this is impossible and $A_n^t \not\cong A_n^{t-2}$.

[1] W. K. Clifford, Am. J. Math. **1**, 350 (1878).
[2] D. Hestenes, *Space Time Algebra* (Gordon and Beach, New York, 1966).
[3] S. Lie and E. Engel, Theorie der Transformations Gruppen (Teubner, Leipzig, 1932).
[4] H. Weyl, *Theory of Groups and Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1931) (English translation).
[5] I. Porteous, *Topological Geometry* (Van Nostrand–Reinhold, London, 1969); and especially D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus* (Reidel, Dordrecht, 1984).
[6] P. K. Rashevskii, Am. Math. Soc. Transl. (Ser. 2) **6**, 1 (1957).
[7] A. Ramakrishnan, "The weak interaction Hamiltonian in L matrix theory," presented at Seminar at Rand Corp., July 1970.
[8] C. P. Poole, Jr. and H. A. Farach, Found. Phys. **12**, 719 (1982).
[9] T. Srivastava, Acta Phys. Austriaca **54**, 287 (1982).
[10] K. R. Greider, Phys. Rev. Lett. **44**, 1718 (1980).
[11] K. R. Greider, Found. Phys. **14**, 467 (1984).
[12] S. A. Basri and A. O. Barut, Int. J. Theor. Phys. **22**, 691 (1983).
[13] P. Jordan and E. P. Wigner, Z. Phys. **47**, 631 (1929).
[14] R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935).
[15] H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1967), Chap. VIII.
[16] M. F. Atiyah, R. Bott, and A. Shapiro, Topology **3** (Suppl. I), 3 (1964).
[17] N. Salingaros and M. Dresden, Adv. Appl. Math. **4**, 1 (1983).
[18] N. Salingaros, Adv. Appl. Math. **4**, 31 (1983).
[19] N. Salingaros, J. Math. Phys. **23**, 1 (1982).
[20] N. Salingaros, J. Math. Phys. **22**, 1919, 2096 (1981); **23**, 31 (1982).
[21] N. Salingaros, J. Math. Phys. **22**, 226 (1981).
[22] R. Coquereaux, Phys. Lett. B **115**, 389 (1982).
[23] W. Greub *Multilinear Algebra* (Springer, New York, 1978), 2nd ed.
[24] N. Salingaros, J. Math. Phys. **25**, 738 (1984).
[25] H. W. Braden, J. Math. Phys. **26**, 613 (1985).

On Diophantine equations and nontrivial Racah coefficients

Andrew Bremner
Department of Mathematics, Arizona State University, Tempe, Arizona 85287

(Received 5 November 1985; accepted for publication 14 January 1985)

Some families of zeros of weight-1 $6j$ coefficients are given, each in terms of four parameters. They arise from a geometrical investigation of certain Diophantine equations. Some general remarks on the solutions of Diophantine equations are also made.

I. INTRODUCTION

In this paper we make some remarks of a general nature concerning an overall appraisal of the problem of solving types of Diophantine equations; and apply specific number-theoretic and geometric methods to solving a Diophantine system that arises in determining nontrivial Racah coefficients of weight 1. This extends results of Brudno and Louck.[1]

A Racah operator (which was originally introduced in work on spectroscopy) is a linear operator acting on a particular abstract Hilbert space, and gives rise to the Racah coefficients. See Biedenharn and Louck[2] for a full discussion, together with motivation for the importance of their study. Considerable interest has been shown in the nontrivial zeros of the Racah coefficients, because these determine vector spaces belonging to the null space of a Racah operator, and accordingly give structural information concerning the operator itself. See Racah[3] and Judd,[4] who extends the Lie algebraic method of the former. Koozekanani and Biedenharn[5] list over a thousand nontrivial zeros of the Racah coefficient, obtained by computer calculation, and Vanden Berghe *et al.*[6] give further examples of nontrivial zeros, again by exploiting ideas from Lie algebra.

A different approach to the classification of the zeros of the Racah coefficients has been by Brudno.[7] Here, it is observed that the explicit expression for each of the coefficients, as given by Racah,[8] is an alternating sum, and the author bases his classification on the number of nonzero terms occurring in this sum. This is shown to be equivalent to a classification by weights of the corresponding Racah operator by Brudno and Louck.[9]

We introduce notation for the $6j$ coefficient $\left\{\begin{smallmatrix} a & b & e \\ d & c & f\end{smallmatrix}\right\}$, which up to sign is equal to a Racah coefficient. The coefficient is given by a polynomial function in the arguments a,b,c,d,e,f, which represent angular momentum quantum numbers; an explicit realization of this polynomial is given in Biedenharn and Louck,[2] p. 142. The domain of definition of a,b,c,d,e, and f is that they must be non-negative integers or non-negative half integers, satisfying the "triangle condition" on (a,b,e), (a,c,f), (b,d,f), and (c,d,e) [where the triangle condition on (p,q,r) is that $-p+q+r, p-q+r$, and $p+q-r$ are all non-negative integers].

An alternative notation for the $6j$ coefficient is the 4×3 array of Bargmann[10]

$$\begin{Bmatrix} a & b & e \\ d & c & f\end{Bmatrix} = \begin{Bmatrix} d+f-b & c+f-a & c+d-e \\ a+f-c & b+f-d & a+b-e \\ d+e-c & b+e-a & b+d-f \\ a+e-b & c+e-d & a+c-f\end{Bmatrix},$$

which has the advantage of displaying clearly the Regge[11] symmetries, corresponding to row interchanges and column interchanges. The smallest entry in the Bargmann array is called the weight of the corresponding $6j$ coefficient, and is equal to the number that is one less than the number of terms in the alternating sum, as mentioned above.

A nontrivial zero of a $6j$ coefficient is now defined to be a sextuple (a,b,c,d,e,f) of non-negative integers or non-negative half integers, such that all entries in the corresponding Bargmann array are non-negative integers. Since coefficients of weight 0 possess no nontrivial zeros, then nontrivial zeros of $6j$ coefficients have corresponding Bargmann arrays with every entry a positive integer.

The first interesting case is that of weight-1 coefficients, having two terms in the alternating sum expression for the coefficient. This has been studied by Brudno and Louck,[1] with the following results. If $\left\{\begin{smallmatrix} a & b & e \\ d & c & f\end{smallmatrix}\right\} = 0$ and is of weight 1, then by using a Regge symmetry if necessary, it follows that there exist positive integers x,y,u,v,w satisfying

$$xy(x+y+u+v+w) = uvw, \tag{1}$$

with

$$\begin{Bmatrix} a & b & e \\ d & c & f\end{Bmatrix} = \begin{Bmatrix} \tfrac12(x+u+v-1) & \tfrac12(y+u+w-1) & \tfrac12(x+y+v+w-2) \\ \tfrac12(x+w) & \tfrac12(y+v) & \tfrac12(x+y+u-1)\end{Bmatrix}. \tag{2}$$

Equation (1) is equivalent to the Diophantine system

$$X^3+Y^3+Z^3 = U^3+V^3+W^3,$$
$$X+Y+Z = U+V+W, \tag{3}$$

under the transformations

$$X = x-y+z,\quad U = u+v-w,$$
$$Y = -x+y+z,\quad V = u-v+w,$$
$$Z = u-v-w,\quad W = x+y+z;$$
$$x = \tfrac12(W-Y),\quad u = \tfrac12(U+V),$$
$$y = \tfrac12(W-X),\quad v = \tfrac12(U-Z),$$
$$z = \tfrac12(X+Y),\quad w = \tfrac12(V-Z).$$

The authors show that, using symmetry, solutions of (1) are in correspondence with 3×3 arrays of positive integers

$$\begin{array}{ccc|c} k & l & m & x \\ n & p & q & y \\ r & s & t & z \\ \hline u & v & w & \end{array}, \tag{4}$$

 0022-2488/86/051181-04$02.50

where $x = klm, y = npq, z = rst, u = knr, v = lps, w = mqt$, and $z = x + y + u + v + w$. They invoke a parametric solution of Eqs. (3), due to Gérardin (see Dickson,[12] pp. 565 and 713) in 1916:

$$\begin{aligned} X &= 2p^2 - 10pq + 12q^2, \quad U = 2pq, \\ Y &= p^2 - 5pq + 6q^2, \quad V = pq, \\ Z &= -2p^2 + 9pq - 6q^2, \quad W = p^2 - 9pq + 12q^2. \end{aligned} \tag{5}$$

Hence they recover an array (4) given by

$$\begin{array}{ccc|c} -\frac{3}{2}q & \frac{1}{3}(2p-3q) & 2 & x \\ -p & 1 & \frac{1}{2}(p-q) & y \\ 1 & \frac{3}{2}(p-2q) & p-3q & z' \\ \hline u & v & w & \end{array}$$

thereby producing a parametrized infinity of nontrivial zeros of weight-1 6j coefficients (necessarily p is even, and $q > p$).

No further results are given, and indeed the authors state "it appears that the general solution of the pair of Diophantine equations (3) is not known."[1] It is the intention of this paper to investigate these equations more carefully, deducing further parametrized infinities of zeros of weight-1 6j coefficients (this time with four parameters).

II. A FIRST SOLUTION

The basic state of knowledge regarding Diophantine equations seems to be that if you have a particular equation representing a geometric curve, then there is a well-established body of theory and you can expect to say almost everything about the rational and integral points upon it. There is a major subdivision of cases, according as to the genus of the curve in question. A curve of genus 0 is said to be rational, and all the points upon it may be described parametrically. In particular, all the rational points are known. For curves of genus greater than 0, there is no polynomial parametrization of all the rational points (indeed, by recent work of Faltings, a curve of genus greater than 1 can have only finitely many rational points), but all the rational points can (at least in theory) be described fully. The corresponding work involved may, of course, require considerable effort. Consider, for example, the curve $y^2 = x(x^2 + 877)$ where the smallest rational solution has for its x coordinate a fraction with a 42-digit numerator! See Bremner and Cassels.[13]

If the Diophantine equation represents a surface, then the theory is far less complete. Certain surfaces are rational, in that the points of the surface are in one to one correspondence with points of a plane, and all the points on such surfaces can be described parametrically. Otherwise, one can hope to say something about the rational points upon a surface, although a complete description is usually beyond reach. And if you have any higher-dimensional geometric object representing a nonrational variety, then very little in general can be said—in particular, to describe fully all the rational points is, at present, a hopeless task.

Equations (3) represent geometrically a cubic threefold, the object beyond a surface in increasing dimension. It is thus not clear that a complete description of the rational points will be at all forthcoming. But not everything is hopeless. The intersection of the threefold (3) with a hyperplane is simply a surface (just as the intersection of a surface with a plane is a curve) and so one might hope to obtain information about the surfaces which lie on (3).

Consider in particular the Gérardin solution (5). As the ratio p/q varies, the locus is a quadratic curve that lies on (3). It also satisfies the linear relation

$$X - U = 2(Y - V), \tag{6}$$

the equation of a hyperplane. Accordingly, the intersection of the threefold (3) with the hyperplane (6) is a surface, which we know contains a rational quadratic curve. Since it contains one rationally parametrizable curve, it is plausible that it may contain others. To investigate this, transform the equations by

$$\begin{aligned} X &= r + 2p, \quad U = r - 2p, \\ Y &= t + p, \quad V = t - p, \\ Z &= s - 3p, \quad W = s + 3p, \end{aligned}$$

so that the equation of the surface is simply

$$\begin{aligned} &(r+2p)^3 + (t+p)^3 + (s-3p)^3 \\ &\quad = (r-2p)^3 + (t-p)^3 + (s+3p)^3, \end{aligned}$$

i.e., $p(2r^2 - 3s^2 + t^2 - 6p^2) = 0$.

The surface is thus reducible, with two components: the plane $p = 0$ and the quadric surface $2r^2 - 3s^2 + t^2 - 6p^2 = 0$. It is easy to find *all* the rational points on the two components and hence on the cubic surface. For the plane $p = 0$, the general point is $(r,s,t,u) = (\alpha,\beta,\gamma,0)$ leading back to $(X,Y,Z,U,V,W) = (\alpha,\beta,\gamma,\alpha,\beta,\gamma)$. This rather dull solution to Eqs. (3) does not give any 6j-coefficient zeros, because the corresponding x,y,v,w at (1) satisfy $x + w = y + v = 0$ and so x,y,v, and w cannot all be positive. However, for the quadric surface $2r^2 + t^2 = 3(s^2 + 2p^2)$, the general solution can be obtained by the parametrization

$$t + r\sqrt{-2} = (1 + \sqrt{-2})(\alpha + \beta\sqrt{-2})(\gamma - \delta\sqrt{-2}),$$
$$s + p\sqrt{-2} = (\alpha + \beta\sqrt{-2})(\gamma + \delta\sqrt{-2}),$$

i.e., by

$$\begin{aligned} r &= \alpha\gamma + 2\beta\delta + \beta\gamma - \alpha\delta, \\ s &= \alpha\gamma - 2\beta\delta, \\ t &= \alpha\gamma + 2\beta\delta - 2\beta\gamma + 2\alpha\delta, \\ p &= \qquad\qquad\quad \beta\gamma + \alpha\delta, \end{aligned}$$

giving

$$\begin{aligned} X &= \alpha\gamma + 2\beta\delta + 3\beta\gamma + \alpha\delta, \quad U = \alpha v + 2\beta\delta - \beta\gamma - 3\alpha\delta, \\ Y &= \alpha\gamma + 2\beta\delta - \beta\gamma + 3\alpha\delta, \quad V = \alpha\gamma + 2\beta\delta - 3\beta\gamma + \alpha\delta, \\ Z &= \alpha\gamma - 2\beta\delta - 3\beta\gamma - 3\alpha\delta, \\ &\quad W = \alpha\gamma - 2\beta\delta + 3\beta\gamma + 3\alpha\delta. \end{aligned}$$

In terms of x,y,z,u,v,w, we have

$$\begin{aligned} x &= -2\beta\delta + 2\beta\gamma, \quad u = \alpha\gamma + 2\beta\delta - 2\beta\gamma - \alpha\delta, \\ y &= -2\beta\delta + \alpha\delta, \quad v = 2\beta\delta + \beta\gamma, \\ z &= \alpha\gamma + 2\beta\delta + \beta\gamma + 2\alpha\delta, \quad w = 2\beta\delta + 2\alpha\delta, \end{aligned}$$

corresponding to the 3×3 array

$$\begin{array}{ccc|c} \gamma-\delta & \beta & 2 & x \\ \alpha-2\beta & 1 & \delta & y \\ 1 & \gamma+2\delta & \alpha+\beta & z \\ \hline u & v & w & \end{array}. \qquad (7)$$

Thus, provided $\alpha>2\beta>0$, $\gamma>\delta>0$, the array has positive entries, and leads to a four-parameter family of zeros of $6j$ coefficients.

Notice that the Gérardin solution is obtained by specializing in (7) to

$$(\alpha,\beta,\gamma,\delta) = (\tfrac{1}{3}(p-6q), \tfrac{1}{3}(2p-3q), \tfrac{1}{2}(p-4q), \tfrac{1}{2}(p-q)).$$

What we have done here is to find *all* solutions of (3), subject to the restriction (6).

III. MORE GENERAL SOLUTIONS

We can work more generally than in Sec. II. The most general hyperplane containing the Gérardin curve (5) is

$$\theta(X-2Y)+\phi(U-2V) + \psi(X+Y+Z-U-V-W)=0,$$

for constants θ,ϕ,ψ. Thus the most general cubic surface arising as an intersection of (3) with a hyperplane, and containing the Gérardin curve, is given by

$$\begin{aligned} X^3+Y^3+Z^3 &= U^3+V^3+W^3, \\ X+Y+Z &= U+V+W, \\ X-2Y &= \lambda(U-2V), \end{aligned}$$

for an arbitrary parameter λ (in the previous instance, $\lambda=1$). Substituting

$$\begin{aligned} &X=\lambda r+2p, \quad U=r-2p, \\ &Y=\lambda t+p, \quad V=t-p, \\ &Z=s-\lambda r-\lambda t-3p, \quad W=s-r-t+3p, \end{aligned}$$

then the resulting cubic surface has equation

$$\begin{aligned} V_\lambda:\quad & 12p^3+(\lambda-1)(5r+8t)p^2 \\ &+p[6s^2-6(\lambda+1)(r+t)s \\ &+(\lambda^2+1)(r^2+6rt+2t^2)] \\ &-(\lambda-1)(r+t)[s^2-(\lambda+1)(r+t)s \\ &+(\lambda^2+\lambda+1)rt]=0. \end{aligned}$$

A basic geometric result concerning cubic surfaces (see Swinnerton-Dyer[14]) is that there is a rational parametrization of *all* the rational points on the surface if and only if the surface contains a set of two, three, or six skew (i.e., nonintersecting) straight lines, the set of lines as a whole being defined over the rationals (so a set of Galois conjugates is allowed, for example). Now in any specific instance, as for example with V_λ, it is straightforward although tedious, to write down the equations of all the straight lines on the surface. Indeed, it is classically known that there are precisely 27 such straight lines.

We do not carry out here these awkward computations, but restrict as in Sec. II to more specific observations. Notice for V_λ that the plane $p=0$ cuts in the line $\{p=0, r=-t\}$ and the quadric $\{p=0,\ s^2-(\lambda+1)s(r+t)+(\lambda^2+\lambda+1)rt=0\}$. In particular, when $\lambda=0$, the intersection is the three lines $\{p=0, r=-t\}$, $\{p=0, r=s\}$, and $\{p=0, s=t\}$. But when $\lambda=0$, V_λ also contains the straight line $\{r=2t, t=2p\}$. Thus in the case $\lambda=0$, we have two skew lines $\{p=0, r=s\}$ and $\{r=2t, t=2p\}$, each of which is defined over the rationals. This leads with the appropriate calculations to the parametrization of all points of V_0 by

$$\begin{aligned} r &= 25\alpha^2\delta^2-10\alpha^2\gamma\delta+\alpha^2\gamma^2+\alpha\beta\gamma^2 \\ &\quad -6\alpha\beta\gamma\delta+8\alpha\beta\delta^2+4\beta^2\gamma\delta-8\beta^2\delta^2, \\ s &= -\alpha^2\gamma\delta+5\alpha^2\delta^2+\alpha\beta\gamma^2-6\alpha\beta\gamma\delta \\ &\quad +8\alpha\beta\delta^2+\beta^2\gamma^2-2\beta^2\gamma\delta, \\ t &= -2\alpha^2\gamma\delta+10\alpha^2\delta^2+\alpha\beta\gamma^2-6\alpha\beta\gamma\delta \\ &\quad +5\alpha\beta\delta^2+\beta^2\gamma^2-4\beta^2\gamma\delta+4\beta^2\delta^2, \\ p &= \delta(-\alpha^2\gamma+5\alpha^2\delta+\beta^2\gamma-2\beta^2\delta). \end{aligned}$$

In terms of the notation at (4), this corresponds to the following array:

$$\begin{array}{ccc|c} \alpha(\gamma-4\delta)+\beta(\gamma-\delta) & -\tfrac{1}{2}(\gamma-5\delta) & \alpha & x \\ \tfrac{1}{2}(\alpha(\gamma-5\delta)+\beta(\gamma-2\delta)) & 1 & -\alpha(\gamma-5\delta)-\beta\delta & y \\ 1 & -\alpha^2(\gamma-5\delta)+\beta^2(\gamma-2\delta) & \tfrac{3}{2}\delta & z \\ \hline u & v & w & . \end{array} \qquad (8)$$

The entries are certainly positive when

$$\alpha,\beta>0, \quad 5\delta>\gamma>4\delta>0, \quad \frac{\gamma-2\delta}{5\delta-\gamma}>\frac{\alpha}{\beta}>\frac{\delta}{5\delta-\gamma},$$

and thus again we obtain a four-parameter family of zeros of $6j$ coefficients.

It may be that for further rational values of λ, the surface V_λ admits a set of two, three, or six skew lines defined over the rationals, and hence admitting a parametrization of all its points. In each such case there will arise a homogeneous four-parameter solution to Eqs. (3) and (1), giving a four-parameter family of zeros of weight-1 $6j$ coefficients. But we have not progressed further in this direction.

When V_λ fails to have a rational set of the required number of lines, there cannot arise a four-parameter solution; but there still may be parametrizable *curves* on V_λ, affording a homogeneous two-parameter solution (as found originally by Gérardin). Finally, there is no particular reason why one has to restrict attention to intersections containing the Gérardin curve as we have done above. Any hyperplane intersection of (3) arises in a cubic surface, and one might try analyzing, for example, solutions of (3) which satisfy in addition, say,

$$X - U = Y - V.$$

It seems plausible that in this manner one can obtain a large number of parametrized solutions to (3), and hence to zeros of weight-1 $6j$ coefficients.

As a final summary, we make explicit the formulas obtained at (7) and (8) for the zeros of weight-1 $6j$ coefficients. Substituting the values given by (7) into (2) produces the $6j$ coefficient

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix}, \tag{9}$$

with

$$a = \tfrac{1}{2}(\alpha\gamma - \alpha\delta + \beta\gamma + 2\beta\delta - 1), \quad b = \tfrac{1}{2}(\alpha\gamma + 2\alpha\delta - 2\beta\gamma + 2\beta\delta - 1),$$
$$c = \tfrac{1}{2}(\alpha\delta + \beta\gamma), \quad d = \alpha\delta + \beta\gamma, \quad e = \tfrac{1}{2}(3\alpha\delta + 3\beta\gamma - 2), \quad f = \tfrac{1}{2}(\alpha\gamma - 2\beta\delta - 1), \tag{9'}$$

corresponding to the Bargmann array

$$\begin{Bmatrix} 2\beta(\gamma-\delta) & \delta(\alpha-2\beta) & 1 \\ \alpha(\gamma-\delta)-1 & \gamma(\alpha-2\beta)-1 & (\gamma-\delta)(\alpha-2\beta) \\ 2(\alpha\delta+\beta\gamma)-1 & 3\alpha\delta-1 & 2\delta(\alpha+\beta) \\ 3\beta\gamma-1 & (\alpha\delta+\beta\gamma)-1 & \beta(\gamma+2\delta) \end{Bmatrix}. \tag{10}$$

Since $(\alpha,\beta,\gamma,\delta)$ and $(-\alpha,-\beta,-\gamma,-\delta)$ correspond to the same coefficient, we may without loss of generality assume that $\alpha > 0$. Then the entries of (10) are positive if and only if

$$\alpha > 2\beta > 0, \quad \gamma > \delta > 0. \tag{11}$$

Further, if

$$\alpha,\beta,\gamma,\delta \in Z, \quad \alpha \equiv \gamma \equiv 1 \bmod 2, \quad \beta \equiv \delta \bmod 2, \tag{12}$$

then a, b, c, d, e, and f at (9') are all integers.

Consequently, (9) and (9'), subject to conditions (11) and (12), provide an explicit realization of a parametrized family of zeros of weight-1 $6j$ coefficients.

As specific numerical illustration we give the nontrivial zeros arising from (9), (9') in which the arguments are at most 20 (with corresponding α, β, γ, δ in following parentheses):

$$\begin{Bmatrix} 5 & 5 & 8 \\ 6 & 3 & 3 \end{Bmatrix} (3,1,3,1); \quad \begin{Bmatrix} 9 & 6 & 11 \\ 8 & 4 & 6 \end{Bmatrix} (3,1,5,1); \quad \begin{Bmatrix} 13 & 7 & 14 \\ 10 & 5 & 9 \end{Bmatrix} (3,1,7,1);$$
$$\begin{Bmatrix} 17 & 8 & 17 \\ 12 & 6 & 12 \end{Bmatrix} (3,1,9,1); \quad \begin{Bmatrix} 7 & 10 & 11 \\ 8 & 4 & 6 \end{Bmatrix} (5,1,3,1); \quad \begin{Bmatrix} 13 & 13 & 14 \\ 10 & 5 & 11 \end{Bmatrix} (5,1,5,1);$$
$$\begin{Bmatrix} 19 & 16 & 17 \\ 12 & 6 & 16 \end{Bmatrix} (5,1,7,1); \quad \begin{Bmatrix} 9 & 15 & 14 \\ 10 & 5 & 9 \end{Bmatrix} (7,1,3,1); \quad \begin{Bmatrix} 17 & 20 & 17 \\ 12 & 6 & 16 \end{Bmatrix} (7,1,5,1);$$
$$\begin{Bmatrix} 11 & 20 & 17 \\ 12 & 6 & 12 \end{Bmatrix} (9,1,3,1); \quad \begin{Bmatrix} 8 & 14 & 20 \\ 14 & 7 & 4 \end{Bmatrix} (3,1,5,3).$$

In similar manner, the values given by (8) may be substituted into (2), resulting in an explicit parametrization for zeros of weight-1 $6j$ coefficients. But the arithmetic is now much more cumbersome, and we refrain from giving any details.

[1] S. Brudno and J. D. Louck, J. Math. Phys. **26**, 2092 (1985).
[2] L. C. Biedenharn and J. D. Louck, *The Racah–Wigner Algebra in Quantum Theory*, Vol. 9, *Encyclopedia of Mathematics and its Applications*, edited by G.-C. Rota (Addison–Wesley, Reading, MA, 1981).
[3] G. Racah, Phys. Rev. **76**, 1352 (1949).
[4] B. R. Judd, "Topics in atomic theory," in *Topics in Atomic and Nuclear Physics* (Caxton, Christchurch, New Zealand, 1970), pp. 1–60.
[5] S. H. Koozekanani and L. C. Biedenharn, Rev. Mex. Fis. **23**, 327 (1974).
[6] G. Vanden Berghe, H. De Meyer, and J. Van der Jeugt, J. Math. Phys. **25**, 751, 2585 (1984); G. Vanden Berghe and H. De Meyer, *ibid.* **25**, 772 (1984); J. Van den der Jeugt, G. Vanden Berghe, and H. De Meyer, J. Phys. A: Math. Gen. **16**, 1377 (1983).
[7] S. Brudno, J. Math. Phys. **26**, 434 (1985).
[8] G. Racah, Phys. Rev. **62**, 438 (1942).
[9] S. Brudno and J. D. Louck, J. Math. Phys. **26**, 1125 (1985).
[10] V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962).
[11] T. Regge, Nuovo Cimento **11**, 116 (1959).
[12] L. E. Dickson, *Theory of Numbers*, Vol. 2 (Chelsea, New York, 1952) (reprint).
[13] A. Bremner and J. W. S. Cassels, Math. Comp. **42**, 257 (1984).
[14] H. P. F. Swinnerton-Dyer, "Applications of algebraic geometry to number theory," in *Proceedings of the Symposia in Pure Mathematics*, Vol. 20 (Am. Math. Soc., Providence, RI, 1971), pp. 1–52.

On the exceptional equivalence of complex Dirac spinors and complex space-time vectors

Patrick L. Nash
Division of Earth and Physical Sciences, The University of Texas at San Antonio, San Antonio, Texas 78285-0663

(Received 23 July 1985; accepted for publication 14 January 1986)

It is well known that there exists an equivalence of $\mathbb{R}^8$ vectors and spinors, which has its roots in the rotational symmetry of the Dynkin diagram for D_4. We endow $\mathbb{R}^8$ with a real action of $\overline{SO(4,4)}$, and restrict to $\overline{SO(3,1)}$. Under this restriction the $\mathbb{R}^8$ spinor decomposes into the direct sum of two real Dirac spinors, while the $\mathbb{R}^8$ vector decomposes into a real space-time vector plus four real scalars. The equivalence is preserved under this restriction; it is shown that it is realized in the (exceptional) equivalence of a complex Dirac spinor and a complex space-time vector.

I. INTRODUCTION

For pedagogical reasons we shall formulate the exceptional equivalence described in the abstract within the context of a problem that arises in classical mechanics. In the setting that we describe, we have no anticommuting operators that might, perhaps, obscure the underlying geometrical relationship that exists between a complex Dirac spinor and a complex space-time vector. However, the implications of this relationship are obviously relevant to the discussion of particle content of multiplets in grand unified and supersymmetric field theories.

In Ref. 1 a Lagrangian model of a classical electron with spin, possessing an intrinsic magnetic dipole moment and vanishing (in a rest frame) electric dipole moment is formulated. In this model, a real-valued eight-component $\overline{O(3,3)}$ spinor ψ is employed to carry the spin degrees of freedom. Several physical observables that, in part, specify the classical state of the electron are constructed from ψ. These include the spin tensor of the electron, as well as an orthonormal space-time frame that is transported along the worldline of the particle. The timelike member of this tetrad is forced to be always parallel to the four-velocity by subjecting ψ to certain constraints. These constraints lead to complications in quantizing this theory.

Accordingly, as a first step in the quantization program, it is natural to attempt to separate the dynamical degrees of freedom possessed by ψ from the constrained components. In this paper we investigate a decomposition of ψ into components with respect to a spin frame, which, when the spin frame is chosen appropriately, may aid in the identification of free versus constrained components. (This identification will be taken up in another paper.) More explicitly, we shall regard ψ as an element of the vector space $\mathbb{R}^8$, which we endow with a $\overline{SO(4,4)}$ invariant metric. We call the resulting manifold $V_{4,4}$. [The $\overline{O(3.3)}$ representation alluded to above is contained in the real irreducible representation of $\overline{SO(4,4)}$ carried by $V_{4,4}$.] We pick a $V_{4,4}$ frame and resolve ψ into components with respect to this frame. One could ordinarily expect that the resulting eight spin frame components would transform as eight scalars under the Lorentz group. However, we originally employ a formalism in which four of these eight frame components comprise a Minkowski space-time vector, while the remaining four spin frame components are scalars. The simple reason that we can make the transition from a set of four real scalars to a real space-time vector is because a spin frame on $V_{4,4}$ directly defines a space-time frame $E^{\alpha}_{(\mu)}$, and thereby provides the mechanism for going from scalars to vector (and back again) via $W^{\alpha} = W^{(\mu)}E^{\alpha}_{(\mu)}$ (and $W_{(\mu)} = W_{\alpha}E^{\alpha}_{(\mu)}$). See Sec. III for details.

In the next section we define reduced generators of the basic spinor representations of $\overline{SO(4,4)}$. These generators, we call them tau matrices, satisfy a remarkable identity, which is formulated in (2.41). This identity leads directly to the proof of the exceptional equivalence to which this paper is devoted: We shall explicitly prove that there exists a one-to-one invertible linear mapping from a pair of real Dirac spinors to a pair of real Minkowski space-time vectors.

II. SPINOR REPRESENTATIONS OF $\overline{SO(4,4)}$

We shall construct a real reducible 16×16 matrix repesentation of $\overline{SO(4,4)}$ utilizing the Clifford algebra approach of Brauer and Weyl.[2] We shall see, as is in fact well known from the general theory, that there are two inequivalent real 8×8 irreducible basic spinor representations $D^{(1)}$ and $D^{(2)}$ of $\overline{SO(4,4)}$. We assume that the real spinor ψ that carries the spin degrees of freedom in our model transforms under $D^{(1)}$ [although in practice, we always restrict our attention to a Lorentz subgroup of $\overline{SO(4,4)}$]. To construct this representation of $\overline{SO(4,4)}$ we define generators $\sigma^{A'}$, $A',B',... = 1,...,8$ of the pseudo-Clifford algebra $C_{4,4}$ that anticommute and have square ± 1. The generators of $\overline{SO(4,4)}$ are defined, up to a constant factor of $-\frac{1}{4}$, as the commutators of the $\sigma^{A'}$. However, prior to the explicit construction of a representation of the $\sigma^{A'}$, it is convenient to first define the generators of $C_{3,3}$, which then may be used in the representation of the $\sigma^{A'}$. Accordingly, let $\Gamma^A \leftrightarrow \Gamma^{Aa}{}_b$ $A,B,... = 1,...,6$ and $a,b,... = 1,...,8$ denote six real 8×8 matrices that generate an irreducible representation of the 2^{3+3}-dimensional pseudo-Clifford algebra $C_{3,3}$. The Γ^A verify

 0022-2488/86/051185-06$02.50

$$\Gamma^A\Gamma^B + \Gamma^B\Gamma^A = 2g^{AB}I, \tag{2.1}$$

where

$$g^{AB} = g_{AB} = \text{diag}(1,1,1,-1,-1,-1), \tag{2.2}$$

and I denotes the 8×8 unit matrix and is sometimes suppressed. We shall also need

$$\Gamma^7 = \Gamma^1\Gamma^2\Gamma^3\Gamma^4\Gamma^5\Gamma^6, \tag{2.3}$$

which verifies

$$\Gamma^A\Gamma^7 + \Gamma^7\Gamma^A = 0 \tag{2.4}$$

and

$$(\Gamma^7)^2 = I. \tag{2.5}$$

The generators of the basic spinor representation of $\overline{SO(3,3)}$ are

$$M^{AB} = -\tfrac{1}{4}[\Gamma^A,\Gamma^B], \tag{2.6}$$

and are fully reducible to the direct sum of two real 4×4 inequivalent irreducible representations of $\overline{SO(3,3)}$ (see Ref. 3).

We introduce a covariant (resp. contravariant) rank two spinor $\sigma_{ab} \leftrightarrow \sigma$ (resp. $\sigma^{ab} \leftrightarrow \sigma^{-1}$) to lower (resp. raise) the $\overline{SO(3,3)}$ spinor indices $a,b,c,...$ as follows: We require that σ satisfy

$$\widetilde{\Gamma}^A\sigma = -\sigma\Gamma^A, \tag{2.7}$$

where the tilde denotes transpose. Then, from (2.3) we have

$$\widetilde{\Gamma}^7\sigma = -\sigma\Gamma^7, \tag{2.8}$$

and from (2.6)

$$\widetilde{M}^{AB}\sigma = -\sigma M^{AB}, \tag{2.9}$$

whence, if $M\in\overline{SO(3,3)}$, then

$$\widetilde{M}\sigma = \sigma M^{-1}, \tag{2.10}$$

so that σ is invariant under $\overline{SO(3,3)}$: $\sigma \mapsto \widetilde{M}\sigma M = \sigma$.

If we set $\tilde{\sigma} = \theta\sigma$, where $\theta = \pm 1$, and examine the symmetry properties of a set of 64 linearly independent real 8×8 matrices, then we find that the 28 linearly independent real 8×8 matrices $\{\sigma\Gamma^A,\sigma\Gamma^7,\sigma\Gamma^A\Gamma^7,\sigma M^{AB}\}$ satisfy $\widetilde{T} = -\theta T$ (e.g., $\widetilde{\Gamma}^A\sigma = \theta\widetilde{\Gamma}^A\tilde{\sigma} = \theta\widetilde{\sigma\Gamma^A} = -\sigma\Gamma^A$, hence $\widetilde{\sigma\Gamma^A} = -\theta\sigma\Gamma^A$), whereas the 36 linearly independent real 8×8 matrices $\{\sigma, \sigma\Gamma^7M^{AB}, \sigma\Gamma^A\Gamma^B\Gamma^C, A<B<C\}$ verify $\widetilde{T} = \theta T$. Since there are 28 (resp. 36) skew-symmetric (resp. symmetric) linearly independent real 8×8 matrices, we conclude that $\theta = 1$, so that σ is symmetric. Hence, for example, $\sigma\Gamma^A \leftrightarrow \sigma_{ac}\Gamma^{Ac}{}_b \equiv \Gamma^A_{ab} = -\Gamma^A_{ba}$ is skew symmetric.

Clearly, all but one of these 64 matrices has vanishing trace.

Lemma 2.1:
tr $T = 0$, for $T\in\{\Gamma^A,\Gamma^7,\Gamma^A\Gamma^7, M^{AB}, \Gamma^7M^{AB}, \Gamma^A\Gamma^B\Gamma^C, A<B<C\}$.

From (2.1) we see that $(\Gamma^4)^2 = -I = (\Gamma^5)^2 = (\Gamma^6)^2$, so that we may assume that each of these matrices is skew symmetric: $\widetilde{\Gamma}^4 = -\Gamma^4$, $\widetilde{\Gamma}^5 = -\Gamma^5$, and $\widetilde{\Gamma}^6 = -\Gamma^6$. Similarly, since $(\Gamma^1)^2 = I = (\Gamma^2)^2 = (\Gamma^3)^2$, we may assume that each of these matrices is symmetric: $\widetilde{\Gamma}^1 = \Gamma^1$, $\widetilde{\Gamma}^2 = \Gamma^2$, and $\widetilde{\Gamma}^3 = \Gamma^3$. Therefore, from (2.7), σ should commute with the skew-symmetric matrices $\{\Gamma^4,\Gamma^5,\Gamma^6\}$ and anticommute with the symmetric matrices $\{\Gamma^1,\Gamma^2,\Gamma^3\}$. Hence one possible choice for σ that satisfies (2.7) is $\sigma = \Gamma^4\Gamma^5\Gamma^6$.

We prove a useful identity satisfied by the Γ^A.

Proposition 2.1: Let $T \leftrightarrow T^a_b$ be an 8×8 matrix with the property that σT is symmetric, i.e., $T_{ab} = \sigma_{ac}T^c{}_b = T_{ba}$, but being otherwise arbitrary. Then

$$I \text{ tr } T = T + \Gamma_A T\Gamma^A + \Gamma^7T\Gamma^7, \tag{2.11}$$

where $\Gamma_B = g_{AB}\Gamma^B$, and I denotes the 8×8 unit matrix.

Proof: Since (2.11) is linear in T, we verify it in turn for $T = I$, Γ^7M^{AB}, and $\Gamma^A\Gamma^B\Gamma^C$, $A<B<C$ (this set of 36 linearly independent real 8×8 matrices verifies $\widetilde{\sigma T} = \sigma T$). We note that $(\Gamma^A)^{-1} = \Gamma_A$, so that $\Gamma_A\Gamma^A = 6I$, and $\Gamma_B\Gamma^A\Gamma^B = (-\Gamma^A\Gamma_B + 2\delta^A_B)\Gamma^B = -4\Gamma^A$.

(i) $T = I$: I tr $T = 8I = I + \Gamma_A I\Gamma^A + \Gamma^7I\Gamma^7 = 8I$.

(ii) $T = \Gamma^7M^{AB}$, or equivalently, $T = \Gamma^A\Gamma^B\Gamma^7$, $A<B$. Here, repeated use of (2.1) and (2.4) yields $\Gamma_C\Gamma^A\Gamma^B\Gamma^7\Gamma^C = -2\Gamma^A\Gamma^B\Gamma^7$, which, when substituted into (2.11), gives

$$\begin{aligned} I \text{ tr } \Gamma^A\Gamma^B\Gamma^7 &= 0 \text{ (by Lemma 2.1)} \\ &= \Gamma^A\Gamma^B\Gamma^7 - 2\Gamma^A\Gamma^B\Gamma^7 + \Gamma^7\Gamma^A\Gamma^B\Gamma^7\Gamma^7 \\ &= 0. \end{aligned}$$

(iii) $T = \Gamma^A\Gamma^B\Gamma^C$, $A<B<C$.

Here repeated use of (2.1) yields $\Gamma_D\Gamma^A\Gamma^B\Gamma^C\Gamma^D = 0$; hence

$$\begin{aligned} I \text{ tr } \Gamma^A\Gamma^B\Gamma^C &= 0 \text{ (by Lemma 2.1)} \\ &= \Gamma^A\Gamma^B\Gamma^C + \Gamma^7\Gamma^A\Gamma^B\Gamma^C\Gamma^7 \\ &= \Gamma^A\Gamma^B\Gamma^C - \Gamma^A\Gamma^B\Gamma^C = 0. \end{aligned}$$ ■

In order to concisely define the $\sigma^{A'}$, $A',B',... = 1,...,8$, we introduce a new set of real 8×8 matrices. We define

$$\tau^{A'} = (\Gamma^A\Gamma^7, -\Gamma^7, I) \tag{2.12}$$

and

$$\bar{\tau}^{A'} = (-\Gamma^A\Gamma^7, \Gamma^7, I), \tag{2.13}$$

where, as before, I denotes the 8×8 unit matrix. By direct evaluation we find that

$$\tilde{\tau}^{A'}\sigma = \sigma\bar{\tau}^{A'}. \tag{2.14}$$

In index notation (2.14) is

$$\tau^{A'}_{ba} = \bar{\tau}^{A'}_{ab}. \tag{2.15}$$

In virtue of (2.1) we also find that

$$\tau^{A'}\bar{\tau}^{B'} + \tau^{B'}\bar{\tau}^{A'} = 2IG^{A'B'} = \bar{\tau}^{A'}\tau^{B'} + \bar{\tau}^{B'}\tau^{A'}, \tag{2.16}$$

where

$$G^{A'B'} = G_{A'B'} = \begin{pmatrix} g_{3,1} & 0 \\ 0 & -g_{3,1} \end{pmatrix} \tag{2.17}$$

and $g_{3,1} = \text{diag}(1,1,1,-1)$ is the metric on Minkowski space-time M_4 (or $M_{3,1}$). Here G (resp. G^{-1}) will be used to lower (resp. raise) primed uppercase Latin indices.

Turning now to the construction of the real reducible 16×16 matrix representation of $\overline{SO(4,4)}$ we define the generators $\sigma^{A'}$, $A',B',... = 1,...,8$ of a real irreducible 16×16 representation of the pseudo-Clifford algebra $C_{4,4}$ according to

$$\sigma^{A'} = \begin{pmatrix} 0 & \tau^{A'}\sigma^{-1} \\ \sigma\bar{\tau}^{A'} & 0 \end{pmatrix}. \tag{2.18}$$

Using (2.16) we find that the $\sigma^{A'}$ verify

$$\sigma^{A'}\sigma^{B'}+\sigma^{B'}\sigma^{A'}=2I_{16\times 16}G^{A'B'}. \tag{2.19}$$

The generators $N^{A'B'}$ of a real reducible 16×16 representation of $\overline{\mathrm{SO}(4,4)}$ may be defined by

$$-4N^{A'B'}=[\sigma^{A'},\sigma^{B'}]. \tag{2.20}$$

Using (2.19) repeatedly, one finds that the $N^{A'B'}$ satisfy the commutation relations for $\overline{\mathrm{SO}(4,4)}$. Using (2.18) and (2.14) we find that the $N^{A'B'}$ may be written as

$$N^{A'B'}=\begin{pmatrix} D^{(1)}(N^{A'B'}) & 0\\ 0 & D^{(2)}(N^{A'B'})\end{pmatrix}, \tag{2.21}$$

where

$$-4D^{(1)}(N^{A'B'})=\tau^{A'}\bar\tau^{B'}-\tau^{B'}\bar\tau^{A'} \tag{2.22}$$

and

$$4\widetilde{D}^{(2)}(N^{A'B'})=\bar\tau^{A'}\tau^{B'}-\bar\tau^{B'}\tau^{A'}, \tag{2.23}$$

which is the direct sum of two real inequivalent 8×8 irreducible representations of $\overline{\mathrm{SO}(4,4)}$. The two representations are not equivalent because, if P is a linear automorphism of $\overline{\mathrm{SO}(4,4)}$ that defines the conjectured equivalence, then $D^{(2)}(N^{A'B'})P=PD^{(1)}(N^{A'B'})$. But from (2.22), (2.23), (2.12), and (2.13) we find that

$$-\widetilde{D}^{(2)}(N^{AB})=D^{(1)}(N^{AB})=M^{AB}, \tag{2.24}$$

$$-\widetilde{D}^{(2)}(N^{A7})=D^{(1)}(N^{A7})=-\tfrac{1}{2}\Gamma^{A}, \tag{2.25}$$

$$\widetilde{D}^{(2)}(N^{A8})=D^{(1)}(N^{A8})=-\tfrac{1}{2}\Gamma^{A}\Gamma^{7}, \tag{2.26}$$

and

$$\widetilde{D}^{(2)}(N^{78})=D^{(1)}(N^{78})=\tfrac{1}{2}\Gamma^{7}. \tag{2.27}$$

Hence from (2.25), P must satisfy $\widetilde{\Gamma}^{A}P=-P\Gamma^{A}$. Using (2.3) we see that this implies that $\widetilde{\Gamma}^{7}P=-P\Gamma^{7}$; but (2.27) demands that $\widetilde{\Gamma}^{7}P=P\Gamma^{7}$, and therefore P is 0, the representations are not equivalent.

In order to prove that σ is invariant under the action of $\overline{\mathrm{SO}(4,4)}$ we define a covariant rank two $\overline{\mathrm{SO}(4,4)}$ spinor Σ by

$$\tilde\sigma^{A'}\Sigma=\Sigma\sigma^{A'}. \tag{2.28}$$

Equations (2.28) and (2.20) imply that

$$\widetilde{N}^{A'B'}\Sigma=-\Sigma N^{A'B'}, \tag{2.29}$$

so that Σ defines a $\overline{\mathrm{SO}(4,4)}$ invariant bilinear form. Since

$$\tilde\sigma^{A'}=\begin{pmatrix} 0 & \sigma\tau^{A'}\\ \bar\tau^{A'}\sigma^{-1} & 0\end{pmatrix}, \tag{2.30}$$

we find that one solution to (2.28) is

$$\Sigma=\begin{pmatrix}\sigma & 0\\ 0 & \sigma^{-1}\end{pmatrix}. \tag{2.31}$$

Hence both $D^{(1)}$ and $D^{(2)}$ preserve σ:

$$\widetilde{D}^{(1)}\sigma=\sigma D^{(1)-1} \tag{2.32}$$

and

$$\widetilde{D}^{(2)}\sigma^{-1}=\sigma^{-1}D^{(2)-1}, \tag{2.33}$$

where $D=D(g)$, $g\in\overline{\mathrm{SO}(4,4)}$ is implicit.

A canonical 2–1 homomorphism $\overline{\mathrm{SO}(4,4)}\to\overline{\mathrm{SO}(4,4)}$, $\overline{\mathrm{SO}(4,4)}\ni g\mapsto L\in\overline{\mathrm{SO}(4,4)}$ may be defined by

$$16L^{A'}{}_{B'}=\operatorname{tr}N^{-1}\sigma^{A'}N\sigma_{B'}, \tag{2.34}$$

where $L\leftrightarrow L^{A'}{}_{B'}$ and

$$N=N(g)=\begin{pmatrix} D^{(1)}(g) & 0\\ 0 & D^{(2)}(g)\end{pmatrix}.$$

Because $L\in\overline{\mathrm{SO}(4,4)}$, the metric G of (2.17) is invariant under automorphism by $\overline{\mathrm{SO}(4,4)}$: $G\to\widetilde{L}GL=G$,

$$L^{A'}{}_{C'}G_{A'B'}L^{B'}{}_{D'}=G_{C'D'}. \tag{2.35}$$

Using (2.20) and (2.34), we find that under the action of $\overline{\mathrm{SO}(4,4)}$

$$L^{A'}{}_{B'}\sigma^{B'}=N^{-1}\sigma^{A'}N. \tag{2.36}$$

Of course, we may also deduce this relationship using the fact that the $L^{A'}{}_{B'}\sigma^{B'}$ satisfy the same anticommutation relations as the $\sigma^{A'}$, and the fact that there is only one irreducible representation of degree 16, up to equivalance.

Using (2.18) in (2.36) we find that, under the action of $\overline{\mathrm{SO}(4,4)}$,

$$L^{A'}{}_{B'}\tau^{B'}=D_1^{-1}\tau^{A'}\widetilde{D}_2^{-1} \tag{2.37}$$

and

$$L^{A'}{}_{B'}\bar\tau^{B'}=\widetilde{D}_2\bar\tau^{A'}D_1, \tag{2.38}$$

where $D=D(g)$ is again implicit, with L given by (2.34), and $D^{(1)}=D_1$, $D^{(2)}=D_2$. By right multiplication of (2.37) by $\bar\tau_{C'}$, and left multiplication of (2.38) by $\tau_{C'}$, and then summing the resulting expressions, we find that

$$2IL^{A'}{}_{C'}=D_1^{-1}\tau^{A'}\widetilde{D}_2^{-1}\bar\tau_{C'}+\tau_{C'}\widetilde{D}_2\bar\tau^{A'}D_1. \tag{2.39}$$

Taking the trace over spinor indices yields

$$8L^{A'}{}_{B'}=\operatorname{tr}\widetilde{D}_2\bar\tau^{A'}D_1\tau_{B'}, \tag{2.40}$$

since

$$\operatorname{tr}D_1^{-1}\tau^{A'}\widetilde{D}_2^{-1}\bar\tau_{B'}=\operatorname{tr}\sigma D_1^{-1}\tau^{A'}\widetilde{D}_2^{-1}\bar\tau_{B'}\sigma^{-1}=\operatorname{tr}\tau_{B'}\widetilde{D}_2\bar\tau^{A'}D_1.$$

If $T\mapsto D_1TD_1^{-1}=T'$ under $\overline{\mathrm{SO}(4,4)}$, and $\widetilde{\sigma T}=\sigma T$, then by (2.32), $\widetilde{\sigma T'}=\sigma T'$; the symmetry property $T_{ba}=T_{ab}$ is preserved under the action of $\overline{\mathrm{SO}(4,4)}$. Moreover, from (2.37) and (2.38) we see that the τ matrices are numerically invariant under combined $\overline{\mathrm{SO}(4,4)}$ and SO(4,4) transformations. For example, $\tau^{A'}=L^{A'}{}_{B'}D_1\tau^{B'}\widetilde{D}_2$. This observation leads us to prove, analogous to (2.11), the following proposition.

Proposition 2.2: If T is any 8×8 matrix satisfying $\widetilde{\sigma T}=\sigma T$ and transforming according to $T\mapsto D_1TD_1^{-1}$ under $\overline{\mathrm{SO}(4,4)}$, then

$$\bar\tau_{A'}T\tau^{A'}=I\operatorname{tr}T. \tag{2.41}$$

Proof: From (2.11)–(2.13), we have

$$\bar{\tau}_{A'} T \tau^{A'} = -\Gamma_A \Gamma^7 T \Gamma^A \Gamma^7 + \Gamma^7 T \Gamma^7 + T$$
$$= (\Gamma^7 T \Gamma^7) + \Gamma_A (\Gamma^7 T \Gamma^7) \Gamma^A + \Gamma^7 (\Gamma^7 T \Gamma^7) \Gamma^7$$
$$= I \operatorname{tr} \Gamma^7 T \Gamma^7 = I \operatorname{tr} T.$$

Equation (2.11) is applicable because $(\sigma \Gamma^7 T \Gamma^7)^T = \sigma \Gamma^7 T \Gamma^7$. ■

For completeness we prove the following proposition.

Proposition 2.3: Let $D^{(1)}(N^{A'B'}) = D_1^{A'B'}$ and $D^{(2)}(N^{A'B'}) = D_2^{A'B'}$; then

$$\bar{\tau}_{C'} D_1^{A'B'} \tau^{C'} = -4 \tilde{D}_2^{A'B'} \tag{2.42}$$

and

$$\tau_{C'} \tilde{D}_2^{A'B'} \bar{\tau}^{C'} = -4 D_1^{A'B'}. \tag{2.43}$$

Proof: Using (2.16) we find that

$$\bar{\tau}_{C'} D_1^{A'B'} = -\tilde{D}_2^{A'B'} \bar{\tau}_{C'} + \delta^{B'}_{C'} \bar{\tau}^{A'} - \delta^{A'}_{C'} \bar{\tau}^{B'} \tag{2.44}$$

and

$$D_1^{A'B'} \tau_{C'} = -\tau_{C'} \tilde{D}_2^{A'B'} + \delta^{A'}_{C'} \tau^{B'} - \delta^{B'}_{C'} \tau^{A'}. \tag{2.45}$$

Right multiplication of (2.44) by $\tau^{C'}$ and summing over C' yields (2.42); right multiplication of (2.45) with $\bar{\tau}^{C'}$ and summing gives (2.43). ■

Under $\overline{\mathrm{SO}(4,4)}$, using (2.37), (2.38), and (2.35), we find that

$$\bar{\tau}_{A'} T \tau^{A'} = \tilde{D}_2 \bar{\tau}_{A'} (D_1 T D_1^{-1}) \tau^{A'} \tilde{D}_2^{-1},$$

or

$$\bar{\tau}_{A'} (D_1 T D_1^{-1}) \tau^{A'} = \tilde{D}_2^{-1} \bar{\tau}_{A'} T \tau^{A'} \tilde{D}_2.$$

If $\widetilde{\sigma T} = \sigma T$, then using (2.41) we find that $\tilde{D}_2^{-1} \bar{\tau}_{A'} T \tau^{A'} \tilde{D}_2 = I \operatorname{tr} T = \bar{\tau}_{A'} T \tau^{A'}$, and hence

$$\bar{\tau}_{A'} (D_1 T D_1^{-1}) \tau^{A'} = \bar{\tau}_{A'} T \tau^{A'}, \tag{2.46}$$

when $\widetilde{\sigma T} = \sigma T$. We shall return to this relationship in the next section.

III. EXCEPTIONAL EQUIVALENCE

In this section we shall use the identity of Proposition 2 to provide a geometrical interpretation of the spinor ψ introduced in Sec. I, which is, in fact, a statement of equivalence of two real Dirac spinors, and a real space-time vector plus four real scalars.

Let $V_{4,4}$ denote the real pseudo-Riemannian space obtained by endowing $\mathbb{R}^8$ with the $\overline{\mathrm{SO}(4,4)}$ invariant metric σ of the previous section. We assume that $V_{4,4}$ carries the real irreducible representation $D^{(1)}$ of $\overline{\mathrm{SO}(4,4)}$. Hence $\psi \in V_{4,4}$ is a real eight-component (reduced) $\overline{\mathrm{SO}(4,4)}$ spinor of the first kind.

We pick a real (constant) spinor $J \in V_{4,4}$ normalized so that

$$\tilde{J} \sigma J = 1, \tag{3.1}$$

but being otherwise arbitrary, and define a real 8×8 matrix T by

$$T = J \tilde{J} \sigma. \tag{3.2}$$

We note that

$$\widetilde{\sigma T} = \sigma T, \tag{3.3}$$

and

$$\operatorname{tr} T = 1, \tag{3.4}$$

the latter property following from $\operatorname{tr} T = \operatorname{tr} J \tilde{J} \sigma = \tilde{J} \sigma J = 1$. For later use we remark that

$$\tilde{J} \sigma \tau^{A'} \bar{\tau}_{B'} J = \delta^{A'}_{B'}, \tag{3.5}$$

which is a consequence of the fact that

$$2 \tilde{J} \sigma \tau^{A'} \bar{\tau}_{B'} J = 2 (\tilde{J} \sigma \tau^{A'} \bar{\tau}_{B'} J)^T = 2 \tilde{J} \tilde{\bar{\tau}}_{B'} \tilde{\tau}^{A'} \sigma J$$
$$= 2 \tilde{J} \sigma \tau_{B'} \bar{\tau}^{A'} J \quad [\text{using } (2.14)]$$
$$= \tilde{J} \sigma \{ \tau^{A'} \bar{\tau}_{B'} + \tau_{B'} \bar{\tau}^{A'} \} J$$
$$= 2 \delta^{A'}_{B'} \tilde{J} \sigma J = 2 \delta^{A'}_{B'}.$$

We substitute (3.2) into (2.41) to obtain a resolution of the identity on $V_{4,4}$:

$$I = \bar{\tau}_{A'} J \tilde{J} \sigma \tau^{A'}. \tag{3.6}$$

Here we have used the facts that $\sigma T = \sigma J \tilde{J} \sigma$ is symmetric, and $\operatorname{tr} T = 1$. From (2.46) we see that this resolution is preserved under the action of $D^{(1)}$ of $\overline{\mathrm{SO}(4,4)}$.

We may obtain a revealing geometrical interpretation of our spinor ψ as follows: Let $\psi \in V_{4,4}$ be arbitrary; then using (3.6) we consider

$$\psi = I \psi = (\bar{\tau}_{A'} J \tilde{J} \sigma \tau^{A'}) \psi = \bar{\tau}_{A'} J (\tilde{J} \sigma \tau^{A'} \psi),$$

that is

$$\psi = \bar{\tau}_{A'} J Z^{A'}, \tag{3.7}$$

where

$$Z^{A'} = \tilde{J} \sigma \tau^{A'} \psi = \tilde{\psi} \sigma \bar{\tau}^{A'} J. \tag{3.8}$$

In order to physically interpret this result, we restrict our attention to a $\overline{\mathrm{SO}(3,1)} \subset \overline{\mathrm{SO}(4,4)}$ subgroup. Under this restriction ψ decomposes into the direct sum of two real four-component Dirac spinors. The $Z^{A'}$ comprise a real space-time vector Z^α, and four real scalars $\{Z^5, Z^6, Z^7, Z^8\}$ under this restriction. Then (3.7) gives the vector and scalar components of ψ with respect to the $V_{4,4}$ frame $\{\bar{\tau}_{A'} J, A' = 1,...,8\}$, these components being given by (3.8).

On the other hand, suppose that we are given quantities $Z^{A'}$, and use them to define a spinor $\psi \in V_{4,4}$ according to (3.7). This map is clearly onto, since an arbitrary $\psi \in V_{4,4}$ is the image of a $Z^{A'}$ given by (3.8). In fact, the map (3.7) is the inverse map of (3.8). To see this we substitute for ψ from (3.7) into (3.8), and obtain an identity:

$$Z^{A'} = \tilde{J} \sigma \tau^{A'} \psi = \tilde{J} \sigma \tau^{A'} (\bar{\tau}_{B'} J Z^{B'}) = Z^{B'} \tilde{J} \sigma \tau^{A'} \bar{\tau}_{B'} J$$
$$= Z^{B'} \delta^{A'}_{B'} \ [\text{using } (3.5)] = Z^{A'}.$$

Therefore we find that the maps (3.7) and (3.8) define a one-to-one invertible linear correspondence between $\overline{\mathrm{SO}(4,4)}$ spinors ψ and the quantities $Z^{A'}$. [Using (2.37) one can easily show that the $Z^{A'}$ do *not* transform as a SO(4,4) vector. Instead the $Z^{A'}$ may be seen to comprise a SO(3,4) vector–scalar pair.] This relation is simply a statement of Cartan's principle of triality.[4–6] It is a special case of the well-known equivalence of $\mathbb{R}^8$ vectors and spinors.[4–8]

We emphasize that the relations (3.7) and (3.8) are of importance in their own right. They assert the equivalence of ψ and $Z^{A'}$. If we restrict our attention to $\overline{\mathrm{O}(3,1)} \subset \overline{\mathrm{SO}(4,4)}$, this is a statement of the exceptional equivalence of a pair of real Dirac spinors (i.e., ψ) and the set of real $\overline{\mathrm{O}(3,1)}$ tensors $\{Z^\alpha, Z^5, Z^6, Z^7, Z^8\}$.

We may state this exceptional equivalence in a slightly different fashion. Let us write the spinor $J \in V_{4,4}$ that gives rise to the $V_{4,4}$ frame $\{\overline{\tau}_{A'}, J\}$ in a form in which the decomposition of $\overline{\mathrm{SO}(3,1)} \subset \overline{\mathrm{SO}(4,4)}$ into a direct sum of irreducible 4×4 representations is explicit:

$$J = \begin{pmatrix} \lambda \\ \tilde{\xi} \end{pmatrix}. \tag{3.9}$$

Here λ and ξ are real four-component Dirac spinors; ξ transforms inversely to λ under $\overline{\mathrm{SO}(3,1)}$. Since $\tilde{J}\sigma J = 2\xi\lambda = 1$, as is well known, J determines an orthogonal space-time frame $E^{\alpha}_{(\mu)}$ given by[1]

$$E^{\alpha}_{(1)} = 2\xi\gamma^\alpha\lambda, \tag{3.10}$$

$$E^{\alpha}_{(2)} = 2\xi\gamma^\alpha\gamma^5\lambda, \tag{3.11}$$

$$E^{\alpha}_{(3)} = n^\alpha - m^\alpha, \tag{3.12}$$

and

$$E^{\alpha}_{(4)} = m^\alpha + n^\alpha, \tag{3.13}$$

where

$$m^\alpha = -\xi\gamma^\alpha\epsilon^{-1}\tilde{\xi} \tag{3.14}$$

and

$$n^\alpha = \tilde{\lambda}\epsilon\gamma^\alpha\lambda. \tag{3.15}$$

Here, the γ^α and γ^5 are the usual Dirac matrices (in a real representation), and ϵ is the symplectic form on Dirac space. With the space-time tetrad so defined, we can take the scalars $Z^{4+\mu}$ and form the vector

$$W^\alpha = E^{\alpha}_{(\mu)} Z^{4+\mu}. \tag{3.16}$$

We may accordingly reformulate the exceptional equivalence of ψ and $Z^{A'}$ as the following theorem.

Theorem: There exists an exceptional equivalence of a pair of real Dirac spinors and a pair of real Minkowski space-time vectors.

Corollary: There exists an exceptional equivalence of a complex Dirac spinor and a complex Minkowski space-time vector.

IV. CONCLUSION

It may be of interest to compare the spinor decomposition (3.7) with the realization of a Dirac spinor as it appears in the Dirac–Kähler formalism,[9–13] namely, as a differential form that is defined as follows: Let ϕ be a Dirac spinor and Ψ be a 4×4 matrix whose first column is ϕ and whose other matrix elements are 0 (more generally, one may take the columns of Ψ to be artibrary Dirac spinors). It is customary to assume that Ψ is an element of the (in general, complexified) Clifford algebra generated by the γ^α (see Refs. 11–13). Accordingly, Ψ may be expanded as a linear combination of a linearly independent basis of this 16-dimensional Clifford algebra. We shall define a convenient basis in terms of γ_0, the 4×4 unit matrix, and Dirac's γ^{AB} matrices.[3,14]

Let $\gamma^{AB} = -\gamma^{BA}, A, B,... = 1,...,6$, be defined as in Eqs. (12)–(14) of Ref. 3, this representation being given explicitly in Table I. As is well known, $-\frac{1}{2}\gamma^{AB}$ are the generators of a real 4×4 irreducible representation of $\overline{\mathrm{SO}(3,3)}$. Moreover, the $\{\gamma_0, \gamma^{AB}\}$ comprise a linearly independent basis for the Clifford algebra generated by the $\gamma^\alpha \equiv \gamma^{\alpha 6}$.

Clearly, any 4×4 matrix ψ admits a decomposition

$$\Psi = \psi_0\gamma_0 + \tfrac{1}{2}\psi_{AB}\gamma^{AB}, \tag{4.1}$$

TABLE I. The γ^{AB} matrices.

$\gamma^{23} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$	$\gamma^{31} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$	$\gamma^{12} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$
$\gamma^{56} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	$\gamma^{64} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$	$\gamma^{45} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$
$\gamma^{14} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$	$\gamma^{24} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$	$\gamma^{34} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
$\gamma^{15} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	$\gamma^{25} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	$\gamma^{35} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$
$\gamma^{16} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	$\gamma^{26} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$	$\gamma^{36} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$

where the 16 $\{\psi_0,\psi_{AB}\}$ are, in general, complex functions on space-time given by

$$4\psi_0 = \text{tr}\,\Psi \tag{4.2}$$

and

$$4\psi_{AB} = \text{tr}\,\Psi\gamma_{AB}. \tag{4.3}$$

It should be clearly understood that the expansion (4.1) can be formulated for any 4×4 matrix, regardless of its particular transformation properties under $\overline{\text{SO}(3,1)}$ (such as an electromagnetic field tensor, a γ matrix, an energy momentum tensor, a matrix whose columns are Dirac spinors, and so forth). As such, the coefficients ψ_{AB} do not possess, in general, any simple transformation properties under $\overline{\text{SO}(3,1)}$.

Let us consider the case that the columns of the 4×4 matrix Ψ are Dirac spinors. In this case one sometimes writes the expansion (4.1) as[11,12]

$$\Psi = \psi_0\gamma_0 + \psi_\alpha\gamma^\alpha + \tfrac{1}{2}\psi_{\alpha\beta}\gamma^{\alpha\beta} + \phi_\alpha\gamma^\alpha\gamma^5 + \phi_0\gamma^5, \tag{4.4}$$

where $\gamma^\alpha = \gamma^{\alpha 6}$, $\gamma^5 = \gamma^{56}$, $\gamma^\alpha\gamma^5 = \gamma^{\alpha 5}$, $\psi_\alpha = \psi_{\alpha 6}$, $\phi_\alpha = \psi_{\alpha 5}$, and $\phi_0 = \psi_{56}$. Since under the action of $\overline{\text{SO}(3,1)}$, $\Psi \to \Psi' = S\Psi$, $S\in\overline{\text{SO}(3,1)}$, and *not* $\Psi \to S\Psi S^{-1}$, the ψ_α cannot be the components of a four-vector. Nevertheless, some authors associate with Ψ a differential form $\psi_0 + \psi_\alpha\, dx^\alpha + \frac{1}{2}\psi_{\alpha\beta}\, dx^\alpha \wedge dx^\beta + \cdots$ in spite of the fact that $\psi_\alpha\, dx^\alpha$ has no $\overline{\text{SO}(3,1)}$ invariant meaning. Equation (4.4) does not express the equivalence of a Dirac spinor with a set of space-time tensors, and should not be confused with the exceptional equivalence of (3.7) and (3.8).

We conclude this paper by giving a concrete real irreducible representation of the tau matrices. Let

$$\tau^8 = \bar\tau^8 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and $\bar\tau^{A'} = -\tau^{A'}$, for $A' = 1,\ldots,7$. We set

$$\tau^7 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Put $6 = 1'$, $5 = 2'$, and $4 = 3'$. For $h,j,k = 1,2,3$, we define

$$\epsilon^{hkj}\tau^j = \begin{pmatrix} 0 & \gamma^{hk} \\ \gamma^{hk} & 0 \end{pmatrix}$$

and

$$\epsilon^{hkj}\tau^{j'} = \begin{pmatrix} 0 & -\gamma^{h'k'} \\ \gamma^{h'k'} & 0 \end{pmatrix}.$$

With

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

one may readily verify that this concrete representation of the tau matrices satisfies the defining relations (2.14) and (2.16).

[1]P. L. Nash, J. Math. Phys. **25**, 2104 (1984).
[2]R. Brauer and H. Weyl, Am. J. Math **57**, 425 (1935).
[3]P. L. Nash, J. Math. Phys. **25**, 204 (1984).
[4]E. Cartan, *Lecons sur la theorie des spineurs* (Hermann and Cie, Paris, 1938), Vols. I and II.
[5]C. C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U.P., New York, 1954).
[6]A. Gamba, J. Math. Phys. **8**, 775 (1967).
[7]P. Truini and L. C. Biedenharn, Hadronic J. **4**, 995 (1981).
[8]P-E. Hagmark and P. Lounesto, "Cayley–Dickson algebras, Clifford algebras, and Walsh functions," preprint.
[9]E. Kähler, Rend. Mat. (3–4) **21**, 425 (1962).
[10]W. Graf, Ann. Inst. H. Poincaré A **29**, 85 (1978).
[11]J. M. Rabin, Nucl. Phys. D **201**, 315 (1982).
[12]P. Becher and H. Joos, Z. Phys. C **15**, 347 (1982).
[13]N. A. Salingaros and G. P. Wene, Acta Appl. Mat. **4**, 271 (1985).
[14]P. A. M. Dirac, J. Math. Phys. **4**, 901 (1963).

Adapted slicings of space-times possessing simply transitive similarity groups

Robert T. Jantzen
Department of Mathematical Sciences, Villanova University, Villanova, Pennsylvania 19085 and Department of Physics, University of Rome, 00185 Rome, Italy
Kjell Rosquist
Institute of Theoretical Physics, University of Stockholm, Vanadisvägen 9, S-113 46, Stockholm, Sweden

(Received 7 November 1985; accepted for publication 3 January 1986)

The relationship between the hypersurface-homogeneous slicing of an exact power law metric space-time and slicings adapted to spatial self-similarities is discussed in a group theoretical setting.

Exact power law metrics, recently introduced by Wainwright[1] and generalized by Jantzen and Rosquist[2] are four-dimensional space-time metrics possessing a transitive group of homothetic motions[3] or "similarity transformations." These metrics arise naturally in the qualitative analysis of the general relativistic dynamics of spatially homogeneous and spatially self-similar space-times as singular points of a Hamiltonian system of ordinary differential equations for conformally invariant variables.[4] Starting with a spatially homogeneous or spatially self-similar space-time, the existence of a homothetic Killing vector field not tangent to the orbits of the three-dimensional symmetry group leads to a simply transitive similarity group of the space-time. The most familiar class of such space-times is the Kasner solution,[5] an exact solution of the vacuum Einstein equations that plays an important role as an asymptotic solution during certain phases of the evolution of more general spatially homogeneous or spatially self-similar space-times.

The simply transitive case may be treated in exactly the same way as the space-time-homogeneous metrics studied by Ozsváth[6] and Farnsworth and Kerr.[7] The space-time manifold may be identified with the four-dimensional manifold of the symmetry group and the group action with the natural left action of the group on itself by left translation. The isometry group of the space-time is necessarily a three-dimensional subgroup[8] acting simply transitively on its orbits, a family of three-dimensional hypersurfaces that are the right cosets of the isometry subgroup. The space-time is therefore hypersurface homogeneous[9] and admits a preferred slicing, namely by the family of orbits of the isometry subgroup. The object of this paper is to discuss other possible adapted slicings of such space-times, namely by the orbits (right cosets) of nontrivial three-dimensional similarity subgroups, when they exist. These subgroups are related to the original isometry subgroup by a certain family of Lie group deformations. The relationship between these various subgroups provides a group theoretical procedure for finding the hypersurface-homogeneous slicing of the space-time associated with a spatially self-similar exact power law metric.

Let H_4 be the similarity group with Lie algebra h_4 of left-invariant vector fields. Let $\tilde{h}_4$ be the Lie algebra of right-invariant vector fields, the generators of the left action of H_4 on itself by left translation, and let $\{e_\alpha\}$ and $\{\tilde{e}_\alpha\}$ be bases of h_4 and $\tilde{h}_4$, respectively, which agree at the tangent space at the identity of the group. Let the dual spaces h_4^* and $\tilde{h}_4^*$ be identified with the spaces of, respectively, left-invariant and right-invariant one-forms on H_4, and let $\{\omega^\alpha\}$ and $\{\tilde{\omega}^\alpha\}$ be the respective dual bases defined by $\omega^\alpha(e_\beta) = \delta^\alpha{}_\beta = \tilde{\omega}^\alpha(\tilde{e}_\beta)$, where Greek letters run from 1 to 4. Both $\{e_\alpha\}$ and $\{\tilde{e}_\alpha\}$ are global frames on H_4. One has

$$\begin{aligned}
&[e_\alpha, e_\beta] = C^\gamma{}_{\alpha\beta} e_\gamma, \quad [\tilde{e}_\alpha, \tilde{e}_\beta] = -C^\gamma{}_{\alpha\beta}\tilde{e}_\gamma,\\
&[e_\alpha, \tilde{e}_\beta] = 0,\\
&d\omega^\alpha = -\tfrac{1}{2}C^\alpha{}_{\beta\gamma}\omega^\beta \wedge \omega^\gamma,\\
&d\tilde{\omega}^\alpha = \tfrac{1}{2}C^\alpha{}_{\beta\gamma}\tilde{\omega}^\beta \wedge \tilde{\omega}^\gamma.
\end{aligned} \tag{1}$$

The similarity condition on the Lorentz metric g

$$\pounds_{\tilde{e}_\alpha} g = 2f_\alpha g, \tag{2}$$

where f_α are constants, not all of which vanish if H_4 is a nontrivial similarity group as assumed, defines a nonzero right-invariant one-form $f = f_\alpha \tilde{\omega}^\alpha$. From the identity $\pounds_{[X,Y]} = [\pounds_X, \pounds_Y]$, one immediately derives the conditions

$$f_\alpha C^\alpha{}_{\beta\gamma} = 0, \tag{3}$$

which has the following consequences for the one-form f:

$$\begin{aligned}
&df = \tfrac{1}{2} f_\alpha C^\alpha{}_{\beta\gamma}\tilde{\omega}^\beta \wedge \tilde{\omega}^\gamma = 0,\\
&\pounds_{\tilde{e}_\alpha} f = f_\beta C^\beta{}_{\alpha\gamma}\tilde{\omega}^\gamma = 0.
\end{aligned} \tag{4}$$

The first relation shows f to be a closed one-form and, in the case in which H_4 is simply connected, exact as well. The second relation shows f to be invariant under the coadjoint representation of $\tilde{h}_4$ on $\tilde{h}_4^*$ and since $\tilde{h}_4$ generates the left translations, f is left invariant and hence bi-invariant. Since e_α and $\tilde{e}_\alpha$ coincide at the identity of the group, f has the same components in either basis

$$\begin{aligned}
f &= f_\alpha \tilde{\omega}^\alpha = f_\alpha \omega^\alpha\\
&\in h_4^* \cap \tilde{h}_4^* = \{\gamma_\alpha \omega^\alpha \mid \gamma_\alpha C^\alpha{}_{\beta\gamma} = 0,\ (\gamma_\alpha) \in \mathbf{R}^4\}.
\end{aligned} \tag{5}$$

One can always choose the basis so that $f_\alpha = \delta^4{}_\alpha$, in which case $C^4{}_{\beta\gamma} = 0$ and the only nonzero structure constant tensor components are of the form $C^a{}_{bc}$ and $C^a{}_{4b}$, where Latin indices run from 1 to 3, leading to the relations

$$\pounds_{\tilde{e}_a} g = 0, \quad \pounds_{\tilde{e}_4} g = 2g, \tag{6}$$

and

$$[\tilde{e}_a, \tilde{e}_b] = -C^c{}_{ab} \tilde{e}_c, \tag{7a}$$

$$[\tilde{e}_4, \tilde{e}_a] = -C^b{}_{4a} \tilde{e}_b. \tag{7b}$$

Thus the $\{\tilde{e}_a\}$ span an isometry Lie subalgebra to which the vector fields $[\tilde{e}_4, \tilde{e}_a]$ belong, while bracketing by $\tilde{e}_4$, already an inner derivation of the Lie algebra $\tilde{g}_4$, becomes a derivation of the Lie subalgebra $\tilde{h}_3 = \text{span}\{\tilde{e}_a\}$; let G_3 be the subgroup of H_4 that corresponds to this Lie subalgebra. One can add any linear combination of the vector fields $\{\tilde{e}_a\}$ to $\tilde{e}_4$ without changing (6), while (7b) changes by the addition of an arbitrary inner derivation of the Lie subalgebra $\tilde{g}_3$,

$$[\tilde{e}_4 + y^c \tilde{e}_c, \tilde{e}_a] = -(C^b{}_{4a} + y^c C^b{}_{ca}) \tilde{e}_b. \tag{8}$$

The essential part of the matrix $(C^b{}_{4a})$ is therefore equivalent to an outer derivation of $\tilde{g}_3$. Recall that the space of outer derivations of a Lie algebra is the vector space quotient of the Lie algebra of derivations of that algebra by the Lie subalgebra of inner derivations of that algebra (adjoint transformations).

If we assume H_4 is simply connected, then the closed bi-invariant one-form $f = \tilde{\omega}^4$ is exact and therefore $\tilde{\omega}^4 = d\zeta$ (which implies $\tilde{e}_4 \zeta = e_4 \zeta = 1$ by duality). The integral submanifolds of $\tilde{\omega}^4$, namely the hypersurfaces $\zeta = \zeta_{(0)}$, are exactly the orbits of the action of the isometry subgroup G_3, namely the right cosets of G_3 in H_4. Note that $\tilde{\omega}^4$ is also invariant under the basis transformations $\tilde{e}_4 \to \tilde{e}_4 + y^a \tilde{e}_a$, which leave (6) invariant.

Except for a few special cases discussed by Eardley,[8] a space-time metric g with a similarity group is always conformally related to a metric that is invariant under the similarity group

$$g = e^{2\psi} g_{(0)}, \quad \pounds_{\tilde{e}_a} g_{(0)} = 0. \tag{9}$$

From (6) one has $\tilde{e}_4 \psi = 1$, $\tilde{e}_a \psi = 0$, so one may assume $\psi = \zeta$. The metric $g_{(0)}$ is space-time homogeneous and is a left-invariant metric on H_4, which therefore may be expressed in the left-invariant frame $\{e_\alpha\}$ with constant components

$$g_{(0)} = g_{(0)\alpha\beta}\, \omega^\alpha \otimes \omega^\beta, \quad d(g_{(0)\alpha\beta}) = 0. \tag{10}$$

Suppose one considers the change of basis of $\tilde{h}_4$,

$$\tilde{\xi}_a = \tilde{e}_a + b_a \tilde{e}_4, \quad \tilde{\sigma}^a = \tilde{\omega}^a, \tag{11a}$$

$$\tilde{\xi}_4 = \tilde{e}_4, \quad \tilde{\sigma}^4 = \tilde{\omega}^4 - b_a \tilde{\omega}^a, \tag{11b}$$

where the b_a are constants satisfying

$$b_c C^c{}_{ab} = 0, \tag{12a}$$

$$b_c C^c{}_{4[a} b_{b]} = 0. \tag{12b}$$

A simple computation shows that

$$\begin{aligned} &[\tilde{\xi}_a, \tilde{\xi}_b] = -(C^c{}_{ab} - C^c{}_{4[a} b_{b]}) \tilde{\xi}_c, \\ &[\tilde{\xi}_4, \tilde{\xi}_a] = -C^b{}_{4a} \tilde{\xi}_b, \\ &d\tilde{\sigma}^4 = -b_a C^a{}_{4b} \tilde{\sigma}^4 \wedge \tilde{\sigma}^b, \quad \tilde{\sigma}^4 \wedge d\tilde{\sigma}^4 = 0, \\ &\pounds_{\tilde{\xi}_a} g = 2 b_a g. \end{aligned} \tag{13}$$

Thus the $\{\tilde{\xi}_a\}$ generate a three-dimensional similarity subgroup H_3 whose three-dimensional orbits (right cosets) are the integral submanifolds of the one-form $\tilde{\sigma}_4$. This new slicing of the space-time is adapted to a "hypersurface self-similarity" of the metric, the natural generalization of Eardley's term "spatial self-similarity[8]" to the case where the causal nature of the hypersurface is arbitrary.

The space

$$g_3^* \cap \tilde{g}_3^* = \{\gamma_a \tilde{\omega}^a | \gamma_c C^c{}_{ab} = 0, \ (\gamma_a) \in \mathbb{R}^3\} \tag{14}$$

(when its elements are restricted to the subgroup G_3) is the space of closed or equivalently bi-invariant one-forms on the subgroup G_3, or equivalently, the subspace invariant under the co-adjoint representation of G_3 on the dual space to $\tilde{g}_3$ or g_3. Since $\tilde{e}_4$ acts as a derivation of the Lie subalgebra $\tilde{g}_3$, with matrix $(-C^a{}_{4b})$, it maps the space $g_3^* \cap \tilde{g}_3^*$ into itself, easily verified by an application of the Jacobi identity. The condition (12b) is equivalent to the requirement that (b_a) be an eigenvector of the matrix $(C^a{}_{4b})$:

$$b_b C^b{}_{4a} = \theta b_a. \tag{15}$$

This guarantees that $b_a \tilde{\omega}^a$ remain bi-invariant under the Lie algebra deformation (11a) of $\tilde{g}_3$ into $\tilde{h}_3$. When the eigenvalue θ is zero, then $\tilde{\sigma}^4$ is also bi-invariant on H_4, and, assuming simple connectivity, also exact.

The classification of four-dimensional similarity groups H_4 for a fixed isometry subgroup G_3 is equivalent to a description of the quotient space of the space of outer derivations of $\tilde{g}_3$ by the natural action of the automorphism group of $\tilde{g}_3$. In more explicit terms this classification is just a description of the equivalence classes of derivation matrices $(C^a{}_{4b})$ under the combined action of the group of matrices that leave invariant the structure constant tensor components $C^a{}_{bc}$ and the addition of adjoint matrices associated with these structure constant tensor matrices

$$\begin{aligned} &C^a{}_{4b} \to A^a{}_c A^{-1d}{}_b (C^c{}_{4d} + y^e C^c{}_{ed}), \\ &A^a{}_d C^d{}_{fg} A^{-1f}{}_b A^{-1g}{}_c = C^a{}_{bc}. \end{aligned} \tag{16}$$

The isometry subgroups G_3 may be classified according to the Bianchi–Behr classification.[10] Only nonsemisimple groups G_3 (of Bianchi types I–VII) admit nontrivial spaces $g_3^* \cap \tilde{g}_3^*$; similarly only these types of groups may act as nontrivial self-similarity groups H_3. Suppose one has a spatially self-similar exact power law metric space-time with spatial self-similarity group H_3 and simply transitive similarity group H_4. Knowing the Lie algebra structure of H_4 one can work backward from (13) to find the Killing vector fields and hence the slicing by homogeneous hypersurfaces. The inverse of the transformation (11) then describes the relationship between the original spatial self-similarity group and the isometry group, which are connected by a family of self-similarity subgroups. Any one of these subgroups of H_4 may be used to slice the space-time; clearly the isometry group is the preferred member of this family of subgroups.

To make this discussion more concrete, it is worth examining an explicit example. Consider a spatially self-similar exact power law metric expressed in coordinates adapted to the orbits of a Bianchi type VI_h subgroup H_3 of the full isometry group $H_4 \sim \mathbb{R}^4$ on which $\{x^1, x^2, x^3, \lambda\}$ are taken to be global coordinates.[11] Using the logarithmic time variable $\lambda = \ln t$, where t is the usual cosmic time function, and the symbol $\mathbf{e}^a{}_b$ for the 3×3 matrix whose only nonzero entry is a 1 in the bth row and ath column, this metric has the form

$$^4g = -\Omega^4 \otimes \Omega^4 + \delta_{ab}\Omega^a \otimes \Omega^b ,$$
$$\Omega^4 = e^{bx^3}\, dt = e^{\lambda + bx^3}\, d\lambda ,$$
$$(\Omega^a) = e^{(\lambda + bx^3)\mathbf{1} + \boldsymbol{\beta}_{(0)}}(\sigma^a) , \qquad (17)$$
$$(\sigma^a) = [\exp\{[(s-1)\lambda - ax^3]\mathbf{I}^{(3)} + (q\lambda - x^3)\mathbf{k}_3^0\} + M\mathbf{e}^3{}_2 + N\mathbf{e}^3{}_1] \begin{pmatrix} dx^1 \\ dx^2 \\ dx^3 \end{pmatrix},$$

where

$$\mathbf{I}^{(3)} = \mathrm{diag}(1,1,0), \quad \boldsymbol{\beta}_{(0)} = \mathrm{diag}(\beta^1{}_{(0)}, \beta^2{}_{(0)}, \beta^3{}_{(0)}) ,$$
$$\mathbf{k}_3^0 = -q_0(\mathbf{e}^1{}_2 + \mathbf{e}^2{}_1) . \qquad (18)$$

The quantities $\{\beta^a{}_{(0)}, s, q, M, N; a, q_0; b\}$ are constants and $h = -a^2 q_0^{-2}$ is the group parameter specifying the Bianchi type of H_3, while $\{\sigma^a, d\lambda\}$ are a basis of left-invariant one-forms on H_4.

The left action of the subgroup H_3 on H_4 is generated by the right-invariant vector fields

$$\{\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3\} = \{\partial_1, \partial_2, \partial_3 + a(x^1\partial_1 + x^2\partial_2) - q_0(x^2\partial_1 + x^1\partial_2)\} ,$$

which satisfy the final relation of (13) with $b_a = b\delta^3{}_a$. Note that $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are Killing vector fields and, when $b \neq 0$, $\tilde{\xi}_3$ is a homothetic Killing vector field. The remaining linearly independent right-invariant vector field is associated with the transformation

$$(x^1, x^2, x^3, \lambda) \to (x^1 e^{(aq - (s-1))\zeta}, x^2 e^{(aq - (s-1))\zeta}, x^3 + q\zeta, \lambda + \zeta) . \qquad (19)$$

under which the metric scales by the constant factor $e^{2(1+bq)\zeta}$. When $1 + bq = 0$, this is an isometry subgroup generated by the vector field

$$\tilde{\xi}_4 = \partial_\lambda + q\,\partial_3 + [aq - (s-1)](x^1\partial_1 + x^2\partial_2) , \qquad (20)$$

and $\{\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_4\}$ span the Killing Lie algebra (a Lie algebra of Bianchi type V, see below) and are tangent to the slicing of the space-time by the orbits of the isometry group. When $1 + bq \neq 0$, correcting for the scale factor so that the generator $\tilde{\xi}_4$ of this one-parameter similarity subgroup satisfies (6), one has instead

$$\tilde{\xi}_4 = (1+bq)^{-1}\{\partial_\lambda + q\,\partial_3 + [aq - (s-1)](x^1\partial_1 + x^2\partial_2)\} , \qquad (21)$$

and consequently

$$\tilde{\sigma}^4 = (1+bq)d\lambda = \sigma^4 . \qquad (22)$$

The Lie brackets of the right invariant basis vector fields are

$$\left[\tilde{\xi}_4, \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \\ \tilde{\xi}_3 \end{pmatrix}\right] = -(1+bq)^{-1}[aq - (s-1)]\mathbf{I}^{(3)} \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \\ \tilde{\xi}_3 \end{pmatrix},$$
$$[\tilde{\xi}_2, \tilde{\xi}_3] = -q_0\tilde{\xi}_1 - a\tilde{\xi}_2 , \qquad (23)$$
$$[\tilde{\xi}_3, \tilde{\xi}_1] = a\tilde{\xi}_1 - q_0\tilde{\xi}_2 ,$$
$$[\tilde{\xi}_1, \tilde{\xi}_2] = 0 .$$

These completely determine the Lie group structure of the simply connected group H_4, once the identity of the group is specified as the point that is the origin of coordinates. The one-forms $\{\sigma^a\}$ defined by (17) and (22) are the left-invariant one-forms dual to the left-invariant basis $\{\xi_\alpha\}$ associated with $\{\tilde{\xi}_\alpha\}$. In terms of them the metric can assume the form (9), (10) with $\psi = \lambda + bx^3$, $(g_{(0)ab}) = e^{2\beta_{(0)}}$ and $g_{(0)4\alpha} = -\delta_{4\alpha}$. Expressed in language more familiar in general relativistic discussions, the one-forms $\{\sigma^\alpha\}$ are invariant under dragging along by the vector field $\tilde{\xi}_4$, which generates the slicing of the space-time from the initial hypersurface $H_3 \subset H_4$ by dragging along.

The derivation matrix $(C^a{}_{4b}) = (1+bq)^{-1} \times [aq - (s-1)]\mathbf{I}^{(3)}$ is invariant under the matrix automorphism group of the Lie algebra $\tilde{h}_3$ of H_3. For all values of h except $h = -1$, which corresponds to the special Bianchi type III, one has

$$h_3^* \cap \tilde{h}_3^* = \{\gamma_a\omega^a = B\,dx^3 | (\gamma_a) = (0,0,B),\ B \in \mathbf{R}\} \qquad (24)$$

since $\omega^3 = \tilde{\omega}^3 = dx^3$. Here $\gamma_a C^a{}_{4b}$ is identically zero for these one-forms, which are therefore also bi-invariant as one-forms on H_4.

The transformation inverse to (11) with b_a replaced by γ_a leads to another exact bi-invariant one-form on H_4:

$$\tilde{\omega}^4 = \tilde{\sigma}^4 + B\tilde{\sigma}^3 = d(\lambda + Bx^3) = d\bar{\lambda} ,$$
$$\bar{\lambda} = \lambda + Bx^3 . \qquad (25)$$

Expressing the metric in terms of $\bar{\lambda}$ leads to a form adapted to the action of the new homothetic subgroup $\bar{H}_3$ of H_4:

$$\Omega^4 = e^{\bar{\lambda} + \bar{b}x^3}(d\bar{\lambda} - B\,dx^3) ,$$
$$(\Omega^a) = e^{(\bar{\lambda} + \bar{b}x^3)\mathbf{1} + \beta_{(0)}}(\exp\{[(s-1)\bar{\lambda} - \bar{a}x^3]\mathbf{I}^{(3)} + (\bar{q}\bar{\lambda} - x^3)\bar{\mathbf{k}}_3^0\} + M\mathbf{e}^3{}_2 + N\mathbf{e}^3{}_1) \begin{pmatrix} dx^1 \\ dx^2 \\ dx^3 \end{pmatrix},$$
$$\bar{b} = b - B, \quad \bar{a} = a - B(s-1) ,$$
$$\bar{q}_0 = (1+qB)q_0, \quad \bar{q} = q(1+qB)^{-1} ,$$
$$\bar{h} = -\bar{a}^2\bar{q}_0^2 . \qquad (26)$$

By choosing $\bar{b} = 0$, one obtains the slicing by the isometry subgroup, which may be a group of Bianchi type $\mathrm{VI}_{\bar{h}}$ ($\bar{h} \neq 0, -\infty$), VI_0 ($\bar{q}_0 \neq 0$, $\bar{a} = 0$), V ($\bar{q}_0 = 0, \bar{a} \neq 0$), or I ($\bar{q}_0 = 0 = \bar{a}$).

For Bianchi type $\mathrm{VI}_{-1} = \mathrm{III}$, setting $a = 1 = q_0$, one has instead

$$h_3^* \cap \tilde{h}_3^* = \{\gamma_a\omega^a | (\gamma_a) = (\mathscr{B}, \mathscr{B}, B) \in \mathbf{R}^3\} \qquad (27)$$

and the automorphism group may be used to reduce (γ_a) to the form (1,1,0). In this case a second linearly independent eigenvector of $(C^b{}_{4b})$ exists, namely (1,1,0) with eigenvalue $\theta = (1+bq)^{-1}(q - s + 1)$, corresponding to the one-form $\sigma^1 + \sigma^2 = e^{-(q-s+1)\lambda}\, d(x^1 + x^2)$, which is bi-invariant when restricted to the subgroup H_3, where $\lambda = 0$. One therefore has the option of considering the transformation inverse to (11) with b_a replaced by $(\mathscr{B}, \mathscr{B}, 0)$, leading to an example in which $\tilde{\omega}^4$ is not bi-invariant when $\theta \neq 0$.

ACKNOWLEDGMENTS

R. J. gratefully acknowledges the hospitality of the Institute of Theoretical Physics at Stockholm and the Vatican Observatory at Castel Gandolfo.

K. R. was supported by the Swedish Natural Research Council.

[1]J. Wainwright, Gen Relativ. Gravit. **16**, 657 (1984).
[2]R. T. Jantzen and K. Rosquist, "Exact power law metrics in cosmology," Class. Quantum Gravit., to appear.
[3]K. Rosquist and R. T. Jantzen, Class. Quantum Gravit. **2**, L129 (1985).
[4]R. Rosquist, Class. Quantum Gravit. **1**, 81 (1984).
[5]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[6]I. Ozsváth, J. Math. Phys. **6**, 590 (1965).
[7]D. L. Farnsworth and R. P. Kerr, J. Math. Phys. **7**, 1625 (1966).
[8]D. M. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[9]D. Kramer, H. Stephani, M. MacCallum, and H. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980).
[10]F. B. Estabrook, H. D. Wahlquist, and C. G. Behr, J. Math. Phys. **9**, 497 (1968).
[11]K. Rosquist and R. T. Jantzen, Phys. Lett. A **107**, 29 (1985).

The most degenerate representation matrix elements of finite rotations of SO($n-2$, 2)

Ansaruddin Syed
Department of Mathematics, University of Karachi, Karachi, Pakistan

(Received 20 May 1985; accepted for publication 20 November 1985)

Using a technique of Strom and Boyer [S. Strom, Ark. Fys. **33**, 465 (1966) and C. P. Boyer, J. Math. Phys. **12**, 1599 (1971)], the matrix elements of finite rotations of the group SO($n-2$, 2) have been computed in the most degenerate principal series of continuous representations.

I. INTRODUCTION

The problem of calculating unitary irreducible representation (UIR) matrix elements of finite rotations of compact and noncompact rotation groups has attracted considerable attention during the past several years, and a number of papers[1–18] have already appeared on the subject. However, most of these papers concern either the compact groups SO(n) or the noncompact groups of the Lorentz type SO($n-1$, 1) only; the more general cases SO(p, q), $p, q \geqslant 2$, usually have not been considered in this context. In a previous paper,[19] the author therefore focused his attention on the simplest of these groups, viz. SO(2, 2), and obtained the matrix elements of its finite rotations in a general UIR. It was possible to do this in a relatively simple manner by using its isomorphism with the direct product SO(2, 1) $\otimes$ SO(2, 1), and a trick of Friedman and Wang.[8] As no such isomorphism obviously exists for the higher groups SO($n-2$, 2), $n \geqslant 5$, other means have to be employed for them. Although the general case of an arbitrary UIR of SO($n-2$, 2) appears somewhat intractable at the moment, a technique of Strom[20] and Boyer[21] for the special class of "most degenerate" representations [which they use for SO(3, 1) and SO(n, 1), respectively] is found to be applicable in the case of these groups also. We prove this in the present paper by explicitly calculating, by this method, the matrix elements of finite rotations of SO($n-2$, 2) in the most degenerate principal series of continuous representations. Let us recall here that by a "most degenerate" representation, we mean one that is labeled by just *one* index, i.e., all the Casimir operators of the group vanish in such a representation except one that is just the Laplace–Beltrami operator.[22] We start by obtaining a particular realization of a set of representations of SO($n-2$, 2), which we identify with the most degenerate principal series of continuous representations as described by Limic, Niederle, and Raczka.[23] The action of operators of these representations on the general element of the function space (on which these operators operate) is then explicitly obtained. It is only then that the actual calculation of the required matrix element is carried out.

II. THE REPRESENTATIONS T^σ OF SO($n-2$, 2)

Consider the Minkowski space $M_{n-2\,2}$ spanned by the points

$$x \equiv (x_1, x_2, \ldots, x_n) \equiv (x_i)$$

(Latin indices i, j, k, etc. take on the values from 1 to n) and having metric

$$r = (x,x)^{1/2} \equiv (x_1^2 + x_2^2 + \cdots + x_{n-2}^2 - x_{n-1}^2 - x_n^2)^{1/2}.$$

Let B^σ be the space of functions $f(x)$ given on the cone

$$C_{n-2\,2}\colon\ r = 0,$$

in $M_{n-2\,2}$, and such that (i) $f(x)$ is infinitely differentiable at every point of the cone; (ii) $f(x)$ is homogeneous in x, of degree σ, i.e.,

$$f(ax) = a^\sigma f(x), \quad \text{for } a > 0;$$

and (iii) $f(x)$ is even in x, i.e.,

$$f(-x) = f(x).$$

Now, if f is an element of B^σ and $g \in \mathrm{SO}(n-2, 2)$, i.e., g is an orthogonal transformation in $M_{n-2\,2}$, with $\det g = 1$, then it is immediate that f_g also belongs to B^σ, where

$$f_g(x) = f(g^{-1}x). \tag{1}$$

Hence

$$(S^\sigma(g))f = f_g \tag{2}$$

defines a (linear) operator

$$S^\sigma(g)\colon\ B^\sigma \to B^\sigma$$

on B^σ. It is easily verified that the set

$$\{S^\sigma(g),\ g \in \mathrm{SO}(n-2,2)\}$$

forms a representation of SO($n-2$, 2) in the space B^σ.

Consider now the ($n-2$)-dimensional hyperboloid $S_{n-3\,1}$ (of one sheet) spanned by the ($n-1$)-tuples of real numbers

$$(\xi_1, \xi_2, \ldots, \xi_{n-1}) \equiv (\xi_\alpha)$$

(Greek indices α, β, γ, etc. take on the values from 1 to $n-1$), satisfying

$$\xi_1^2 + \xi_2^2 + \cdots + \xi_{n-2}^2 - \xi_{n-1}^2 = 1. \tag{3}$$

For the sake of convenience, we shall denote the points of $S_{n-3\,1}$ by primed Greek letters such as ξ', η', ζ', etc.; thus

$$\xi' \equiv (\xi_\alpha) \equiv (\xi_1, \xi_2, \ldots, \xi_{n-1}),$$

where the ξ_α satisfy (3). Let $D^\infty(S_{n-3\,1})$ be the set of infinitely differentiable functions on $S_{n-3\,1}$.

We now associate with each function f of B^σ, a function $F \equiv Qf$ of $D^\infty(S_{n-3\,1})$ by the rule [note that $(\xi_\alpha) \in S_{n-3\,1} \Rightarrow (\xi_\alpha, 1) \in C_{n-2\,2}$]

$$F(\xi') \equiv F(\xi_1, \xi_2, \ldots, \xi_{n-1}) = f(\xi_1, \xi_2, \ldots, \xi_{n-1}, 1). \tag{4}$$

In other words, Q is a mapping

$$Q\colon\ B^\sigma \to D^\infty(S_{n-3\,1})$$

 0022-2488/86/051195-11$02.50

defined by (4) with $F = Qf$.

Now homogeneity of $f \in B^\sigma$ implies that, for $x_n > 0$,

$$f(x_1, x_2, \ldots, x_n) = x_n^\sigma f\left(\frac{x_1}{x_n}, \frac{x_2}{x_n}, \ldots, \frac{x_{n-1}}{x_n}, 1\right) = x_n^\sigma F\left(\frac{x_1}{x_n}, \frac{x_2}{x_n}, \ldots, \frac{x_{n-1}}{x_n}\right), \quad (5a)$$

where $F = Qf$, as

$$(x_1, x_2, \ldots, x_n) \in C_{n-22} \Rightarrow \left(\frac{x_1}{x_n}, \frac{x_2}{x_n}, \ldots, \frac{x_{n-1}}{x_n}\right) \in S_{n-31}.$$

Similarly, for $x_n < 0$,

$$f(x_1, x_2, \ldots, x_n) = (-x_n)^\sigma f\left(-\frac{x_1}{x_n}, -\frac{x_2}{x_n}, \ldots, -\frac{x_{n-1}}{x_n}, -1\right) = (-x_n)^\sigma f\left(\frac{x_1}{x_n}, \frac{x_2}{x_n}, \ldots, \frac{x_{n-1}}{x_n}, 1\right) = (-x_n)^\sigma F\left(\frac{x_1}{x_n}, \frac{x_2}{x_n}, \ldots, \frac{x_{n-1}}{x_n}\right), \quad (5b)$$

where we again have $F = Qf$, and evenness of the function $f(x)$ has been used.

Thus each $F \in D^\infty(S_{n-31})$ uniquely determines, according to the rules (5a) and (5b), the $f \in B^\sigma$ such that $F = Qf$; we express this by saying that $f = Q^{-1}F$. Note that we encounter some difficulty in the rules (5a) and (5b) when $x_n = 0$; however, this can be overcome, as shown by Bander and Iztykson[24] and mentioned by Boyer and Ardalan,[25] by adding to S_{n-31} extra points at infinity, i.e., by compactifying S_{n-31} by adjunction of a surface at infinity.

We thus see that there exists a one to one correspondence between B^σ and $D^\infty(S_{n-31})$, and therefore, operators of the representation $S^\sigma(g)$ lead to operators of the representation

$$T^\sigma(g) = Q(S^\sigma(g))Q^{-1} \quad (6)$$

in the function space $D^\infty(S_{n-31})$.

III. RELATIONSHIP OF T^σ TO THE REPRESENTATIONS $C^{\Lambda+}_{n-22}$ OF LNR

The best way to find the values of σ for which the representations T^σ are unitary and irreducible is by relating them to the representations $C^{\Lambda+}_{n-22}$ of Limic, Niederle, and Raczka[23] (we denote them by LNR), which we now do. Let $\Box$ be the Laplace operator

$$\Box = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_{n-2}^2} - \frac{\partial^2}{\partial x_{n-1}^2} - \frac{\partial^2}{\partial x_n^2}$$

in the space M_{n-22}. If we consider $\Box$ inside the cone C_{n-22}, i.e., for

$$r \equiv (x,x)^{1/2} > 0,$$

and introduce some polar coordinates

$$(r, \theta_1, \theta_2, \ldots, \theta_{n-1})$$

in M_{n-22} such that

$$x_i = r\,\Theta_i(\theta_1, \theta_2, \ldots, \theta_{n-1}), \quad i = 1, 2, \ldots, n,$$

then $\Box$ can be written as

$$\Box = \frac{1}{r^{n-1}}\frac{\partial}{\partial r}\left(r^{n-1}\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\Box_0,$$

where $\Box_0$, the Laplace operator over the hyperboloid $r = 1$, is a differential operator in the angular variables $\{\theta_\alpha\}$. (This can be easily seen by writing $\Box$ = div grad and then using for divergence and gradient, the expressions in orthogonal curvilinear coordinates in M_{n-22}, which are immediate generalizations of the corresponding expressions in $\mathbf{R}^3$.) Suppose now that $f(x)$ is an even homogeneous function of $x \in M_{n-22}$, of degree σ, and satisfies $\Box f = 0$. Then, by homogeneity,

$$f(x) = f(r\cdot(x/r)) = r^\sigma f(x/r) = r^\sigma \hat{f}(\theta_\alpha)$$

for some function $\hat{f}$ of the angular variables $\{\theta_\alpha\}$. Hence

$$0 = \Box f = \hat{f}\cdot\frac{1}{r^{n-1}}\frac{d}{dr}\left(r^{n-1}\frac{d}{dr}r^\sigma\right) + \frac{r^\sigma}{r^2}\Box_0\hat{f}$$

$$\Rightarrow \hat{f}\cdot\sigma(n+\sigma-2)r^{\sigma-2} + r^{\sigma-2}\Box_0\hat{f} = 0.$$

$$\Rightarrow \Box_0\hat{f} = -\sigma(n+\sigma-2)\hat{f}.$$

This proves the following statement.

Statement (A): If $f(x)$ is an even homogeneous function of degree σ and satisfies $\Box f = 0$, then $f(x/r)$ is an even eigenfunction of $\Box_0$ corresponding to the eigenvalue $-\sigma(n+\sigma-2)$.

Next, it can be checked that $\Box_0$ commutes with the elements of $SO(n-2, 2)$, i.e., if ξ belongs to the hyperboloid $r = 1$ and

$$L(g)F(\xi) = F(g^{-1}\xi),$$

then

$$\Box_0 L(g) = L(g)\Box_0.$$

As $L(g)$ obviously transforms even functions into even functions, we get the following statement.

Statement (B): An even eigenfunction of $\Box_0$ corresponding to a given eigenvalue is transformed by $L(g)$ into an even eigenfunction corresponding to the same eigenvalue.

Now, just as in the case of $SO(n)$ (see Ref. 26) and $SO(n-1, 1)$ (see Ref. 27), the representations T^σ of $SO(n-2, 2)$ are equivalent to the representations Q^σ obtained as follows: First, let $\mathscr{P}^\sigma$ be the set of even homogeneous functions $f(x)$ of degree σ such that $\Box f = 0$. Next, we introduce a bit of notation: we use unprimed Greek letters ξ, η, ζ, etc. to denote points of M_{n-22} that lie on the hyperboloid $r = 1$; thus

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n)$$

means that

$$\xi_1^2 + \xi_2^2 + \cdots + \xi_{n-2}^2 - \xi_{n-1}^2 - \xi_n^2 = 1.$$

And now, we suppose that $\mathscr{H}^\sigma$ is the set of even functions $f(\xi)$ on this hyperboloid such that

$$r^\sigma f(x/r) \in \mathscr{P}^\sigma.$$

Then, by the statement (A) above, $\mathscr{H}^\sigma$ is simply the set of even eigenfunctions of $\Box_0$ corresponding to the eigenvalue

$-\sigma(n+\sigma-2)$, and by (B) above, it is invariant under $SO(n-2,2)$. The representations Q^σ are then just the restriction of $L(g)$ to $\mathscr{H}^\sigma$.

If we now look at the paper[23] of LNR, we see that the definition of our $\mathscr{H}^\sigma$ makes it identical with their space $\mathscr{H}^{\Lambda+}_{n-22}$ of the representations $C^{\Lambda+}_{n-22}$. Hence our representations Q^σ are the same as their representations $C^{\Lambda+}_{n-22}$. The permissible set of values of σ is therefore obtained by expressing it in terms of Λ (whose range is given in Ref. 23); this is best done by equating the expressions for the eigenvalues of $\Box_0$ in terms of σ and Λ, which gives

$$\begin{aligned}-\sigma(n+\sigma-2) &= \Lambda^2+((n-2)/2)^2\\ &= ((n-2)/2+i\Lambda)((n-2)/2-i\Lambda)\\ &= (i\Lambda-(n-2)/2)(i\Lambda-(n-2)/2+n-2)\\ \Rightarrow \sigma &= i\Lambda-(n-2)/2.\end{aligned}$$

As Λ ranges[23] between 0 and ∞, we see that the representations T^σ are unitary and irreducible for

$$\sigma = i\rho-(n-2)/2,\quad \rho>0. \tag{7}$$

IV. EXPLICIT FORM OF THE OPERATORS T^σ

Let us now find the effect of the operator T^σ on F, i.e., if

$$(T^\sigma(g))F=F_g,$$

we wish to find out $F_g(\xi')$ in terms of $F(\xi')$ [recall that

$$\xi'\equiv(\xi_1,\xi_2,\ldots,\xi_{n-1})$$

is a general point of S_{n-31}]. We have

$$\begin{aligned}F_g &= (T^\sigma(g))F = QS^\sigma(g)Q^{-1}F\\ &= Q(S^\sigma(g))f \quad (f=Q^{-1}F)\\ &= Qf_g,\\ \Rightarrow F_g(\xi_\alpha) &= (Qf_g)(\xi_\alpha)=f_g(\xi_\alpha,1)\\ &= f(g^{-1}(\xi_\alpha,1))\\ &= f((g^{-1})_{\alpha\beta}\xi_\beta+(g^{-1})_{\alpha n},(g^{-1})_{n\beta}\xi_\beta\\ &\quad +(g^{-1})_{nn})\\ &= |(g^{-1})_{n\beta}\xi_\beta+(g^{-1})_{nn}|^\sigma\\ &\quad \times f\left(\frac{(g^{-1})_{\alpha\beta}\xi_\beta+(g^{-1})_{\alpha n}}{(g^{-1})_{n\beta}\xi_\beta+(g^{-1})_{nn}},1\right)\\ \Rightarrow ((T^\sigma(g))F)(\xi_\alpha) &\equiv F_g(\xi_\alpha)\\ &= |(g^{-1})_{n\beta}\xi_\beta+(g^{-1})_{nn}|^\sigma\\ &\quad \times F\left(\frac{(g^{-1})_{\alpha\beta}\xi_\beta+(g^{-1})_{\alpha n}}{(g^{-1})_{n\beta}\xi_\beta+(g^{-1})_{nn}}\right).\end{aligned} \tag{8}$$

If $g=h$, an element of the subgroup $SO(n-2,1)$ of $SO(n-2,2)$, which keeps x_n invariant, then

$$\begin{aligned}((T^\sigma(h))F)(\xi_\alpha) &\equiv ((T^\sigma(h))F)(\xi')\\ &= F((h^{-1})_{\alpha\beta}\xi_\beta)=F(h^{-1}\xi'),\end{aligned}$$

i.e., $T^\sigma(h)$ becomes the quasiregular representation of $SO(n-2,1)$. As its decomposition into irreducible components is known[28] and the matrix elements of h in these components have already been calculated,[21] it is sufficient to consider only those $g\in SO(n-2,2)$ that do not keep x_n invariant. Now it is known[29] that every such element has one of the following two representations: (i) if $g_{nn}<1$ then

$$g=h_1r_{n-1\,n}(\alpha)h_2,\quad h_1,h_2\in SO(n-2,1),$$

where $r_{n-1\,n}(\alpha)$ is the rotation by an angle α in the $((n-1),n)$ plane of M_{n-22}; and (ii) if $g_{nn}>1$ then

$$g=h_1\,l_{n-2\,n}(\alpha)h_2,\quad h_1,h_2\in SO(n-2,1),$$

where $l_{n-2\,n}(\alpha)$ is the usual Lorentz transformation by a (hyperbolic) angle α in the $((n-2),n)$ plane of M_{n-22}. However, it is easy to verify by the direct multiplication of matrices, that

$$l_{n-2,n}(\alpha)=r_{n-1\,n}(-\pi/2)l_{n-2\,n-1}(\alpha)r_{n-1\,n}(\pi/2),$$

so that in the case (ii), g is expressible in the form

$$\begin{aligned}&g=h_1r_{n-1\,n}(-\pi/2)h_0r_{n-1\,n}(\pi/2)h_2,\\ &\quad h_0,h_1,h_2\in SO(n-2,1).\end{aligned}$$

It therefore follows that we need only to find the effect of $T^\sigma(T_{n-1\,n}(\alpha))$ on functions F in order to be able to obtain the effect of $T^\sigma(g)$ on them for arbitrary $g\in SO(n-2,2)$ and hence also the matrix elements of $T^\sigma(g)$ between various basis elements of $D^\infty(S_{n-31})$.

Now with $g=r_{n-1\,n}(\alpha)$, (8) gives

$$\begin{aligned}&T^\sigma(r_{n-1\,n}(\alpha))F(\xi_1,\xi_2,\ldots,\xi_{n-1})\\ &\quad =|A|^\sigma F\left(\frac{\xi_1}{A},\frac{\xi_2}{A},\ldots,\frac{\xi_{n-2}}{A},\frac{\sin\alpha-\xi_{n-1}\cos\alpha}{A}\right),\end{aligned} \tag{9}$$

where

$$A=\cos\alpha-\xi_{n-1}\sin\alpha.$$

Next, we transform, with LNR,[30] to the "biharmonic" coordinates

$$(\varphi_1,\varphi_2,\vartheta_2,\ldots,\varphi_r,\vartheta_r,\vartheta_{r+1},\theta),$$
$$n \text{ odd and } =2r+3,$$

on the hyperboloid S_{n-31}, defined by

$$\begin{aligned}\xi_1 &= \cos\varphi_1\sin\vartheta_2\cdots\sin\vartheta_{r+1}\cosh\theta,\\ \xi_2 &= \sin\varphi_1\sin\vartheta_2\cdots\sin\vartheta_{r+1}\cosh\theta,\\ \xi_3 &= \cos\varphi_2\cos\vartheta_2\sin\vartheta_3\cdots\sin\vartheta_{r+1}\cosh\theta,\\ \xi_4 &= \sin\varphi_2\cos\vartheta_2\sin\vartheta_3\cdots\sin\vartheta_{r+1}\cosh\theta,\\ &\vdots\\ \xi_{2(r-t)-1} &= \cos\varphi_{r-t}\cos\vartheta_{r-t}\sin\vartheta_{r-t+1}\\ &\quad \cdots\sin\vartheta_{r+1}\cosh\theta,\\ \xi_{2(r-t)} &= \sin\varphi_{r-t}\cos\vartheta_{r-t}\sin\vartheta_{r-t+1}\\ &\quad \cdots\sin\vartheta_{r+1}\cosh\theta,\\ &\vdots\\ \xi_{2r-1} &= \cos\varphi_r\cos\vartheta_r\sin\vartheta_{r+1}\cosh\theta,\\ \xi_{2r} &= \sin\varphi_r\cos\vartheta_r\sin\vartheta_{r+1}\cosh\theta,\\ \xi_{2r+1} &= \cos\vartheta_{r+1}\cosh\theta,\\ \xi_{2r+2} &= \sinh\theta,\end{aligned} \tag{10a}$$

with

$$\varphi_j \in [0, 2\pi], \quad j = 1, 2, \ldots, r,$$
$$\vartheta_j \in [0, \pi/2], \quad j = 2, 3, \ldots, r,$$
$$\vartheta_{r+1} \in [0, \pi],$$
$$\theta \in (-\infty, \infty),$$

i.e., by

$$\sinh\theta = \xi_{2r+2},$$
$$\cosh\theta = (\xi_1^2 + \xi_2^2 + \cdots + \xi_{2r+1}^2)^{1/2},$$
$$\cos\vartheta_{r+1} = \frac{\xi_{2r+1}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2r+1}^2)^{1/2}},$$
$$\sin\vartheta_{r+1} = \frac{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2r}^2)^{1/2}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2r+1}^2)^{1/2}},$$
$$\cos\vartheta_r = \frac{(\xi_{2r-1}^2 + \xi_{2r}^2)^{1/2}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2r}^2)^{1/2}},$$
$$\sin\vartheta_r = \frac{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2r-2}^2)^{1/2}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2r}^2)^{1/2}}, \tag{10b}$$
$$\vdots$$
$$\cos\vartheta_{r-t} = \frac{(\xi_{2(r-t)-1}^2 + \xi_{2(r-t)}^2)^{1/2}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2(r-t)}^2)^{1/2}},$$
$$\sin\vartheta_{r-t} = \frac{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2(r-t)-2}^2)^{1/2}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_{2(r-t)}^2)^{1/2}},$$
$$\vdots$$
$$\cos\vartheta_2 = \frac{(\xi_3^2 + \xi_4^2)^{1/2}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_4^2)^{1/2}},$$
$$\sin\vartheta_2 = \frac{(\xi_1^2 + \xi_2^2)^{1/2}}{(\xi_1^2 + \xi_2^2 + \cdots + \xi_4^2)^{1/2}},$$
$$\left.\begin{aligned} \sin\varphi_j &= \frac{\xi_{2j}}{(\xi_{2j-1}^2 + \xi_{2j}^2)^{1/2}}, \\ \cos\varphi_j &= \frac{\xi_{2j-1}}{(\xi_{2j-1}^2 + \xi_{2j}^2)^{1/2}}, \end{aligned}\right\} \quad j = 1, 2, \ldots, r.$$

Then if

$$(\varphi_1, \varphi_2, \ldots, \varphi_r, \vartheta_2, \vartheta_3, \ldots, \vartheta_{r+1}, \theta)$$

are the biharmonic coordinates of

$$(\xi_1, \xi_2, \ldots, \xi_{n-1}),$$

those of

$$\left(\frac{\xi_1}{A}, \frac{\xi_2}{A}, \ldots, \frac{\xi_{n-2}}{A}, \frac{\sin\alpha + \xi_{n-1}\cos\alpha}{A}\right)$$

will be

$$(\varphi_1, \varphi_2, \ldots, \varphi_r, \vartheta_2, \vartheta_3, \ldots, \vartheta_{r+1}, \theta'),$$

where

$$\sinh\theta' = \frac{\xi_{n-1}\cos\alpha + \sin\alpha}{-\xi_{n-1}\sin\alpha + \cos\alpha},$$

i.e.,

$$\sinh\theta' = \frac{\sinh\theta\cos\alpha + \sin\alpha}{-\sinh\theta\sin\alpha + \cos\alpha}. \tag{11}$$

Thus we finally get

$$T^\sigma(r_{n-1\,n}(\alpha))F(\varphi_1, \varphi_2, \ldots, \varphi_r, \vartheta_2, \vartheta_3, \ldots, \vartheta_{r+1}, \theta) \tag{12}$$
$$= |-\sinh\theta\sin\alpha + \cos\alpha|^\sigma \times F(\varphi_1, \varphi_2, \ldots, \varphi_r, \vartheta_2, \vartheta_3, \ldots, \vartheta_{r+1}, \theta')$$

where θ' is given by (11).

V. MATRIX ELEMENTS OF FINITE ROTATIONS

We now come to the calculation of the matrix elements of finite rotations of $SO(n-2, 2)$ in the representation T^σ. As seen earlier, we need to carry it out only for those of the rotation $r_{n-1\,n}(\alpha)$. For this purpose, we need, of course, a set of basis vectors (functions) spanning the space $D^\infty(S_{n-3\,1})$; we take this as the following one given by LNR[28]:

$$|L, +, l_2, l_3, \ldots, l_{r+1}, m_1, m_2, \ldots, m_r\rangle,$$

with

$$L = -r+1, -r+2, \ldots, \tag{13a}$$
$$l_{r+1} = L+2, L+4, \ldots;$$

$$|L, -, l_2, l_3, \ldots, l_{r+1}, m_1, m_2, \ldots, m_r\rangle,$$

with

$$L = -r+1, -r+2, \ldots, \tag{13b}$$
$$l_{r+1} = L+1, L+3, \ldots;$$

$$|\Lambda, +, l_2, l_3, \ldots, l_{r+1}, m_1, m_2, \ldots, m_r\rangle,$$

with

$$0 \leqslant \Lambda < \infty, \tag{13c}$$
$$l_{r+1} = 0, 1, 2, \ldots;$$

$$|\Lambda, -, l_2, l_3, \ldots, l_{r+1}, m_1, m_2, \ldots, m_r\rangle,$$

with

$$0 \leqslant \Lambda < \infty, \tag{13d}$$
$$l_{r+1} = 0, 1, 2, \ldots;$$

where

$$|L, \pm, l_2, l_3, \ldots, l_{r+1}, m_1, m_2, \ldots, m_r\rangle = V_{l_{r+1}}^{\pm L}(\theta)\, Y_{m_1 m_2 \cdots m_r}^{l_2 l_3 \cdots l_{r+1}}(\omega),$$

with

$$V_{l_{r+1}}^{+L}(\theta) = -(2/\sqrt{N_1})\tanh\theta\cosh^{(-L+2r)}\theta \times {}_2F_1(\tfrac{1}{2}(L + l_{r+1} + 2r + 1), \tfrac{1}{2}(L - l_{r+1} + 2); \tfrac{3}{2}; \tanh^2\theta), \tag{14a}$$

$$V_{l_{r+1}}^{-L}(\theta) = (1/\sqrt{N_2})\cosh^{-(L+2r)}\theta \times {}_2F_1(\tfrac{1}{2}(L + l_{r+1} + 2r), \tfrac{1}{2}(L - l_{r+1} + 1); \tfrac{1}{2}; \tanh^2\theta), \tag{14b}$$

and

$$|\Lambda, \pm, l_2, l_3, \ldots, l_{r+1}, m_1, m_2, \ldots, m_r\rangle = V_{l_{r+1}}^{\pm\Lambda}(\theta)\, Y_{m_1 m_2 \cdots m_r}^{l_2 l_3 \cdots l_{r+1}}(\omega),$$

with

$$V_{l_{r+1}}^{+\Lambda}(\theta) = -\frac{2}{\sqrt{K_1}} \tanh\theta \cosh^{-(r+i\Lambda)}\theta \times {}_2F_1(\tfrac{1}{2}(i\Lambda + l_{r+1} + r + 1), \tfrac{1}{2}(i\Lambda - l_{r+1} - r + 2); \tfrac{3}{2}; \tanh^2\theta), \quad (14c)$$

$$V_{l_{r+1}}^{-\Lambda}(\theta) = \frac{1}{\sqrt{K_2}} \cosh^{-(r+i\Lambda)}\theta \times {}_2F_1(\tfrac{1}{2}(i\Lambda + r), \tfrac{1}{2}(i\Lambda - l_{r+1} - r + 1); \tfrac{1}{2}; \tanh^2\theta). \quad (14d)$$

Here ω stands for the set of angles

$$\{\varphi_1, \varphi_2, \ldots, \varphi_r, \vartheta_2, \vartheta_3, \ldots, \vartheta_{r+1}\},$$

while

$$Y_{m_1 m_2 \cdots m_r}^{l_2 l_3 \cdots l_{r+1}}(\omega)$$

are the harmonic functions on the sphere S_{n-3} given in Ref. 30, where the ranges of values of $l_2, l_3, \ldots, l_r, m_1, m_2, \ldots, m_r$ are also given.

Let us now calculate the matrix elements of $r_{n-1\,n}(\alpha)$ between two basis vectors of the type (12a). We have

$$\begin{aligned} M &= \langle L, +, l_2, l_3, \ldots, l_{r+1}, m_1, m_2, \ldots, m_r| \\ &\quad \times T^\sigma(r_{n-1\,n}(\alpha)) \\ &\quad \times |L', +, l'_2, l'_3, \ldots, l'_{r+1}, m'_1, m'_2, \ldots, m'_r\rangle \\ &= \int d\mu(\omega,\theta) V_{l_{r+1}}^{+L}(\theta) Y_{m_1\cdots m_r}^{l_2\cdots l_{r+1}}(\omega) T^\sigma(r_{n-1\,n}(\alpha)) \\ &\quad \times \{V_{L'_{r+1}}^{+L'}(\theta) Y_{m'_1\cdots m'_r}^{l'_2\cdots l'_{r+1}}(\omega)\}, \end{aligned}$$

where $d\mu(\omega,\theta)$ is the invariant measure on $S_{n-3\,1}$ and is therefore given by[30]

$$d\mu(\omega,\theta) = d\mu(\omega)\cosh^2\theta\, d\theta$$

with $d\mu(\omega)$ being the invariant measure on S_{n-3} (see Ref. 30). Now using the formula (12) and the orthonormality of Y's (see Ref. 28), we get

$$\begin{aligned} M &= \int d\mu(\omega) Y_{m_1\cdots m_r}^{l_2\cdots l_{r+1}}(\omega) Y_{m'_1\cdots m'_r}^{l'_2\cdots l'_{r+1}}(\omega) \\ &\quad \times \int_{-\infty}^{\infty} \cosh^2\theta\, d\theta\, V_{l_{r+1}}^{+L}(\theta) V_{l'_{r+1}}^{+L'}(\theta') \\ &\quad \times |-\sinh\theta\sin\alpha + \cos\alpha|^\sigma \\ &= \prod_{a=2}^{r+1} \delta_{l_a l'_a} \prod_{b=1}^{r} \delta_{m_b m'_b} \\ &\quad \times \int_{-\infty}^{\infty} d\theta \cosh^2\theta\, |-\sinh\theta\sin\alpha + \cos\alpha|^\sigma \\ &\quad \times V_{l_{r+1}}^{+L}(\theta) V_{l'_{r+1}}^{+L'}(\theta'). \end{aligned}$$

Denoting the product of Kronecker deltas by δ, we write this as

$$M = \delta \int_{-\infty}^{\infty} d\theta \cosh^2\theta\, |-\sinh\theta\sin\alpha + \cos\alpha|^\sigma \times V_{l_{r+1}}^{+L}(\theta) V_{l'_{r+1}}^{+L'}(\theta'). \quad (15)$$

Remembering that

$$0<\alpha<\pi \Rightarrow \sin\alpha > 0,$$

we note that

$$-\sinh\theta\sin\alpha + \cos\alpha$$

is a monotonically decreasing function of θ, starting from $+\infty$ at $\theta = -\infty$, going to the value 0 at

$$\theta = \theta_0 = \sinh^{-1}(\cot\alpha),$$

and then decreasing to $-\infty$ at $\theta = \infty$. Hence

$$(-\sinh\theta\sin\alpha + \cos\alpha) = \begin{cases} -(-\sinh\theta\sin\alpha + \cos\alpha), & \theta > \theta_0, \\ -\sinh\theta\sin\alpha + \cos\alpha, & \theta < \theta_0. \end{cases}$$

It follows that

$$M = M_1 + M_2,$$

where

$$M_1 = \delta \int_{-\infty}^{\theta_0} d\theta \cosh^2\theta (-\sinh\theta\sin\alpha + \cos\alpha)^\sigma \times V_{l_{r+1}}^{+L}(\theta) V_{l'_{r+1}}^{+L'}(\theta'), \quad (16a)$$

$$M_2 = \delta \int_{\theta_0}^{\infty} d\theta \cosh^2\theta (-1)^\sigma (-\sinh\theta\sin\alpha + \cos\alpha)^\sigma \times V_{l_{r+1}}^{+L}(\theta) V_{l'_{r+1}}^{+L'}(\theta'). \quad (16b)$$

It will be shown in the Appendix that

$$M_2 = (-1)^{\sigma+1} M_1 \quad (17)$$

so that

$$M = \{1 - (-1)^\sigma\} M_1, \quad (18)$$

i.e., it is sufficient to calculate the integral M_1. In order to carry out this rather complicated integration, we proceed as follows.

We put

$$L - l_{r+1} = -(2k+2),$$
$$L' - l'_{r+1} = -(2k'+2),$$

so that k, k' take on the values

$$0, 1, 2, \ldots .$$

Then

$$\begin{aligned} L + l_{r+1} + 2r + 1 &= L + L + 2k + 2 + 2r + 1 \\ &= 2(L + k + r + \tfrac{3}{2}), \\ L - l_{r+1} + 2 &= L - L - 2k - 2 + 2 = -2k. \end{aligned}$$

Hence, remembering that

$$F(a, b; c; z) = F(b, a; c; z),$$

we get from (14a),

$$\begin{aligned} V_{l_{r+1}}^{+L}(\theta) &= -(2/\sqrt{N_1})\tanh\theta\,\text{sech}^{L+2r}\theta F(-k, L + k + r + \tfrac{3}{2}; \tfrac{3}{2}; \tanh^2\theta) \\ &= -(2/\sqrt{N_1})\tanh\theta\,\text{sech}^{L+2r}\theta F(-k, -L - r - k; \tfrac{3}{2}; -\sinh^2\theta), \end{aligned}$$

using Eq. (22), p. 64 of Ref. 31;

$$= -(2/\sqrt{N_1})\tanh\theta\,\mathrm{sech}^{L+2r+2k}\theta\,\frac{(-1)^k k!}{(-2k-L-r-1)_{k+1}}\,\frac{1}{2i\sinh\theta}\,C_{2k+1}^{-2k-L-r-1}(i\sinh\theta),$$

using Eq. (22), p. 176 of Ref. 32, $C_n^\alpha(x)$ being the Gegenbauer polynomial;

$$=\frac{i}{\sqrt{N_1}}\frac{(-1)^k k!}{(-2k-L-r-1)_{k+1}}\,\mathrm{sech}^{L+2r+2k+1}\theta\, C_{2k+1}^{-2k-L-r-1}(i\sinh\theta)$$

$$=\frac{i}{\sqrt{N_1}}\frac{(-1)^k k!}{(-2k-L-r-1)_{k+1}}\,\mathrm{sech}^{L+2r+2k+1}\theta\,\frac{\Gamma(-2k-L-r-\frac{1}{2})\Gamma(-2k-2L-2r-1)}{\Gamma(-4k-2L-2r-2)\Gamma(-L-r+\frac{1}{2})}$$
$$\times P_{2k+1}^{(-2k-L-r-3/2,\,-2k-L-r-3/2)}(i\sinh\theta),$$

using the expression for Gegenbauer polynomials in terms of Jacobi polynomials given in Ref. 21;

$$=\frac{i}{\sqrt{N_1}}\frac{(-1)^k k!}{(-2k-L-r-1)_{k+1}}\,\mathrm{sech}^{L+2r+2k+1}\theta\,\frac{\Gamma(-2k-L-r-\frac{1}{2})\Gamma(-2k-2L-2r-1)}{\Gamma(-4k-2L-2r-2)\Gamma(-L-r+\frac{1}{2})}$$
$$\times 2^{-2k-1}\sum_{m=0}^{2k+1}\frac{\Gamma(-L-r+\frac{1}{2})\Gamma(-L-r+\frac{1}{2})}{\Gamma(m+1)\Gamma(2k+2-m)\Gamma(-L-r-m-\frac{1}{2})\Gamma(-L-r-2k+m-\frac{3}{2})}$$
$$\times(i\sinh\theta-1)^{2k+1-m}(i\sinh\theta+1)^m,$$

using this time the standard expansion of Jacobi polynomials.[32] Thus

$$V_{l_{r+1}}^{+L}(\theta)=\frac{i}{\sqrt{N_1}}\frac{(-1)^k k!2^{-2k-1}}{(-2k-L-r-1)_{k+1}}\frac{\Gamma(-2k-L-r-\frac{1}{2})}{\Gamma(-4k-2L-2r-2)}$$
$$\times\Gamma(-2k-2L-2r-1)\Gamma(-L-r+\tfrac{1}{2})\mathrm{sech}^{L+2r+2k+1}\theta$$
$$\times\sum_m\frac{(i\sinh\theta+1)^m(i\sinh\theta-1)^{2k+1-m}}{\Gamma(m+1)\Gamma(2k+2-m)\Gamma(-L-r-m-\frac{1}{2})\Gamma(-L-r-2k+m-\frac{3}{2})}$$
$$=i\mathscr{N}_1\,\mathrm{sech}^{L+2r+2k+1}\theta\sum_m\frac{(i\sinh\theta+1)^m(i\sinh\theta-1)^{2k+1-m}}{\Gamma(m+1)\Gamma(2k+2-m)\Gamma(-L-r-m-\frac{1}{2})\Gamma(-L-r-2k+m-\frac{3}{2})},\tag{19}$$

where

$$\mathscr{N}_1=\frac{1}{\sqrt{N_1}}\frac{(-1)^k k!2^{-2k-1}}{(-2k-L-r-1)_{k+1}}\frac{\Gamma(-2k-L-r-\frac{1}{2})}{\Gamma(-4k-2L-2r-2)}\Gamma(-2k-2L-2r-1)\Gamma(-L-r+\tfrac{1}{2}).\tag{20}$$

It follows that

$$V_{l'_{r+1}}^{+L'}(\theta')=i\mathscr{N}'_1\,\mathrm{sech}^{L'+2r+2k'+1}\theta'$$
$$\times\sum_{m'=0}^{2k'+1}\frac{(i\sinh\theta'+1)^{m'}(i\sinh\theta'-1)^{2k'+1-m'}}{\Gamma(m'+1)\Gamma(2k'+2-m')\Gamma(-L'-r-m'-\frac{1}{2})\Gamma(-L'-r-2k'+m'-\frac{3}{2})},\tag{21}$$

where $\mathscr{N}'_1$ is obtained from $\mathscr{N}_1$ by replacing L by L' and k by k'.

Let us now put

$$t=\tfrac{1}{2}(i\sinh\theta-1)\tag{22}$$

$$\Rightarrow -2i\,dt=\cosh\theta\,d\theta,\quad 1+t=\tfrac{1}{2}(i\sinh\theta+1).\tag{23}$$

Then it is easy to check that, with $z=1-e^{-2i\alpha}$,

$$-\sinh\theta\sin\alpha+\cos\alpha=e^{i\alpha}(1+zt),\tag{24}$$

$$i\sinh\theta'+1=2(1+t)/(1+zt),\tag{25a}$$

$$i\sinh\theta'-1=2e^{-2i\alpha}t/(1+zt),\tag{25b}$$

$$\mathrm{sech}^{L+2r+2k}\theta=(1+i\sinh\theta)^{-r-k-L/2}(1-i\sinh\theta)^{-r-k-L/2}$$
$$=2^{-L-2r-2k}(-1)^{-r-k-L/2}t^{-r-k-L/2}(1+t)^{-r-k-L/2},\tag{26a}$$

$$\mathrm{sech}^{L'+2r+2k'+1}\theta'=2^{-L'-2r-2k'-1}(-1)^{-r-k'-(L'+1)/2}e^{2i(r+k'+(L'+1)/2)\alpha}$$
$$\times t^{-r-k'-(L'+1)/2}(1+t)^{-r-k'-(L'+1)/2}(1+zt)^{L'+2r+2k'+1}.\tag{26b}$$

Using Eqs. (19)–(26), (14a) gives ($t_0=\sinh\theta_0$)

$$M_1=\delta\int_{-1/2-i\infty}^{-1/2+it_0} -2i\,dt\,e^{i\alpha\sigma}(1+zt)^{\sigma}i\mathcal{N}_1 i\mathcal{N}_1' 2^{-L-2r-2k}(-1)^{-r-k-L/2}t^{-r-k-L/2}(1+t)^{-r-k-L/2}$$
$$\times 2^{-L'-2r-2k'-1}e^{2i(r+k'+(L'+1)/2)\alpha}(-1)^{-r-k'-(L'+1)/2}t^{-r-k'-(L'+1)/2}(1+t)^{-r-k'-(L'+1)/2}$$
$$\times(1+zt)^{L'+2r+2k'+1}\sum_m\sum_{m'}\frac{2^m(1+t)^m\cdot 2^{2-k-m}t^{2k+1-m}}{\Gamma(m+1)\Gamma(2k+2-m)\Gamma(-L-r-m-\frac{1}{2})\Gamma(-L-r-2k+m-\frac{3}{2})}$$
$$\times\frac{2^{m'}(1+t)^{m'}(1+zt)^{-m'}2^{2k'+1-m'}e^{-2i(2k'+1-m')\alpha}t^{2k'+1-m'}(1+zt)^{-2k'-1+m'}}{\Gamma(m'+1)\Gamma(2k'+2-m')\Gamma(-L'-r-m'-\frac{1}{2})\Gamma(-L'-r-2k'+m'-\frac{3}{2})}$$
$$=i\delta\mathcal{N}_1\mathcal{N}_1' 2^{-L-L'-4r+2}(-1)^{-k-k'-2r-(L+L'+1)/2}e^{i\alpha(\sigma+2r-2k'+L'-1)}$$
$$\times\sum_m\sum_{m'}e^{2im'\alpha}\Big\{\Gamma(m+1)\Gamma(2k+2-m)\Gamma\Big(-L-r-m-\frac{1}{2}\Big)\Gamma\Big(-L-r-2k+m-\frac{3}{2}\Big)$$
$$\times\Gamma(m'+1)\Gamma(2k'+2-m')\Gamma\Big(-L'-r-m'-\frac{1}{2}\Big)\Gamma\Big(-L'-r-2k'+m'-\frac{3}{2}\Big)\Big\}^{-1}\times I, \tag{27}$$

where I is the integral

$$I=\int_{-1/2-i\infty}^{-1/2+it_0} dt\,t^{-(L+L'-3)/2-2r-m-m'+k+k'}(1+t)^{-(L+L'+1)/2-2r-k-k'+m+m'}(1+zt)^{L'+\alpha+2r}. \tag{28}$$

It is calculated in the Appendix for the general exponents α, β, γ. As here,

$$\alpha=-(L+L'-3)/2-2r-m-m'+k+k',$$
$$\beta=-(L+L'+1)/2-2r-k-k'+m+m',$$
$$\gamma=L'+\alpha+2r\equiv L'+i\rho-(n-1)/2+2r,$$

and the definition of k, k' [given after Eq. (18)] implies that

$L+L'$ is an even integer $\Rightarrow (L+L')/2$ is an integer,

we shall have

$$\exp(2\pi i\alpha)=\exp(i\pi)=-1,$$
$$\exp(2\pi i\beta)=\exp(i\pi)=-1,$$
$$\exp(2\pi i\gamma)=\exp(in\pi)\exp(-2\pi\rho)=(-1)^n\exp(-2\pi\rho).$$

Hence Eq. (A6) gives

$$I=\frac{1}{-1+(-1)^n e^{-2\pi\rho}}\Big[2\frac{\Gamma(-(L+L'-5)/2-2r-m-m'+k+k')}{\Gamma(-\sigma-(L'-L-1)/2+2r+k+k'-m-m')}\Gamma(-\sigma+L+2r-1)$$
$$\times{}_2F_1\Big(-\sigma-L'-2r,-\frac{L+L'-5}{2}-2r-m-m'+k+k';-\sigma-\frac{L'-L-1}{2}+k+k'-m-m';e^{-2i\alpha}\Big)$$
$$-2\frac{\Gamma(-(L+L'-1)/2-2r-k-k'+m+m')}{\Gamma-\sigma-(L-L'+3)/2+m+m'-k-k')}\Gamma(-\sigma+L+2r-1)e^{-2i\alpha(L'+\sigma+2r)}$$
$$\times{}_2F_1\Big(-\sigma-L'-2r,-\frac{L+L'-1}{2}-2r-k-k'+m+m';-\sigma-\frac{L'-L+3}{2}-m-m'+k+k';e^{2i\alpha}\Big)\Big]. \tag{29}$$

We thus see that I depends only on $(m+m')$ and not on m or m' individually. If we therefore change the (m, m') double summation in (27) to the (μ, m') summation, where $\mu=m+m'$, I will come out of the m' summation. Using

$$\Gamma(z)\Gamma(1-z)=\pi\csc\pi z,$$

this summation will become

$$\sum_{m'}\frac{e^{2im'\alpha}}{m'!}\frac{\Gamma(-\mu+m')\Gamma(L+r+2k-\mu+m'+\frac{3}{2})\Gamma(-2k'+m'-1)}{\Gamma(2k+2+m'-\mu)\Gamma(-L-r-\mu+m'-\frac{1}{2})\Gamma(-L'-r-2k'+m'-\frac{3}{2})}$$
$$\times\Gamma(L'+r+m'+\tfrac{3}{2})\frac{1}{\pi^4}\sin\pi(1+\mu-m')\sin\pi(-L-r-2k+\mu-m'-\tfrac{3}{2})$$
$$\times\sin\pi(-L'-r-m'-\tfrac{1}{2})\sin\pi(2k'+2-m')$$
$$=\Big[\sum_{m'}\frac{(-\mu)_{m'}(L+r+2k-\mu+\frac{3}{2})_{m'}(-2k-1)_{m'}(L'+r+\frac{3}{2})_{m'}e^{2im'\alpha}}{(2k+2-\mu)_{m'}(-L-r-\mu-\frac{1}{2})_{m'}(-L-r-2k'+\frac{3}{2})_{m'}m'!}\Big]$$

$$\times\frac{\Gamma(-\mu)\Gamma(L+r+2k-\mu+\frac{5}{2})\Gamma(-2k'-1)\Gamma(L'+r+\frac{3}{2})}{\Gamma(2k+2-\mu)\Gamma(-L-r-\mu-\frac{1}{2})\Gamma(-L'-r-2k'-\frac{3}{2})}$$

$$\times\frac{1}{\pi^4}\sin\pi(1+\mu)\sin\pi(-L-r-2k+\mu-\tfrac{3}{2})\sin\pi(-L'-r-\tfrac{1}{2})\sin\pi(2k'+2)$$

$$={}_4F_3\begin{pmatrix}-\mu, L+r+2k-\mu+\frac{5}{2}, -2k-1, L'+r+\frac{3}{2}; & e^{2i\alpha}\\ 2k+2-\mu, -L-r-\mu-\frac{1}{2}, -L'-r-2k-\frac{3}{2}; & \end{pmatrix}$$

$$\times\{\Gamma(1+\mu)\Gamma(-L-r-2k+\mu-\tfrac{3}{2})\Gamma(-L'-r-\tfrac{1}{2})\Gamma(2k'+2)$$

$$\times\Gamma(2k+2-\mu)\Gamma(-L-r-\mu-\tfrac{1}{2})\Gamma(-L'-r-2k'-\tfrac{3}{2})\}^{-1}.$$

Using this and (29), (27) becomes

$$M_1=\frac{i\delta\mathcal{N}_1\mathcal{N}_1'}{-1+(-1)^n e^{-2\pi p}}2^{-L-L'-4r-3}(-1)^{-k-k'-2r-(L+L'+1)/2}$$

$$\times\frac{\Gamma(-\sigma+L+2r-1)e^{-i(2k'+1)\alpha}}{\Gamma(-L'-r-\frac{1}{2})\Gamma(2k'+2)\Gamma(-L'-r-2k'-\frac{3}{2})}$$

$$\times\sum_{\mu}\Big[\{\Gamma(1+\mu)\Gamma(-L-r-2\mu+\mu-\tfrac{3}{2})\Gamma(2k+2-\mu)\Gamma(-L-r-\mu-\tfrac{1}{2})\}^{-1}$$

$$\times{}_4F_3\begin{pmatrix}-\mu, L+r+2k-\mu+\frac{5}{2}, -2k-1, L'+r+\frac{3}{2}; & e^{2i\alpha}\\ 2k+2-\mu, -L-r-\mu-\frac{1}{2}, -L'-r-2k-\frac{3}{2}; & \end{pmatrix}$$

$$\times\Big\{\frac{\Gamma(-(L+L'-5)/2-2r-\mu+k+k')}{\Gamma(-\sigma-(L'-L-1)/2+2r+k+k'-\mu)}e^{i(\sigma+2r+L')\alpha}$$

$$\times{}_2F_1\Big(-\sigma-L'-2r-\frac{L+L'-5}{2}-2r-\mu+k+k';-\sigma-\frac{L'-L-1}{2}+k+k'-\mu;e^{-2i\alpha}\Big)$$

$$-\frac{\Gamma(-(L+L'-1)/2-2r-k-k'+\mu)}{\Gamma(-\sigma-(L-L'+3)/2+\mu-k-k')}e^{-i(\sigma+2r+L')\alpha}$$

$$\times{}_2F_1\Big(-\sigma-L'-2r,-\frac{L+L'-1}{2}-2r-k-k'+\mu;-\sigma-\frac{L'-L+3}{2}-\mu+k+k';e^{2i\alpha}\Big)\Big\}\Big]. \quad (30)$$

This value of M_1 gives the required matrix element M through (18).

The above computation has been carried out only for the matrix elements between two states (basis functions) of the type (13a); for completeness' sake, we should really carry them out for all possible pair of states (13a)–(13d). In addition, the same thing should be repeated for even n. However, as these large number of lengthy calculations (leading to equally lengthy results) are not going to give us any additional insight, we content ourselves with just one illustrative calculation and the remark that others can be carried out in exactly the same way.

ACKNOWLEDGMENTS

The author wishes to thank (Pakistan) University Grants Commission for sponsoring a two weeks visit to the Quaid-e-Azam University, Islamabad, during the progress of the present work, and to the Chairman of the Mathematics Department of that University for his kind hospitality during his stay there.

APPENDIX: EVALUATION OF AN IMPORTANT INTEGRAL

In this appendix, we prove Eq. (17) of the text and evaluate the integral

$$J=\int_{-1/2-i\infty}^{-1/2+it_0}t^\alpha(1+t)^\beta(1+zt)^\gamma\,dt,$$

$$z=1-e^{-2i\alpha},\quad t_0=\tfrac{1}{2}\cot\alpha,$$

which appears in Eq. (28). Note first of all that

$$-\frac{1}{2}-it_0=-\frac{1}{2}+\frac{1}{2}i\cot\alpha=-\frac{1}{2}\Big(1+\frac{e^{i\alpha}+e^{-i\alpha}}{e^{i\alpha}-e^{-i\alpha}}\Big)$$

$$=-\frac{1}{2}\frac{2e^{i\alpha}}{e^{i\alpha}-e^{-i\alpha}}=-\frac{1}{1-e^{-2i\alpha}}=-\frac{1}{z},$$

so that the point A: $-\frac{1}{2}+it_0$ is the only zero of $(1+zt)$ in the complex t plane. We denote the other two zeros $t=0$ and $t=-1$ of the integrand by 0 and B, respectively.

Consider now the integral

$$\hat{J}=\int_D t^\alpha(1+t)^\beta(1+zt)^\gamma\,dt,$$

where D is the contour shown in Fig. 1. As the integrand has no singularity within D, $\hat{J}=0$. Also, for

$$\mathrm{Re}(\alpha+\beta+\gamma)<-1, \quad (A1)$$

contributions to $\hat{J}$ from portions of D consisting of the infinite circle $|t|=R\to\infty$, vanish so that we get

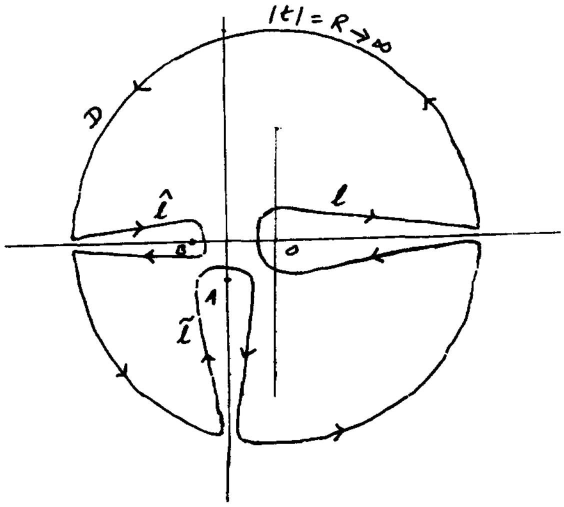

FIG. 1. The contour D in the complex t plane.

$$-\int_{\tilde l} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=\int_{l} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$+\int_{\hat l} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt. \qquad \text{(A2)}$$

Suppose now that

$$\operatorname{Re}\alpha>-1,\quad \operatorname{Re}\beta>-1,\quad \operatorname{Re}\gamma>-1. \qquad \text{(A3)}$$

Then the portions l, $\hat l$, and $\tilde l$ of D can be taken as in Fig. 2. Now

$$t^{\alpha}|_{-l_2}=t^{\alpha}|_{l_1}e^{2\pi i\alpha},$$
$$(1+t)^{\beta}(1+zt)^{\gamma}|_{-l_2}=(1+t)^{\beta}(1+zt)^{\gamma}|_{l_1}$$
$$\Rightarrow\int_{l} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=\left(\int_{l_1}-\int_{-l_2}\right)\{t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\}dt$$
$$=(1-e^{2\pi i\alpha})\int_{l_1} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=(1-e^{2\pi i\alpha})\int_{0}^{\infty} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt,$$

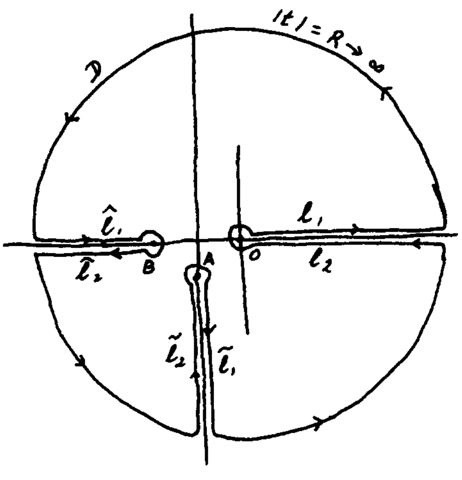

FIG. 2. Possible form of the contour D.

where $\arg t=0$ in the integrand. Similarly

$$-\int_{\tilde l} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=-(1-e^{2\pi i\gamma})\int_{\tilde l_1} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=(1-e^{2\pi i\gamma}\int_{-1/2-i\infty}^{-1/2+it_0} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt,$$

with $\arg(1+zt)=-\pi/2$ in the integrand, and

$$\int_{\hat l} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=(1-e^{2\pi i\beta})\int_{-1}^{-\infty} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt,$$

with $\arg(1+t)=-\pi$ in the integrand,

$$=(1-e^{2\pi i\beta})(-1)^{\alpha+\beta+\gamma+1}z^{\gamma}$$
$$\times\int_{1}^{\infty} t^{\alpha}(t-1)^{\beta}(t-z^{-1})^{\gamma}\,dt.$$

Thus we get

$$\int_{-1/2-i\infty}^{-1/2+it_0} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=(1-e^{2\pi i\gamma})^{-1}\Big[(1-e^{2\pi i\alpha})\int_{0}^{\infty} t^{\alpha}(1+t)^{\beta}$$
$$\times(1+zt)^{\gamma}\,dt+(1-e^{2\pi i\beta})(-1)^{\alpha+\beta+\gamma+1}z^{\gamma}$$
$$\times\int_{1}^{\infty} t^{\alpha}(t-1)^{\beta}(t-z^{-1})^{\gamma}\,dt. \qquad \text{(A4)}$$

$$\hat J'=\int_{D'} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt,$$

where D' is the contour shown in Fig. 3, we can show that

$$\int_{-1/2+it_0}^{-1/2+i\infty} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$
$$=-(1-e^{2\pi i\gamma})^{-1}\Big[(1-e^{2\pi i\alpha})$$
$$\times\int_{0}^{\infty} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\,dt$$

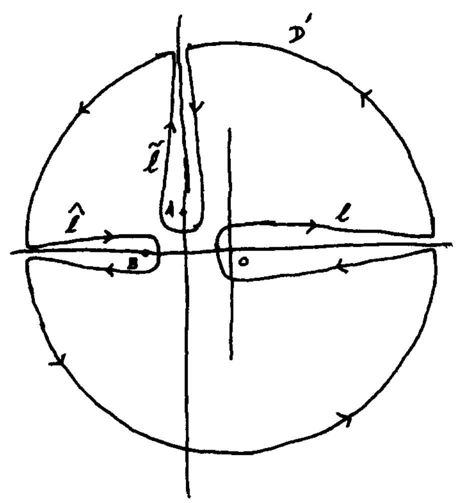

FIG. 3. The contour D' in the complex t plane.

$$+ (1 - e^{2\pi i\beta})(-1)^{\alpha+\beta+\gamma+1} z^{\gamma} \times \int_1^{\infty} t^{\alpha}(t-1)^{\beta}(t-z^{-1})^{\gamma}\, dt \Big],$$

which proves that

$$\int_{-1/2-i\infty}^{-1/2+it_0} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\, dt = -\int_{-1/2+it_0}^{-1/2+i\infty} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\, dt. \qquad \text{(A5)}$$

If we now compare the definitions (16a) and (16b) (of the text) of M_1 and M_2, we see that $M_2/(-1)^{\sigma}$ will be given by the expression (27) for M_1 except for the fact that the range of integration for I in (28) will be changed to

$$-\tfrac{1}{2} + it_0 \quad \text{to} \quad -\tfrac{1}{2} + i\infty.$$

Hence (A5) implies that

$$M_2/(-1)^{\sigma} = -M_1,$$

which proves (17).

To evaluate the right-hand side of (A4), we proceed as follows: Using Eq. (5), p. 115 of Ref. 31, with

$$\alpha = b - 1, \quad \beta = a - c, \quad \gamma = -a,$$

we have

$$\int_0^{\infty} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\, dt = \frac{\Gamma(1+\alpha)\Gamma(-\alpha-\beta-\gamma-1)}{\Gamma(-\beta-\gamma)} \times F(-\gamma, 1+\alpha; -\beta-\gamma; 1-z), \quad |\arg z| < \pi,$$

for

$$\mathrm{Re}(-\beta-\gamma) > \mathrm{Re}(\alpha+1) > 0,$$

i.e., for

$$\mathrm{Re}\,\alpha > -1, \quad \mathrm{Re}(\alpha+\beta+\gamma) < -1,$$

which is certainly satisfied if (A1) and (A3) are satisfied.

Next, using Eq. (6), p. 115 of Ref. 31, with

$$\alpha = a - c, \quad \beta = c - b - 1, \quad \gamma = -a,$$

we have

$$\int_1^{\infty} t^{\alpha}(t-1)^{\beta}(t-z^{-1})^{\gamma}\, dt = \frac{\Gamma(1+\beta)\Gamma(-\alpha-\beta-\gamma-1)}{\Gamma(-\gamma-\alpha)} \times F(-\gamma, -\alpha-\beta-\gamma-1; -\gamma-\alpha; z^{-1}), \quad |\arg(z-1)| < \pi,$$

for

$$1 + \mathrm{Re}(-\gamma) > \mathrm{Re}(-\gamma-\alpha) > \mathrm{Re}(-\alpha-\beta-\gamma-1),$$

i.e., for

$$\mathrm{Re}\,\alpha > -1, \quad \mathrm{Re}\,\beta > -1,$$

which is again certainly satisfied if (A1) and (A3) are satisfied.

Hence if (A1) and (A3) are satisfied, we will have

$$\int_{-1/2-i\infty}^{-1/2+it_0} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\, dt = (1-e^{2\pi i\gamma})^{-1}\Big[(1-e^{2\pi i\alpha}) \times \frac{\Gamma(1+\alpha)\Gamma(-\alpha-\beta-\gamma-1)}{\Gamma(-\beta-\gamma)} \times F(-\gamma, 1+\alpha; -\beta-\gamma; 1-z) + (1-e^{2\pi i\beta})(-1)^{\alpha+\beta+\gamma+1} \times \frac{\Gamma(1+\beta)\Gamma(-\alpha-\beta-\gamma-1)}{\Gamma(-\alpha-\gamma)} z^{\gamma} F(-\gamma, -\alpha-\beta-\gamma-1; -r-\alpha; z^{-1})\Big].$$

Now, from Eq. (27), p. 64 of Ref. 31, we have

$$F(-\gamma, -\alpha-\beta-\gamma-1; -\gamma-\alpha; z^{-1}) = (1-z^{-1})^{\gamma} F(-\gamma, 1+\beta; -\alpha-\gamma; (1-z)^{-1}),$$

so that

$$\int_{-1/2-i\infty}^{-1/2+it_0} t^{\alpha}(1+t)^{\beta}(1+zt)^{\gamma}\, dt = (1-e^{2\pi i\gamma})^{-1}\Big[(1-e^{2\pi i\alpha}) \times \frac{\Gamma(1+\alpha)\Gamma(-\alpha-\beta-\gamma-1)}{\Gamma(-\beta-\gamma)} \times F(-\gamma, 1+\alpha; -\beta-\gamma; 1-z) + (1-e^{2\pi i\beta})(-1)^{\alpha+\beta+1} \times \frac{\Gamma(1+\beta)\Gamma(-\alpha-\beta-\gamma-1)}{\Gamma(-\alpha-\gamma)}(1-z)^{\gamma} \times F(-\gamma, 1+\beta; -\alpha-\gamma; (1-z)^{-1})\Big]. \qquad \text{(A6)}$$

We have proved this relationship only when the conditions (A1) and (A3) are satisfied. However, as both of its sides are analytic functions of α, β, and γ, it follows by the principle of analytic continuation that the two sides will be equal for all those values of α, β, γ for which they do not have any singularity. Thus we finally have evaluated the integral J [Eq. (A6)] for all those values of the exponents α, β, γ that do not make it singular.

[1]E. P. Wigner, *Groupentheorie* (Vieweg, Braunschweig, 1931), pp. 180, 233.
[2]V. Bargmann, Ann. Math. **48**, 568 (1947).
[3]S. Strom, Ark. Fys. **34**, 215 (1967).
[4]Dao Vong Due and Nguyen Van Hieu, Dokl. Akad. Nauk. SSSR **137**, 1281 (1967); Ann. Inst. H. Poincaré **6**, 17 (1967).
[5]G. I. Kuznetsov, M. A. Liberman, A. A. Markov, and Ya. A. Smorodinski, Sov. J. Nucl. Phys. **10**, 370 (1970).
[6]K. Toth, Trieste preprint I. C. (69), 1969, p. 128.
[7]A. Markov and G. I. Shepelev, Sov. J. Nucl. Phys. **12**, 596 (1971).
[8]D. Z. Friedman and J. M. Wang, Phys. Rev. **160**, 1560 (1967).
[9]L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961).
[10]J. W. Holman, J. Math. Phys. **10**, 1710 (1969).
[11]J. W. Holman, J. Math. Phys. **10**, 888 (1969).
[12]S. Strom, Ark. Fys. **30**, 455 (1965); A. Kihlberg and S. Strom, Ark. Fys. **31**, 491 (1966).
[13]R. Takahashi, Bull. Soc. Math. France **91**, 289 (1963).
[14]S. C. Peng and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[15]M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967).
[16]K. B. Wolf, J. Math. Phys. **12**, 197 (1971).

[17]M. K. F. Wong, J. Math. Phys. **15**, 25 (1974).
[18]T. Mackawa, J. Math. Phys. **16**, 344 (1975).
[19]A. Syed, J. Math. Phys. **25**, 2132 (1984).
[20]S. Strom, Ark. Fys. **33**, 465 (1966).
[21]C. P. Boyer, J. Math. Phys. **12**, 1599 (1971).
[22]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962), Chap. X, Sec. 2.
[23]N. Limic, J. Niederle, and R. Raczka, J. Math. Phys. **7**, 2026 (1966).
[24]M. Bander and C. Iztykson, Rev. Mod. Phys. **38**, 346 (1966).
[25]C. P. Boyer and F. Ardalan, J. Math. Phys. **12**, 2070 (1971).
[26]N. Ja Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968), Chap. IX, Sec. 2.
[27]Reference **26**, Chap. X, Sec. 5.
[28]N. Limic, J. Niederle, and R. Raczka, J. Math. Phys. **8**, 1079 (1967).
[29]A. Syed, J. Math. Phys. **24**, 463 (1983).
[30]N. Limic, J. Niederle, and R. Raczka, J. Math. Phys. **7**, 1861 (1966).
[31]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.
[32]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II.

Relativistic plasma dispersion functions

P. A. Robinson
School of Physics, University of Sydney, New South Wales 2006, Australia

(Received 27 September 1985; accepted for publication 3 January 1986)

The known properties of plasma dispersion functions (PDF's) for waves in weakly relativistic, magnetized, thermal plasmas are reviewed and a large number of new results are presented. The PDF's required for the description of waves with small wave number perpendicular to the magnetic field (Dnestrovskii and Shkarofsky functions) are considered in detail; these functions also arise in certain quantum electrodynamical calculations involving strongly magnetized plasmas. Series, asymptotic series, recursion relations, integral forms, derivatives, differential equations, and approximations for these functions are discussed as are their analytic properties and connections with standard transcendental functions. In addition a more general class of PDF's relevant to waves of arbitrary perpendicular wave number is introduced and a range of properties of these functions are derived.

I. INTRODUCTION

Electron cyclotron waves in magnetized plasmas have a wide range of applications. These include electron cyclotron resonance heating of laboratory plasmas to achieve ignition and plasma profile control[1,2] and electron cyclotron current drive to enable continuous operation of tokamaks.[2,3] Electron cyclotron instabilities also have been investigated widely in recent years with applications to plasmas occurring in tokamaks,[4,5] magnetic mirrors, planetary and stellar magnetospheres,[6,7] solar flares,[8] and elsewhere. The general question of dispersion of electron cyclotron waves (including electron Bernstein waves) is also of current interest in studies of propagation, absorption, and mode conversion.[9–12]

Essential to each of the above applications is a knowledge of the dielectric properties of the plasma and the resulting dispersion of the relevant waves. Analytic treatment of these properties leads to expressions for the dielectric tensor in terms of *relativistic plasma dispersion functions* (henceforth, relativistic PDF's) analogous to the well-known plasma dispersion function discussed by Fried and Conte,[13] which is appropriate to waves in unmagnetized thermal plasmas. The most important such PDF's are those relevant to the case of weakly relativistic, magnetized, thermal plasmas; it is with these functions that the present paper will be concerned. We note that these functions also occur in quantum electrodynamical calculations involving strongly magnetized plasmas.[14,15]

The theory of relativistic PDF's presently consists largely of results scattered through a wide literature relating to the various applications mentioned above. Consequences of this situation are as follows.

(i) Several different definitions (and many notations) have been introduced for the relevant PDF's, thereby making comparison of papers by different authors difficult.

(ii) Having been obtained piecemeal by many authors, the results do not form a coherent whole. Furthermore, they suffer collectively from a number of omissions and shortcomings such as the lack of connection with standard transcendental functions of mathematical physics and lack of a detailed treatment of the analytic properties of the PDF's.

(iii) The PDF's required to treat electrostatic cyclotron waves such as Bernstein waves have been discussed only very briefly and few of their properties are known.

In this paper we collect and systematize the known properties of the PDF's for weakly relativistic thermal plasmas in one reference. In addition we include a large number of new properties of the most commonly discussed PDF's and a section on the hitherto inadequately treated PDF's appropriate to electrostatic cyclotron waves. We also take this opportunity to harmonize the many notations found in the literature and to clarify the relationships of various alternative PDF's to the ones considered here.

In Secs. II and III we discuss the relativistic PDF's appropriate to waves with small perpendicular (to the magnetic field) wave number in magnetized thermal plasmas—the *Dnestrovskii* and *Shkarofsky* functions. Interrelations, differential equations, analytic properties, series expansions, asymptotic forms, approximations, relations to the standard higher transcendental functions, and other properties are discussed. Section IV is concerned with a class of more general PDF's relevant to waves of arbitrary perpendicular wave number in weakly relativistic thermal plasmas. A number of useful integral forms are established for these functions and, in Sec. V, for the functions considered in Sec. II. Relationships between the PDF's considered in this paper and those introduced by other authors are discussed in Sec. VI.

II. PDF'S FOR WAVES WITH SMALL PERPENDICULAR WAVE NUMBER

In this section we discuss the properties of the relativistic PDF's appropriate to waves with small perpendicular wave number in weakly relativistic, thermal, magnetized plasmas—the Dnestrovskii and Shkarofsky functions.[16,17]

A. Definitions, notations

We define the *generalized Shkarofsky functions* of indexes q and r as follows,[18] if $\mathrm{Im}(z-a)>0$:

 0022-2488/86/051206-09$02.50

$$\mathscr{F}_{q,r}(z,a) = -i\int_0^\infty dt \frac{(it)^r}{(1-it)^q} \exp\left[izt - \frac{at^2}{1-it}\right] \quad (1)$$

$$= -ie^{-a}\int_0^\infty dt \frac{(it)^r}{(1-it)^q} \times\exp\left[i(z-a)t + \frac{a}{1-it}\right], \quad (2)$$

where q is real, r is a non-negative integer, and z and a are complex. Analytic continuation is used to extend this definition to $\mathrm{Im}(z-a)\leqslant 0$. (Alternatively, the definition may be extended by deforming the contour of integration to ensure convergence.) The corresponding *generalized Dnestrovskii functions*[16,18] are defined by

$$F_{q,r}(z) = \mathscr{F}_{q,r}(z,0), \quad (3)$$

while the usual Shkarofsky and Dnestrovskii functions[16,17] are

$$\mathscr{F}_q(z,a) = \mathscr{F}_{q,0}(z,a), \quad (4a)$$

$$F_q(z) = F_{q,0}(z), \quad (4b)$$

respectively.

The functions $\mathscr{F}_q(z,a)$ defined here are identical (apart from notation) to the functions used by Airoldi and Orefice,[19] Krivenski and Orefice,[20] Wong *et al.*,[21] Wu *et al.*,[22] Maroli and Petrillo,[23] Bornatici *et al.*,[2] Imre and Weitzner,[12] and Robinson,[18,24] among others. The functions $F_q(z)$ are identical with the original functions introduced by Dnestrovskii *et al.*[16] The connection between the present functions and those introduced by other authors is discussed in Sec. VI.

B. Derivatives

Derivatives of $\mathscr{F}_{q,r}(z,a)$ may be obtained immediately from (1), giving

$$\frac{\partial^{j+k}}{\partial z^j \partial a^k}\mathscr{F}_{q,r}(z,a) = \mathscr{F}_{q+k,r+j+2k}(z,a), \quad (5)$$

which relation has the special case

$$\mathscr{F}_{q,r}(z,a) = \frac{\partial^r}{\partial z^r}\mathscr{F}_q(z,a). \quad (6)$$

Derivatives of $e^{-z}F_q(z)$ are given in (18).

C. Sums

Equation (2) implies that the generalized Shkarofsky functions $\mathscr{F}_{q,r}$ may be reexpressed in terms of the usual Shkarofsky functions ($r=0$) thus:

$$\mathscr{F}_{q,r}(z,a) = \sum_{j=0}^{r}(-1)^j\binom{r}{j}\mathscr{F}_{q-j}(z,a), \quad (7)$$

where the binomial expansion has been used. Similarly, if the numerator and denominator in (1) are multiplied by $(1-it)^s$, where s is a non-negative integer, we find

$$\mathscr{F}_{q,r}(z,a) = \sum_{j=0}^{s}(-1)^j\binom{s}{j}\mathscr{F}_{q+s,r+j}(z,a). \quad (8)$$

Expansion of the factors $\exp[at^2/(1-it)]$ and $\exp[a/(1-it)]$ in powers of a in (1) and (2), respectively, leads to the following relations:

$$\mathscr{F}_q(z,a) = \sum_{j=0}^{\infty}\frac{a^j}{j!}F_{q+j,2j}(z) \quad (9)$$

and[17]

$$\mathscr{F}_q(z,a) = e^{-a}\sum_{j=0}^{\infty}\frac{a^j}{j!}F_{q+j}(z-a). \quad (10)$$

Similar results are easily obtained for $\mathscr{F}_{q,r}(z,a)$.

The generating function for the modified Bessel functions $I_j(z)$ is

$$\exp[a(u+u^{-1})] = \sum_{j=-\infty}^{\infty}u^jI_j(2a). \quad (11)$$

If (1) is rewritten in the alternative form

$$\mathscr{F}_q(z,a) = -ie^{-2a}\int_0^\infty dt(1-it)^{-q} \times\exp[izt + a(1-it) + a/(1-it)], \quad (12)$$

then (11) yields the identity

$$\mathscr{F}_q(z,a) = \sum_{j=-\infty}^{\infty}e^{-2a}I_j(2a)F_{q-j}(z). \quad (13)$$

Results obtained later [Eqs. (21)–(23)] imply that the series in (9) and (13) converge for $|a|<|z|$.

D. Recursion relations

The recursion relation for $F_q(z)$ can be obtained by integrating (1) by parts (with $a=0$, $r=0$). This yields[19]

$$(q-1)F_q(z) = 1 - zF_{q-1}(z), \quad (14)$$

for $q\neq 1$.[19] The more general recursion relation for $\mathscr{F}_q(z,a)$ may be obtained by considering the following integral:

$$e^{z-2a}\int_1^\infty d(1-it)(1-it)^{-q} \times\exp\left[(a-z)(1-it) + \frac{a}{1-it}\right] \times\left(a-z-\frac{q}{(1-it)}-\frac{a}{(1-it)^2}\right), \quad (15)$$

for $\mathrm{Im}(a-z)>0$. This integral can be evaluated directly and also by comparison with (12). Equating the two results gives

$$a\mathscr{F}_{q+2}(z,a) = 1 + (a-z)\mathscr{F}_q(z,a) - q\mathscr{F}_{q+1}(z,a), \quad (16)$$

which contains (14) as a special case. As usual, analytic continuation is used to extend this relation to the entire complex plane. Krivenski and Orefice[20] derived (16) by less direct means. [We note in passing that the method used here easily can be employed to generate an infinite hierarchy of relations similar to (16).]

The following two identities are easily proved using (14) and mathematical induction:

$$qF_{q+1,m+1}(z) = -[(m+1)F_{q,m}(z) + zF_{q,m+1}(z)], \quad (17)$$

$$\frac{d^n}{dz^n}[e^{-z}F_q(z)] = (-1)^ne^{-z}F_{q-n}(z), \quad (18)$$

where m is a non-negative integer. Equation (17) is a corrected version of a result obtained by Imre and Weitzner.[12]

E. Differential equations, relations to other functions

Equations (6) and (7) with $r = 1$ and $a = 0$ imply

$$\frac{dF_q(z)}{dz} = F_q(z) - F_{q-1}(z). \tag{19}$$

The following first-order linear differential equation satisfied by $F_q(z)$ then follows from (14) and (19):

$$0 = \frac{dF_q(z)}{dz} + \left(\frac{1-q}{z} - 1\right)F_q(z) + \frac{1}{z}. \tag{20}$$

Equation (20) is easily integrated to give

$$F_q(z) = z^{q-1}e^z \int_z^\infty du\, u^{-q}e^{-u}$$

$$= z^{q-1}e^z\Gamma(1-q, z) \tag{21}$$

$$= \Gamma(1-q)e^z[z^{q-1} - \gamma^*(1-q, z)], \tag{22}$$

where $\Gamma(1-q, z)$ and $\gamma^*(1-q, z)$ are the usual incomplete gamma functions (Ref. 25, Eqs. 6.5.2 and 6.5.4). Thus $F_q(z)$ is singular at the origin if $q \leqslant 1$. This point is a branch point of $F_q(z)$ (which is then multivalued) unless $q = 0$, -1, -2,... (Ref. 26, Eq. 8.351.3). An alternative form of (21) is

$$F_q(z) = \begin{cases} e^z E_q(z), & q \geqslant 0, \\ e^z \alpha_{-q}(z), & q < 0, \end{cases} \tag{23}$$

where E_q and α_{-q} are exponential integral functions defined by

$$\int_1^\infty du\, u^{-q}e^{-zu} = \begin{cases} E_q(z), & q \geqslant 0, \\ \alpha_{-q}(z), & q < 0, \end{cases} \tag{24}$$

for $\mathrm{Re}(z) > 0$ with appropriate analytic continuation for $\mathrm{Re}(z) \leqslant 0$.

The relation of $F_q(z)$ to the confluent hypergeometric function is (Ref. 25, Eqs. 13.6.28 and 13.1.29)

$$F_q(z) = U(1, 2-q, z) = z^{q-1}U(q, q, z) \tag{25a}$$

and hence (Ref. 25, Eq. 13.4.22)

$$F_{q,r}(z) = (-1)^r\Gamma(r+1)U(1+r, 2-q+r, z). \tag{25b}$$

The analytic continuation of the confluent hypergeometric function is given by Ref. 25 (Eqs. 13.1.9 and 13.1.10).

Together Eqs. (10), (23), and (24) imply the following integral identity if $\mathrm{Re}(z-a) > 0$:

$$\mathscr{F}_q(z, a) = e^{z-2a}\int_1^\infty dt\, t^{-q}\exp\left[(a-z)t + \frac{a}{t}\right]. \tag{26}$$

A similar result for $\mathrm{Im}(z-a) > 0$ may be obtained from (12) by a change of variable. Equation (26) may also be used to derive the recursion relation (16) for $\mathscr{F}_q(z, a)$.

Equations (6) and (7), with $r = 1, 2$, yield

$$\frac{\partial \mathscr{F}_q}{\partial z} = \mathscr{F}_q - \mathscr{F}_{q-1},$$

$$\frac{\partial^2 \mathscr{F}_q}{\partial z^2} = \mathscr{F}_q - 2\mathscr{F}_{q-1} + \mathscr{F}_{q-2}, \tag{27}$$

where the arguments have been omitted. Eliminating $\mathscr{F}_{q-1}$ and $\mathscr{F}_{q-2}$ between these equations and the recursion relation (16) gives

$$0 = 1 - (z+q-2)\mathscr{F}_q - [2(a-z) - q + 2]\mathscr{F}_q' + (a-z)\mathscr{F}_q'', \tag{28}$$

where the primes denote partial differentiation with respect to z. Other partial differential equations satisfied by $\mathscr{F}_q(z, a)$ may be obtained by similar means.

F. Series expansions

Series representations of $F_q(z)$ follow from (21) and (22):

$$F_q(z) = z^{q-1}e^z\Gamma(1-q) - \sum_{j=0}^\infty \frac{z^j\Gamma(1-q)}{\Gamma(j+2-q)} \tag{29}$$

$$= z^{q-1}e^z\Gamma(1-q) - e^z\sum_{j=0}^\infty \frac{(-z)^j}{\Gamma(j+q-1)j!} \tag{30}$$

(Ref. 25, Eq. 6.5.29). A further identity, which may be of use in applications of these functions to the quantum electrodynamical calculations mentioned in the Introduction, follows if z is real and positive and lies on the principal branch of $F_q(z)$ (Ref. 26, Eq. 8.354.5):

$$F_q(z) = \sum_{j=0}^\infty \frac{L_j^{(1-q)}(z)}{j+1}, \tag{31}$$

where $L_j^{(1-q)}$ is a generalized Laguerre polynomial (Ref. 25, Eqs. 22.5.16 and 22.5.17).

G. Asymptotic forms, continued fraction

The relations (21) and (23) between $F_q(z)$, the incomplete gamma function, and the exponential integral function lead immediately to the following asymptotic forms, which are valid for $|z| \gg 1$ provided $|\arg(z)| < 3\pi/2$ (Ref. 25, Eq. 6.5.32):

$$\Gamma(q)F_q(z) \sim \sum_{j=0}^\infty (-1)^j z^{-1-j}\Gamma(q+j), \tag{32}$$

$$\Gamma(q)F_{q,r}(z) \sim \sum_{j=0}^\infty (-1)^{j+r}(j+1)_r\, z^{-1-r-j}\Gamma(q+j), \tag{33}$$

with

$$(j+1)_r = \begin{cases} (j+1)(j+2)\cdots(j+r), & r \geqslant 1, \\ 1, & r = 0. \end{cases}$$

A further asymptotic form is given in Eq. (49). These forms are essential when calculating $F_q(z)$ for $|z| \gg 1$ in order to avoid subtraction errors in the relations (7) and (14).

The following continued fraction representation of $F_q(z)$ follows immediately from Ref. 25 (Eq. 5.1.22) or Ref. 26 (Eq. 8.358):

$$F_q(z) = \frac{1}{z+}\,\frac{q}{1+}\,\frac{1}{z+}\,\frac{q+1}{1+}\,\frac{2}{z+}\,\frac{q+2}{1+\cdots}, \tag{34}$$

provided $|\arg(z)| < \pi$.

Imre and Weitzner[12] obtained the following asymptotic form for $\mathscr{F}_q(z, a)$:

$$\mathscr{F}_q(z,a) = -\sum_{j=0}^{\infty} \frac{C_{q,j}}{(a-z)^{j+1}}, \quad (35)$$

with

$$C_{q,0} = 1, \quad (36a)$$

$$C_{q,j} = (j+q-1)C_{q,j-1} + aC_{q+1,j-1}. \quad (36b)$$

Although not noted by these authors, (35) is valid only for $|\arg(z)| < 3\pi/2$. This asymptotic form is required when the recursion relation (16) or the sum rule (7) suffers from subtraction errors.

Having obtained asymptotic expansions of $F_q(z)$ and $\mathscr{F}_q(z,a)$ for large $|z|$ and $|a|$ it remains to consider the case of large q. Maroli and Petrillo[23] obtained the following expressions for $q>0$ in terms of the usual PDF Z and its derivatives[13]:

$$F_q(z) = -\frac{Z(\eta)}{(2q)^{1/2}} - \frac{Z^{(3)}(\eta)}{12q} - \frac{Z^{(4)}(\eta)}{16q(2q)^{1/2}} - \frac{Z^{(6)}(\eta)}{144q(2q)^{1/2}} + O(q^{-2}), \quad (37)$$

with $Z^{(n)}(\eta) = [d^nZ/dz^n]_{z=\eta}$, $\eta = (z+q)/(2q)^{1/2}$, and

$$\mathscr{F}_q(z,a) = -\frac{Z(\psi)}{(4a+2q)^{1/2}} - \frac{3a+q}{3(4a+2q)^2} Z^{(3)}(\psi) - \frac{4a+q}{4(4a+2q)^{5/2}} Z^{(4)}(\psi) - \frac{(3a+q)^2}{18(4a+2q)^{7/2}} Z^{(6)}(\psi) + O(q^{-2}, a^{-2}), \quad (38)$$

with $\psi = (z+q)/(4a+2q)^{1/2}$. Equations (37) and (38) are valid on the principal branches of $F_q(z)$ and $\mathscr{F}_q(z,a)$ respectively, if z and a are real.

H. Approximation

If z lies on the positive real axis in the principal branch of $F_q(z)$ with $q>0$ then Ref. 25, Eq. 5.1.52, gives

$$F_q(z) = \frac{1}{z+q}\left(1 + \frac{q}{(z+q)^2} + \frac{q(q-2z)}{(z+q)^4} + \frac{q(6z^2-8qz+q^2)}{(z+q)^6} + R(q,z)\right), \quad (39)$$

with $|R(q,z)| \leqslant q^{-4}$. A simpler, related approximation is that used by Robinson,[24] which was of the form $F_q(z) \simeq (z+q-1)^{-1}$ for $q>1$ and $z \geqslant 0$.

I. Special cases

In a number of special cases the form of $\mathscr{F}_q(z,a)$ simplifies considerably.

(i) If $a=0$, (3) gives $\mathscr{F}_q(z,a) = F_q(z)$.

(ii) If $z=a=0$ and $q>1$, then (21) yields

$$\mathscr{F}_q(0,0) = F_q(0) = (q-1)^{-1}. \quad (40)$$

Equation (14) then gives

$$F_0(z) = 1/z. \quad (41)$$

(iii) If $q = \frac{1}{2}, \frac{3}{2}, \ldots$ and z is real, the imaginary part of $F_q(z)$ (on the principal branch) can be obtained from (27) to give[17]

$$\mathrm{Im}[F_q(z)] = \begin{cases} 0, & 0 \leqslant z, \\ -\pi(-z)^{q-1}e^z/\Gamma(q), & z<0. \end{cases} \quad (42)$$

Substitution of (43) into the sum (10) then implies[17]

$$\mathrm{Im}[\mathscr{F}_q(z,a)] = \begin{cases} 0, & a \leqslant z, \\ -\pi e^{z-2a}[(a-z)/a]^{(q-1)/2} \times I_{q-1}[2a^{1/2}(a-z)^{1/2}], & z<a, \end{cases} \quad (43)$$

provided z and a are real and $z-a$ lies on the principal branch of F_q.

(iv) If $z=a$ we find that the change of variable $t=1/u$ in (26) gives

$$\mathscr{F}_q(a,a) = e^{-a}\int_0^1 du\, u^{q-2}e^{au}$$

$$= e^{-a}(-a)^{1-q}\gamma(q-1,-a) \quad (44)$$

$$= e^{-a}(-a)^{1-q}\Gamma(q-1) - F_{-q}(-a), \quad (45)$$

and hence in particular,

$$\mathscr{F}_{3/2}(a,a) = e^{-a}(-a)^{-1/2}\pi^{1/2}\,\mathrm{erf}[(-a)^{1/2}].$$

Note that $\mathscr{F}_{3/2}(a,a)$ is double valued owing to there being two possible choices for $(-a)^{1/2}$.

(v) If $q=-r$ then (25b) and Eq. 13.6.24 of Ref. 25 imply

$$F_{-r,r}(z) = (-1)^r\Gamma(r+1) \times \pi^{-1/2}e^{z/2}z^{-r-1/2}K_{r+1/2}(\tfrac{1}{2}z), \quad (46)$$

where $K_{r+1/2}$ is a modified spherical Bessel function.

III. PDF'S OF HALF-INTEGER AND INTEGER INDEX

Those cases in which q is either a half-integer or an integer are the ones of most interest in plasma physics and quantum electrodynamics. A number of additional properties of the PDF's F_q and $\mathscr{F}_q$ are known in these special cases. In this section we present these additional properties together with a number of simplified forms of the expressions in Sec. II valid for integer and half-integer q. We also discuss the analytic properties of $F_q(z)$ in more detail in these special cases.

A. Half-integer q

Shkarofsky[17] obtained the following expression for $F_q(z)$ for positive half-integer q in terms of the PDF Z (see Ref. 13), a number of whose properties are summarized in the Appendix:

$$\Gamma(q)F_q(z) = \sum_{j=0}^{q-3/2}(-z)^j\Gamma(q-1-j) + \pi^{1/2}(-z)^{q-3/2} \times [iz^{1/2}Z(iz^{1/2})] \quad (47a)$$

$$= \sum_{j=0}^{q-3/2}(-z)^j\Gamma(q-1-j) - \pi(-z)^{q-3/2} \times z^{1/2}e^z\,\mathrm{erfc}(z^{1/2}). \quad (47b)$$

Equation (47a) determines $F_q(z)$ throughout the complex plane and leads to the series representation[17]

$$\Gamma(q)F_q(z) = \sum_{j=0}^{\infty} (-z)^j \Gamma(q-1-j) - i\pi(-z)^{q-1}e^z, \tag{48}$$

and to the asymptotic expression[17]

$$\Gamma(q)F_q(z) \sim -\sum_{j=0}^{\infty} \Gamma(q+j)(-z)^{-1-j} - i\sigma\pi(-z)^{q-1}e^z, \tag{49}$$

with

$$\sigma = \begin{cases} 0, & |\arg(z)| < \pi, \\ 1, & |\arg(z)| = \pi, \\ 2, & \pi < |\arg(z)| < 2\pi. \end{cases}$$

The case of negative half-integer q can be treated using these results and (14).

Analytic properties: Krivenski and Orefice[20] derived the following expressions for $\mathscr{F}_{1/2}$ and $\mathscr{F}_{3/2}$ in terms of Z:

$$\mathscr{F}_{1/2}(z,a) = -iZ^{+}/(z-a)^{1/2}, \tag{50a}$$

$$\mathscr{F}_{3/2}(z,a) = -Z^{-}/a^{1/2}, \tag{50b}$$

with

$$Z^{\pm} = \tfrac{1}{2}[Z\{a^{1/2} + i(z-a)^{1/2}\} \pm Z\{-a^{1/2} + i(z-a)^{1/2}\}]. \tag{51}$$

In practice only one root $a^{1/2}$ is relevant in Eqs. (50) and (51) since the other root leads to identical expressions for $\mathscr{F}_{1/2}(z,a)$ and $\mathscr{F}_{3/2}(z,a)$. In general, however, both roots $(z-a)^{1/2}$ must be considered, implying that $\mathscr{F}_q(z,a)$ is double valued with a branch point at $z=a$ [or, equivalently, that $\mathscr{F}_q(z,a)$ is defined on the same two-sheeted Riemann surface as $(z-a)^{1/2}$]. Equations (16), (50), and (51) together enable the functions $\mathscr{F}_q$ of half-integer index to be reexpressed in terms of the PDF Z in a similar manner to $F_q(z)$ [(47a)].

The expressions (51) and (A3) imply[20]

$$\mathscr{F}_q(z,a) \simeq -\tfrac{1}{2}a^{-1/2}Z(\tfrac{1}{2}za^{-1/2}), \tag{52}$$

provided $1 \ll |a|$, $|z| \ll |a|$, and $-5\pi/4 < \arg(a^{1/2}) < \pi/4$. The restriction on $\arg(a^{1/2})$ in (52) was first noted by Robinson.[18]

The double valuedness of $F_q(z)$ implies $F_q(z^*) \neq [F_q(z)]^*$ in general, where the asterisk denotes complex conjugation. However, we can write

$$F_q(w^{*2}) = [F_q(w^2)]^* \tag{53}$$

for half-integer q. A modulus-argument diagram of $F_{5/2}(w^2)$ is shown in Fig. 1 for $|w|<2.8$, $\mathrm{Im}(w)>0$ as an illustration of the qualitative behavior of $F_q(z)$ for $q=\frac{3}{2}, \frac{5}{2}, \ldots$ [by contrast $F_{1/2}(z)$ has a singularity at the origin]. The asymptotic behavior (32) holds in the region $|\arg(w)| < 3\pi/4$ while $F_{5/2}(w^2)$ diverges exponentially for large $|w|$ if $3\pi/4 < |\arg(w)| < \pi$. Complicated behavior persists for large values of $|w|$ when $|\arg(w)| \simeq 3\pi/4$; we have not investigated this behavior in detail. Figures illustrating other aspects of the behavior of $F_q(z)$ and $\mathscr{F}_q(z,a)$ have been published by Maroli and Petrillo,[23] Airoldi and Orefice,[19] Krivenski and Orefice,[20] and Bornatici *et al.*[2]

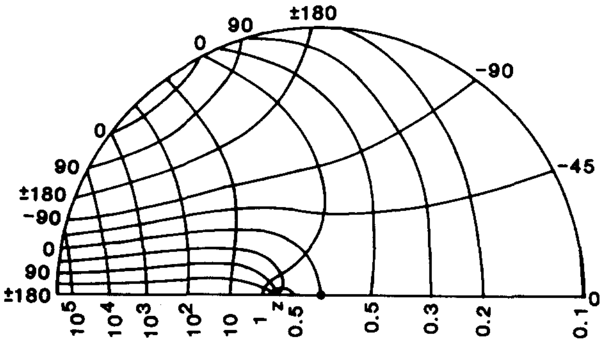

FIG. 1. Modulus-argument diagram of $F_{5/2}(w^2)$ for $\mathrm{Im}(w)>0$ and $|w|<2.8$. Contours of constant modulus and constant argument (in degrees) are as labeled along the real axis and the semicircular boundary, respectively. The origin is denoted by a large dot. The zero of $F_{5/2}(w^2)$ on the real axis is indicated by the letter z.

Choice of branch: Any line passing through the origin in Fig. 1 divides the w plane into two half-planes, each of whose image under the mapping $w \to w^2$ is the entire complex plane. Hence there is an infinite set of possible pairs of branches for F_q if q is a half-integer. The two most symmetric locations for the line of separation in the w plane are along either the real axis or the imaginary axis. These choices, however, correspond to branch cuts along the positive real axis and negative real axis, respectively, in the z plane (with $z=w^2$); they are thus inconvenient for calculational purposes if $F_q(z)$ is to be investigated near the real z axis, as is usually the case. A more convenient but less symmetric choice moves the branch cut to the negative imaginary axis in the z plane. This choice corresponds to the line of separation $\mathrm{Im}(w) = -\mathrm{Re}(w)$ in the w plane.

B. Integer q

The functions $\mathscr{F}_q(z,a)$, $F_q(z)$ of integer index are of interest in studies of the dispersion of Bernstein waves,[9,10] the dielectric properties of two-dimensional thermal plasmas,[27] and quantum electrodynamics.[14,15]

Explicit expressions for $F_q(z)$ for integer q are

$$\Gamma(q)F_q(z) = (-1)^{q-1}\left[z^{q-1}e^zE_1(z) - \sum_{j=0}^{q-2}(-1)^j j! z^{q-j-2}\right], \tag{54}$$

for $q = 1, 2, \ldots$ with

$$E_1(z) = -\gamma - \ln z - \sum_{j=1}^{\infty} \frac{(-z)^j}{j!j}, \tag{55}$$

where $\gamma = 0.5772\ldots$ is Euler's constant, and

$$F_q(z) = \Gamma(1-q)z^{q-1}\sum_{j=0}^{-q}\frac{z^j}{j!}, \tag{56}$$

for $q = 0, -1, -2, \ldots$ (Ref. 25, Eq. 5.1.8).

Analytic properties: If $q = 1, 2, \ldots$ then the Riemann surface for F_q has a countably infinite number of sheets due to the logarithmic contribution to $E_1(z)$ in (55); a branch point exists at the origin. The location of the principal branch is arbitrary except that it is usually chosen to include those values of z satisfying $\arg(z) = \pi$ but not those satisfying $\arg(z) = -\pi$. If $q = 0, -1, -2, \ldots$ then $F_q(z)$ is single valued, as seen explicitly from (56). The symmetry property of $E_n(z)$ (Ref. 25, Eq. 5.1.13), the explicit expression (56), and (10) together imply

$$F_q(z^*) = [F_q(z)]^*, \tag{57a}$$

$$\mathcal{F}_q(z^*, a^*) = [\mathcal{F}_q(z, a)]^*, \tag{57b}$$

for any integer q.

IV. PDF'S FOR WAVES WITH ARBITRARY PERPENDICULAR WAVE NUMBER

The dielectric tensor of a weakly relativistic thermal plasma can be written in terms of integrals of the following form[2,28] for waves with arbitrary perpendicular wave number:

$$\Psi = -i\int_0^\infty dt\,\frac{(it)^m}{(1-it)^l}\exp\left[izt - \Lambda(1-\cos at) - \frac{at^2}{1-it}\right], \tag{58}$$

with

$$\Lambda = \lambda/(1-it), \tag{59}$$

with $m = 0, 1, 2,\ l = \frac{3}{2}, \frac{5}{2}, \ldots,$ and where λ and a are complex constants. These integrals may be reexpressed using the identity

$$e^{\Lambda\cos\vartheta} = \sum_{j=-\infty}^{\infty} e^{-ij\vartheta} I_j(\Lambda) \tag{60}$$

to give

$$\Psi = \sum_{j=-\infty}^{\infty} \mathcal{R}_{l,m}(z - aj, a, \lambda, j), \tag{61}$$

with

$$\mathcal{R}_{l,m}(z,a,\lambda,j) = -i\int_0^\infty dt\,\frac{(it)^m}{(1-it)^l}\exp\left[izt - \frac{at^2}{1-it}\right]e^{-\Lambda}I_j(\Lambda). \tag{62}$$

We make the following definitions, which are analogous to (3), (4a), and (4b):

$$R_{l,m}(z,\lambda,s) = \mathcal{R}_{l,m}(z,0,\lambda,s), \tag{63a}$$

$$\mathcal{R}_l(z,a,\lambda,s) = \mathcal{R}_{l,0}(z,a,\lambda,s), \tag{63b}$$

$$R_l(z,\lambda,s) = R_{l,0}(z,\lambda,s). \tag{63c}$$

A. Sums, recursion relations, series

Equations (63a) and (63b) lead to

$$\mathcal{R}_{l,m}(z,a,\lambda,s) = \sum_{j=0}^{m}(-1)^j\binom{m}{j}R_{l-j}(z,a,\lambda,s), \tag{64}$$

which is analogous to (7). Analogs of (8)–(10) also are derived easily.

The Bessel functions satisfy the identity

$$I_s(z) = (z/2s)[I_{s-1}(z) - I_{s+1}(z)], \tag{65}$$

which implies the following recursion relation:

$$\mathcal{R}_{l,m}(z,a,\lambda,s) = (\lambda/2s)[\mathcal{R}_{l+1,m}(z,a,\lambda,s-1) - \mathcal{R}_{l+1,m}(z,a,\lambda,s+1)]. \tag{66}$$

The ascending series for $I_s(z)$ is

$$I_s(z) = \sum_{j=0}^{\infty}\frac{(\frac{1}{2}z)^{2j+s}}{j!\Gamma(s+j+1)}. \tag{67}$$

Substitution of this series into (62) leads to the identity

$$\mathcal{R}_l(z,a,\lambda,s) = \sum_{j=0}^{\infty}\left(\frac{\lambda}{2}\right)^{2j+s}\times\frac{e^{-\lambda}}{j!\Gamma(s+j+1)}\mathcal{F}_{l+s+2j}(z-\lambda, a-\lambda), \tag{68}$$

which in turn implies

$$\mathcal{R}_l(z,a,\lambda,s) = (\tfrac{1}{2}\lambda)^s[1/\Gamma(s+1)]\mathcal{F}_{l+s}(z,a),$$

for $|\lambda| \ll 1$. A further series expansion is given by Eq. (71).

B. Integral representations

Use of the series (68) is cumbersome in general and may result in subtraction errors for some combinations of parameters. To avoid such problems it is desirable to have available integral representations of $\mathcal{R}_l(z,a,\lambda,s)$. We derive such forms here.

The quantity $e^{-\Lambda}I_s(\Lambda)$ may be replaced in (62) using

$$e^{-\Lambda}I_s(\Lambda) = 2\int_0^\infty dx\,x\exp[-x^2(1-it)]\times J_s^2[(2\lambda)^{1/2}x](1-it) \tag{69}$$

for $s > -1$ (Ref. 26, Eq. 6.633.2). Upon reversing the order of the resulting integrals this gives

$$\mathcal{R}_l(z,a,\lambda,s) = 2\int_0^\infty dx\,x\exp(-x^2)\times J_s^2[(2\lambda)^{1/2}x]\mathcal{F}_{l-1}(z+x^2, a). \tag{70}$$

If the square of the Bessel function in Eq. (70) is expanded in powers of λ we obtain

$$\mathcal{R}_l(z,a,\lambda,s) = \sum_{j=0}^{\infty}(-1)^j\left(\frac{\lambda}{2}\right)^{s+j}\times\frac{\Gamma(2s+2j+1)}{\Gamma(s+j+1)\Gamma(2s+j+1)j!}\times\mathcal{F}_{l+s+j}(z,a), \tag{71}$$

where we have anticipated Eq. (80) in simplifying this result. Use of the result

$$e^{-\Lambda}I_s(\Lambda) = 2\pi^{-1/2}\int_0^\infty dx\exp[-x^2(1-it)]\times J_{2s}[(8\lambda)^{1/2}x](1-it)^{1/2}, \tag{72}$$

for $s > -\frac{1}{2}$ (Ref. 26, Eq. 6.618.1) enables the following expression to be derived:

$$\mathscr{R}_l(z,a,\lambda,s) = 2\pi^{-1/2}\int_0^\infty dx\,\exp(-x^2)\times J_{2s}[(8\lambda)^{1/2}x]\mathscr{F}_{l-1/2}(z+x^2,a). \quad (73)$$

Computational advantages of (70) and (73) are (i) the factors $\exp(-x^2)$ in these equations restrict the effective ranges of integration to relatively small intervals in x provided $|\arg(z+x^2-a)|<3\pi/2$; (ii) if λ, z, and a have small imaginary parts (a case often considered in applications) the integrand in (70) does not oscillate rapidly in sign and so (70) is relatively straightforward to evaluate; and (iii) if λ is real, the real and imaginary parts of $\mathscr{R}_l$ separate as integrals of the real and imaginary parts of $\mathscr{F}_{l-1}$ or $\mathscr{F}_{l-1/2}$.

A further integral identity can be derived if the following integral representation for $I_s(\Lambda)$ with $s>-\frac{1}{2}$ (Ref. 25, Eq. 9.6.18) is substituted into (62):

$$I_s(\Lambda) = \frac{(\frac{1}{2}\Lambda)^s}{\pi^{1/2}\Gamma(s+\frac{1}{2})}\int_{-1}^{1} du(1-u^2)^{s-1/2}e^{-\Lambda u}. \quad (74)$$

Using (2), this relation yields the representation

$$\mathscr{R}_l(z,a,\lambda,s) = \frac{(\frac{1}{2}\lambda)^s}{\pi^{1/2}\Gamma(s+\frac{1}{2})}\int_{-1}^{1} du(1-u^2)^{s-1/2}e^{-\lambda(1+u)}\times\mathscr{F}_{l+s}(z-\lambda-\lambda u, a-\lambda-\lambda u). \quad (75)$$

Equation (75) has the particular advantage that the range of integration is finite.

Note that the restrictions $s>-1$ in (69) and $s>-\frac{1}{2}$ in (72) and (74) do not diminish the usefulness of (70), (73), and (75) in evaluating $\mathscr{R}_l(z,a,\lambda,j)$ for negative integer j since then $I_j(z)=I_{-j}(z)$ and $J_j(z)=(-1)^jJ_{-j}(z)$.

C. Limiting cases

The classical quantity corresponding to $\mathscr{R}_l$ (see Refs. 2 and 29) is reproduced if $(1-it)$ is replaced by unity wherever it occurs explicitly in Eqs. (59) and (62):

$$\mathscr{R}_l(z,a,\lambda,s)\simeq -\tfrac{1}{2}a^{-1/2}e^{-\lambda}I_s(\lambda)Z(\tfrac{1}{2}za^{-1/2}). \quad (76)$$

The approximation made to $\mathscr{R}_l$ by many authors[30–32] when $a=0$ is recovered from (70) in the limit $|z|\gg 1$, $|a|$ if $|\arg(z)|<3\pi/2$:

$$\mathscr{R}_l(z,a,\lambda,s)\simeq e^{-\lambda}I_s(\lambda)/z. \quad (77)$$

This result may also be obtained by replacing the factor $(1-it)$ by unity in (59) and (62).

If we substitute the asymptotic form $I_s(\Lambda)$ for $|\arg(\Lambda)|<\pi/2$ (Ref. 25, Eq. 9.7.1),

$$I_s(\Lambda)\simeq(2\pi\Lambda)^{-1/2}e^{\Lambda}[1-(4s^2-1)/8\Lambda+\cdots], \quad (78)$$

into (62) we obtain the asymptotic series

$$\mathscr{R}_l(z,a,\lambda,s)\sim(2\pi\lambda)^{-1/2}[\mathscr{F}_{l-1/2}(z,a) - (4s^2-1)\mathscr{F}_{l-3/2}(z,a)/8\lambda+\cdots]. \quad (79)$$

This result generalizes those of Lazzaro and Orefice[33] and Airoldi-Crescentini *et al.*[9]; it requires $\lambda\gg s^2$ for its validity. Note, however, that the analytic properties of (79) are different from those which arise from an expansion of (62) in powers of λ (cf. Sec. III).

V. INTEGRALS INVOLVING PDF'S

If we assume $|\lambda|\ll 1$, then each of the Bessel functions in (62), (70), and (73) may be approximated by the first term in its ascending series:

$$J_s(u)\simeq(\tfrac{1}{2}u)^s/\Gamma(s+1)\simeq I_s(u).$$

If we evaluate these three equations using this approximation and equate the results we obtain

$$\mathscr{F}_{r+s}(z,a) = \frac{2}{\Gamma(s+1)}\int_0^\infty dx\,x^{2s+1}\exp(-x^2)\times\mathscr{F}_{r-1}(z+x^2,a), \quad r\geqslant 1, \quad (80)$$

$$= \frac{2^{2s+1}\Gamma(s+1)}{\pi^{1/2}\Gamma(2s+1)}\int_0^\infty dx\,x^{2s}\exp(-x^2)\times\mathscr{F}_{r-1/2}(z+x^2,a), \quad r\geqslant\tfrac{1}{2}. \quad (81)$$

Setting $r=q+1$ and $s=-\frac{1}{2}$ in (80) or $r=q+\frac{1}{2}$ and $s=0$ in (81) we find

$$\mathscr{F}_{q+1/2}(z,a) = 2\pi^{-1/2}\int_0^\infty dx\,\exp(-x^2)\mathscr{F}_q(z+x^2,a). \quad (82)$$

This expression is a generalization of a result obtained by Airoldi and Orefice[19] in the case $a=0$. Setting $r=1$ and $s=q-1$ in (80) or $r=\frac{1}{2}$ and $s=q-\frac{1}{2}$ in (81) we find

$$\mathscr{F}_q(z,a) = \frac{2}{\Gamma(q)}\int_0^\infty dx\,x^{2q-1}\exp(-x^2)\times\mathscr{F}_0(z+x^2,a), \quad (83)$$

which remains valid in the limit $q\to 0$. If $a=0$ then (41) implies[19]

$$F_q(z) = \frac{1}{\Gamma(q)}\int_0^\infty du\,u^{q-1}\frac{e^{-u}}{u+z}. \quad (84)$$

Differentiation of (84) yields[12]

$$\Gamma(q)F_{q,r}(z) = (-1)^r r!\int_0^\infty du\,u^{q-1}e^{-u}(u+z)^{-r-1}. \quad (85)$$

Airoldi and Orefice[19] obtained the integral relation

$$\mathscr{F}_{q+1/2}(z,a) = \pi^{-1/2}\int_{-\infty}^\infty dx\,\exp(-x^2)\times F_q(z+x^2-2a^{1/2}x), \quad (86)$$

which provides a useful link between the Shkarofsky and Dnestrovskii functions. An equivalent relationship was obtained by Fidone *et al.*[1] in the special case $q=2$.

Further integrals: Upon setting $r=\frac{3}{2}$ and $a=0$ in (80) and replacing $F_{1/2}(z)$ using (50a) and (A3) we find

$$F_{s+3/2}(z) = \frac{2\pi^{1/2}}{\Gamma(s+1)}e^z\int_0^\infty dx\times x^{2s+1}(z+x^2)^{1/2}\,\mathrm{erfc}[(z+x^2)^{1/2}], \quad (87)$$

which generalizes Eq. 6.281 of Ref. 26.

Using $F_1(z)=e^zE_1(z)$ in (81) with $r=\frac{3}{2}$ yields

$$F_{s+3/2}(z) = \frac{2^{2s+1}\Gamma(s+1)e^z}{\pi^{1/2}\Gamma(2s+1)}\int_0^\infty dx\,x^{2s}E_1(z+x^2). \quad (88)$$

Further integrals may be obtained by reference to the sections on the incomplete gamma function, the exponential integral function, and the error function in Refs. 25 and 26.

VI. CONNECTIONS WITH PDF'S USED BY OTHER AUTHORS

In this section we make explicit the relation between the PDF's defined here and those introduced by a number of other authors. In several cases we modify the notations used by these other workers to make the parameter dependence of their functions explicit. We recall that the functions (1) used in this paper correspond to those employed in most discussions of electron cyclotron waves, including those by Dnestrovskii *et al.*,[16] Airoldi and Orefice,[19] Krivenski and Orefice,[20] and Robinson.[18,24] Moreover, the function $W_q(z, a)$ used by Maroli and Petrillo[23] and Bornatici *et al.*[2] is identical with $\mathscr{F}_q(z, a)$ as defined in (4a). The notations $\mathscr{F}_{q,r}(z, a)$ and $F_{q,r}(z)$ were introduced by Robinson[18]; $\mathscr{F}_{q,r}(z, a)$ corresponds closely to the H function used by Airoldi and Orefice.[19]

The original functions introduced by Shkarofsky[17] result from a series expansion of Trubnikov's[34] integral formulation of the plasma dielectric tensor; they are defined by

$$\mathscr{F}_q^{\text{Shkar}}(z, a, \alpha) = -i\int_0^\infty dt\,[\chi(z,a)]^{-q} \times \exp[z\{1-\chi(z,a)\} - iazt], \quad (89a)$$

where z, a, and α are constants and with

$$\chi(z, a) = [(1-it)^2 + 2at^2/z]^{1/2}. \quad (89b)$$

In the limit $|a| \ll 1$ $\mathscr{F}_q^{\text{Shkar}}(z, a, \alpha) = \mathscr{F}_q(z - z\alpha, a)$. De Barbieri[28] showed that a slightly different approach to the calculation of the weakly relativistic dielectric tensor yields plasma dispersion functions of the form (1); these appear to be the only functions to have been extensively employed in numerical calculations.

The functions $\overline{W}_q(z, a)$ were defined by De Barbieri[28] to approximate $\mathscr{F}_q(z, a)$:

$$\overline{W}_q(z, a) = -i\int_0^\infty dt(1-it)^{-q}\exp[izt - at^2]. \quad (90)$$

These functions satisfy many analogous relations to those discussed in this paper for $\mathscr{F}_q(z, a)$ but their analytic properties are somewhat different.

Imre and Weitzner[12] used functions closely related to $F_{q,r}$ and $\mathscr{F}_{q,r}$. In a modified notation their functions are defined by

$$F_q^{(r)}(z) = \frac{2r!}{\Gamma(q)}\int_0^\infty dw\, w^{2q-1} \times (w^2 - z^2)^{-r-1}\exp(-w^2), \quad (91a)$$

$$\mathscr{F}_q^{(r)}(z, \alpha) = \sum_{j=0}^{\infty}\frac{\alpha^{2j}}{j!}F_{q+j}^{(r+2j)}(z), \quad (91b)$$

where z and α are complex. We find

$$F_q^{(r)}(z) = (-1)^r F_{q,r}(-z^2), \quad (92a)$$

$$\mathscr{F}_q^{(r)}(z, \alpha) = (-1)^r \mathscr{F}_{q,r}(-z^2, \alpha^2). \quad (92b)$$

Imre and Weitzner[12] also introduced the more general PDF's defined (with a change of notation) by

$$Z_{n,m}(z, \alpha) = \pi^{-3/2}\int d^3\mathbf{x}\, x_\perp^n x_\parallel^m \frac{\exp(-x^2)}{z^2 - \frac{1}{2}x^2 + \alpha x_\parallel}, \quad (93)$$

where $\parallel$ and $\perp$ denote components of $\mathbf{x}$ parallel and perpendicular to an arbitrary axis in three-dimensional $\mathbf{x}$ space. These functions reproduce those introduced by Stix (Ref. 35, p. 177) and the usual plasma dispersion function[13] for appropriate choice of z, α, m, and n.

The function $\mathscr{R}_l(z, a, \lambda, s)$ was introduced by Lazzaro and Orefice[33] and Airoldi-Crescentini *et al.*[9] in the special case $l = \frac{5}{2}$.

VII. CONCLUSION

The properties of relativistic plasma dispersion functions relevant to the description of cyclotron waves in weakly relativistic, magnetized, thermal plasmas have been reviewed and a substantial number of new results have been obtained. We have obtained new sum rules, differential equations, and integral relations for the Dnestrovskii and Shkarofsky functions and have investigated in detail the analytic properties of these functions and their connection with standard transcendental functions such as the incomplete gamma function. A more general class of PDF's relevant to waves with arbitrary perpendicular wave number also has been defined and investigated for the first time and many properties of these functions have been obtained.

The results of this paper should be of considerable use in analytic work on electron cyclotron waves of all types and, as mentioned in the text, in quantum electrodynamics in strong magnetic fields. Numerical work will be facilitated by the variety of calculational tools presented (sums, recursion relations, continued fractions, integral forms, series, etc.). Comparison of existing papers also will be made easier by the results in Sec. VI, which relate many of the different PDF's currently to be found in the literature to those discussed in detail here.

ACKNOWLEDGMENT

The author thanks R. G. Hewitt for his comments on the manuscript of this paper.

APPENDIX: THE PDF *Z*

The PDF $Z(u)$ is defined as follows for $\text{Im}(u) > 0$ [with analytic continuation to $\text{Im}(u) \leqslant 0$][13]:

$$Z(u) = \pi^{-1/2}\int_{-\infty}^{\infty} dt\frac{\exp(-t^2)}{t-u} \quad (\text{A1a})$$

$$= 2\pi^{-1/2}u\int_0^\infty dt\frac{\exp(-t^2)}{t^2-u^2}. \quad (\text{A1b})$$

An alternative definition valid for all u is

$$Z(u) = 2i\exp(-u^2)\int_{-\infty}^{iu} dt\exp(-t^2), \quad (\text{A2})$$

which yields[13]

$$Z(u) = i\pi^{1/2}\exp(-u^2)[1 + \text{erf}(iu)]. \quad (\text{A3})$$

The differential equation satisfied by $Z(u)$ is[13]

$$\frac{dZ(u)}{du} = -2[1 + uZ(u)]. \quad \text{(A4)}$$

The following series and asymptotic series apply[13]:

$$Z(u) = i\pi^{1/2}\exp(-u^2) - u\sum_{j=0}^{\infty}\frac{(-u^2)^j\pi^{1/2}}{\Gamma(j+\frac{3}{2})}, \quad \text{(A5)}$$

$$Z(u) \sim i\pi^{1/2}\sigma\exp(-u^2) - \sum_{j=0}^{\infty}\frac{u^{-1-2j}\Gamma(j+\frac{1}{2})}{\pi^{1/2}},$$

$$|u| \gg 1 \quad \text{(A6)}$$

with

$$\sigma = \begin{cases} 0, & \mathrm{Im}(u) > 0, \\ 1, & \mathrm{Im}(u) = 0, \\ 2, & \mathrm{Im}(u) < 0. \end{cases}$$

Symmetry relations and other properties of $Z(u)$ are discussed by Fried and Conte.[13] The general behavior of $Z(u)$ for complex u may be seen from the modulus-argument plot of the function $w(u) = Z(u)/i\pi^{1/2}$ given in Ref. 25 (Fig. 7.3, p. 298).

[1] I. Fidone, G. Granata, and R. L. Meyer, Phys. Fluids **25**, 2249 (1982).
[2] M. Bornatici, R. Cano, O. De Barbieri, and F. Engelmann, Nucl. Fusion **23**, 1153 (1983).
[3] N. J. Fisch and A. H. Boozer, Phys. Rev. Lett. **45**, 720 (1980).
[4] R. F. Gandy, I. H. Hutchinson, and D. H. Yates, Phys. Rev. Lett. **54**, 800 (1985).
[5] C. S. Liu, Y. C. Mok, K. Papadopoulos, F. Engelmann, and M. Bornatici, Phys. Rev. Lett. **39**, 701 (1977).
[6] C. S. Wu and L. C. Lee, Astrophys. J. **230**, 621 (1979).
[7] D. B. Melrose, K. G. Rönnmark, and R. G. Hewitt, J. Geophys. Res. **87**, 5140 (1982).
[8] D. B. Melrose and G. A. Dulk, Astrophys. J. **259**, 844 (1982).
[9] A. Airoldi-Crescentini, E. Lazzaro, and A. Orefice, *Proceedings of the 2nd Joint Grenoble–Varenna Symposium on Heating in Toroidal Plasmas* (C.E.C., Brussels, 1980), p. 225.
[10] M. Bornatici, C. Maroli, and V. Petrillo, *Proceedings of the 3rd Joint Varenna–Grenoble Symposium on Heating in Toroidal Plasmas* (C.E.C., Brussels, 1982), p. 691.
[11] K. Imre and H. Weitzner, Phys. Fluids **28**, 133 (1985).
[12] K. Imre and H. Weitzner, Phys. Fluids **28**, 1757 (1985).
[13] B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).
[14] M. C. Allen, D. B. Melrose, and A. J. Parle, Proc. Astron. Soc. Aust. **6**, 211 (1985).
[15] M. C. Storey, "Collisions in magnetic fields," submitted to Aust. J. Phys.
[16] Y. N. Dnestrovskii, D. P. Kostomarov, and N. V. Skrydlov, Sov. Phys. Tech. Phys. **8**, 691 (1964).
[17] I. P. Shkarofsky, Phys. Fluids **9**, 561 (1966).
[18] P. A. Robinson, "Dielectric tensor of weakly relativistic electron distributions separable in momentum and pitch angle," Aust. J. Phys., in press.
[19] A. C. Airoldi and A. Orefice, J. Plasma Phys. **27**, 515 (1982).
[20] V. Krivenski and A. Orefice, J. Plasma Phys. **30**, 125 (1983).
[21] H. K. Wong, C. S. Wu, F. J. Ke, R. S. Schneider, and L. F. Ziebell, J. Plasma Phys. **28**, 503 (1982).
[22] C. S. Wu, C. S. Lin, H. K. Wong, S. T. Tsai, and R. L. Zhou, Phys. Fluids **24**, 2191 (1981).
[23] C. Maroli and F. Petrillo, Phys. Scripta **24**, 955 (1981).
[24] P. A. Robinson, "Electron cyclotron waves: Dispersion and accessibility conditions in isotropic and anisotropic plasmas," J. Plasma Phys., in press.
[25] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[26] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980).
[27] B. A. Trubnikov and V. B. Yakubov, Plasma Phys. **5**, 7 (1963).
[28] O. De Barbieri, Euratom-CEA, Grenoble, Report EUR-CEA-FC-1035, 1980.
[29] J. A. Tataronis and F. W. Crawford, J. Plasma Phys. **4**, 249 (1970).
[30] I. B. Bernstein, Phys. Rev. **109**, 10 (1958).
[31] J. A. Tataronis and F. W. Crawford, J. Plasma Phys. **4**, 231 (1970).
[32] S. Puri, F. Leuterer, and M. Tutter, J. Plasma Phys. **14**, 169 (1975).
[33] E. Lazzaro and A. Orefice, Phys. Fluids **23**, 2330 (1980).
[34] B. A. Trubnikov, Ph.D. thesis, Moscow University, 1958 (English translation, USAEC Tech. Information Service No. AEC-tr-4073).
[35] T. H. Stix, *The Theory of Plasma Waves* (McGraw–Hill, New York, 1962).

On the integrability of systems of nonlinear ordinary differential equations with superposition principles

T. C. Bountis and V. Papageorgiou
Department of Mathematics and Computer Science, Clarkson University, Potsdam, New York 13676

P. Winternitz
Centre de Recherches Mathématiques, Université de Montréal, C. P. 6128 Montréal, Quebec, Canada H3C 3J7

(Received 3 April 1985; accepted for publication 5 December 1985)

A new class of "solvable" nonlinear dynamical systems has been recently identified by the requirement that the ordinary differential equations (ODE's) describing each member of this class possess nonlinear superposition principles. These systems of ODE's are generally *not* derived from a Hamiltonian and are classified by associated pairs of Lie algebras of vector fields. In this paper, all such systems of $n\leqslant 3$ ODE's are integrated in a unified way by finding explicit integrals for them and relating them all to a "pivotal" member of their class: the projective Riccati equations. Moreover, by perturbing two parametrically driven projective Riccati equations (thus making them nonsolvable in the above sense) evidence is discovered of chaotic behavior on the Poincaré surface of section—in the form of sensitive dependence on initial conditions—near a boundary separating bounded from unbounded motion.

I. INTRODUCTION

In recent years there has been great progress in the mathematical analysis of nonlinear dynamical systems of n first order ordinary differential equations (ODE's)

$$\frac{dx_i}{dt}\equiv\dot{x}_i=F_i(\mathbf{x},t),\quad i=1,2,...,n, \tag{1.1}$$

$\mathbf{x}\equiv(x_1,...,x_n)$, which are coupled in a nontrivial way. For example, it has been widely recognized that most (nonlinear) dynamical systems (1.1) are, in general, nonsolvable, by the known analytical methods, and possess classes of solutions, which depend extremely sensitively on the initial conditions, giving rise to regions of so-called chaotic behavior in phase space.[1–5] The presence of these chaotic regions is, in fact, a consequence of the nonintegrability of (1.1), i.e., of the nonexistence, in general, of n global, independent, analytic integrals of the motion.[1–5]

There are, however, many physically interesting integrable examples of systems (1.1), which have attracted the attention of many researchers: they correspond to special choices of the functions $F_i(\mathbf{x},t)$, and are of practical importance, in that minor deviations from these choices are not likely to bring about major changes in the overall behavior of the system. Recently, there has been considerable progress in the identification and analysis of completely integrable Hamiltonian systems,[5–9] where (1.1) are Hamiltonian's equations of motion, with $n=2N$, and (x_i,x_{i+N}) N canonically conjugate pairs.[3]

Our interest here is in the integration of a class of nonlinear systems (1.1), which are not derived from a Hamiltonian, and which have been identified by the requirement that they possess *nonlinear superposition principles*[10–13]; this means that their general solution can be expressed in terms of a finite number of particular solutions.

In this paper, we identify by their Lie algebras, all systems (1.1) with $n\leqslant 3$ having superposition principles, find simple integrals of motion for them, and obtain their general solution by relating them all to a "pivotal" member of their class: a system of n projective Riccati equations.[11] In some cases we were even able to do this for arbitrary values of n.

We thus completely integrate, in a unified way, all indecomposable systems of $n=3$ ODE's (1.1) with superposition principles, identified in the Appendix by the general classification of Ref. 13. It is worth noting that these systems could also have been integrated using the particular superposition rule valid in each case, or by embedding them all in higher-dimensional linear systems with time dependent coefficients.[11–13]

However, the new integration method presented here has several advantages of its own: it is simple, direct, and reveals a deep connection between all these systems and a single one of them—the projective Riccati equations. Moreover, it is reminiscent of a similar approach, recently introduced, to find exact integrals of two-degrees-of-freedom Hamiltonian systems.[9]

It was recently shown that all indecomposable systems of ODE's with (nonlinear) superposition principles are related to the transitive primitive action of a Lie group G on a homogeneous space G/G_0 (see Ref. 13). Thus, to each algebra–subalgebra pair $\{\mathscr{L},\mathscr{L}_0\}$, defining a transitive primitive Lie algebra, we can associate a family of ODE's, whose representative is unique up to a choice of coordinates on G/G_0, i.e., up to an arbitrary invertible change of dependent variables in the equations.

The Lie algebra $\mathscr{L}$, corresponding to the Lie group G, is the algebra of vector fields in Lie's theorem.[10,11] The subalgebra $\mathscr{L}_0$, corresponding to the isotropy group G_0 of the origin in G/G_0 consists of vector fields vanishing at the origin. The pair $\{\mathscr{L},\mathscr{L}_0\}$ is said to determine a transitive primitive Lie algebra if (i) $\mathscr{L}_0$ is a maximal subalgebra of $\mathscr{L}$ and (ii) $\mathscr{L}_0$ does not contain an ideal of $\mathscr{L}$.

The number of ODE's corresponding to a pair $\{\mathscr{L},\mathscr{L}_0\}$ is[12,13]

$$n = \dim \mathscr{L} - \dim \mathscr{L}_0. \tag{1.2}$$

All the families of transitive primitive Lie algebras and their corresponding systems of ODE's integrated in this paper are explicitly constructed in the Appendix starting from $\{\mathscr{L},\mathscr{L}_0\}$ pairs with $n \leqslant 3$ in (1.2). For $n = 3$, we find four types of indecomposable systems of real equations with superposition formulas: (a) projective Riccati equations (PRE's) associated with the Lie algebras $\mathscr{L} = \mathrm{sl}(4,\mathbb{R})$ and $\mathscr{L}_0 = \mathrm{aff}(3,\mathbb{R})$ (affine transformations of a real, three-dimensional vector space); (b) one type of conformal Riccati equations (CRE's) associated with the Lie algebra pair $\mathscr{L} = \mathrm{o}(4,1)$ and $\mathscr{L}_0 = \mathrm{sim}(3)$; (c) one type of CRE's associated with $\mathscr{L} = \mathrm{o}(3,2)$ and $\mathscr{L}_0 = \mathrm{sim}(2,1)$ [$\mathrm{sim}(p,q)$ is the similitude algebra of $p+q$-dimensional Minkowski space]; and (d) a system of three real equations with fourth-order polynomial nonlinearities associated with the Lie algebra $\mathscr{L} = \mathrm{su}(2,1)$ and a subalgebra $\mathscr{L}_0$ of affine transforms.

In addition to these four systems of equations we consider two more types, one related to the action of $\mathscr{L} = \mathrm{sp}(4,\mathbb{R})$ and the other to $\mathscr{L} = \mathrm{sp}(2,\mathbb{R}) \oplus \mathrm{sp}(2,\mathbb{R})$. These "symplectic Riccati equations" (SRE's) for $n = 3$ are equivalent to one of those discussed above, but their integration presents new features and the equations have independent physical interest. In any case, for $n > 3$, the SRE's are not equivalent to any other of the systems described here. Finally, we also consider the only two independent types of complex ODE's with $n = 3$, having superposition formulas: the $\mathrm{sl}(4,\mathbb{C})$ PRE's and the $\mathrm{o}(5,\mathbb{C})$ CRE's.

We have successfully applied our integration method to *all* indecomposable systems of ($n =$) 3 ODE's with superposition principles. Whether it is more generally applicable, for n arbitrary, is still an open question. So far, we have integrated, by "dimensional reduction" (see Sec. II), systems of n projective Riccati equations. We have reduced systems of n conformal Riccati equations to $n+1$ PRE's (Sec. III) and shown (see the Appendix and Sec. IV) that a system of $2(p-1)$ real "pseudounitary Riccati equations" [based on $\mathrm{SU}(p,1)$] has the form of a system of p complex PRE's with one additional constraint. Since the complexity (and number) of these systems increases significantly with increasing n, we have chosen to postpone the analysis of other $n > 3$ cases to a future publication.

In Sec. II we integrate a system of n real (or complex) projective Riccati equations by a successive "dimensional reduction" to a system of $(n-1)$, $(n-2)$, etc. PRE's, down to a single Riccati equation. To do this, all we need to know at the k th stage is one particular solution of the corresponding system of k PRE's.

In Secs. III and IV we reduce a system of n conformal Riccati equations, two systems of three symplectic Riccati equations and one system of pseudounitary ODE's to systems of projective Riccati's, to which the dimensional reduction of Sec. II can be applied. This is done in a unified way as follows: Consider the subgroup $G_L \subset G$ that acts linearly on the space G/G_0. Find its quadratic invariant $u(x_1,x_3,\dots,x_n)$ (if it exists) expressed in terms of the coordinates on G/G_0, and use it as an $(n+1)$st dependent variable. The resulting ODE's describe the time evolution of $x_i(t)$ and $u(t)$ in a consistent manner, and are cast in the form of $(n+1)$ projective Riccati equations.

Finally, in Sec. V, we discuss our results in connection with recent work on the integrability of non-Hamiltonian systems possessing the Painlevé property[14,15] (i.e., whose solutions have no movable singularities in the complex t plane other than poles).[16] In fact, all systems discussed in this paper do possess the Painlevé property, since they are linearizable, in a precise sense, in spaces of higher dimensions.[11–13]

To demonstrate, however, what can happen to the solutions of our systems, when the equations are slightly modified away from their integrable Painlevé-like form, we integrate numerically, in Sec. V, a set of two parametrically driven, perturbed projective Riccati equations. What we find is that in "sensitive regions" of phase space—e.g., near a boundary separating bounded from unbounded motion—the solutions appear to exhibit, indeed, a chaotic behavior. In particular, we observe an extremely sensitive dependence on the choice of initial conditions,[1–4] on a Poincaré surface of section.

II. INTEGRATION OF PROJECTIVE RICCATI EQUATIONS

Since they turn out to play a central role in the integration of all other systems of ODE's with superposition principles discussed in this paper, we integrate first a system of projective Riccati's (PRE's), written in its most general form as[11]

$$\dot{x}_\mu = a_\mu + \sum_{\nu=1}^{n} b_{\mu\nu} x_\nu + x_\mu \sum_{\nu=1}^{n} c_\nu x_\nu, \tag{2.1}$$

$\mu = 1,2,\dots,n$, where a_μ, $b_{\mu\nu}$, and c_ν are arbitrary functions of t.

It has already been shown[11] that the general solution of (2.1) can be expressed algebraically in terms of $n+2$ generically chosen particular solutions. Moreover, it is also known that, with the substitution $x_\mu = y_\mu/y_0$, $\mu = 1,2,\dots,n$, (2.1) is transformed into a system of $n+1$ linear equations. This transformation shows by itself that the PRE's are of Painlevé type, since the only possible movable singularities they can have are poles [zeros of $y_0(t)$]. For special choices of a_μ, $b_{\mu\nu}$, and c_ν, PRE's commonly arise in physical applications.[15]

In this section, we present an alternative approach to integrating (2.1) by "dimensional reduction," i.e., by using at each step one particular solution of k PRE's to reduce the system to $k-1$ equations of the same type. Thus, with n particular solutions (one for each system of n, $n-1,\dots,2,1$ PRE's), the general solution of (2.1) can be obtained by quadratures.

To see this, all we need to demonstrate here is the first step of this procedure, since all subsequent steps can be performed in exactly the same way: Denoting a particular solution of (2.1) by $x_\mu = \alpha_\mu(t)$, we introduce new variables y_μ in (2.1),

$$x_\mu \equiv y_\mu + \alpha_\mu(t), \quad \mu = 1,2,\dots,n, \tag{2.2}$$

in terms of which (2.1) becomes

$$\dot{y}_\mu = \sum_{\nu=1}^{n} B_{\mu\nu} y_\nu + y_\mu \sum_{\nu=1}^{n} c_\nu y_\nu, \quad \mu = 1,...,n, \tag{2.3a}$$

where

$$B_{\mu\nu} \equiv b_{\mu\nu} + \alpha_\mu c_\nu + \delta_{\mu\nu} \sum_{\lambda=1}^{n} c_\lambda \alpha_\lambda, \tag{2.3b}$$

and the inhomogeneous terms have dropped out, since $\alpha(t)$ is a particular solution of (2.1).

Multiply now the μth equation (2.3) by y_n and the nth equation by y_μ and subtract to obtain

$$\dot{y}_\mu y_n - \dot{y}_n y_\mu = y_n \sum_{\nu=1}^{n} B_{\mu\nu} y_\nu - y_\mu \sum_{\nu=1}^{n} B_{n\nu} y_\nu, \quad \mu = 1,2,...,n-1. \tag{2.4}$$

Dividing both sides of (2.4) by y_n^2 and introducing the new variables

$$z_\mu = y_\mu / y_n, \quad \mu = 1,2,...,n-1, \tag{2.5}$$

we find that (2.4) takes the form of $(n-1)$ PRE's:

$$\dot{z}_\mu = B_{\mu n} + \sum_{\nu=1}^{n-1} C_{\mu\nu} z_\nu - z_\mu \sum_{\nu=1}^{n-1} B_{n\nu} z_\nu, \quad \mu = 1,...,n-1, \tag{2.6}$$

with $C_{\mu\nu} \equiv B_{\mu\nu} - \delta_{\mu\nu} B_{nn}$.

This procedure can now be repeated $(n-1)$ times down to a single Riccati equation, which, after being made homogeneous by a particular solution, can be solved easily by quadratures, i.e., by a linear ODE of first order. The general solution is now obtained by going up the "ladder" of the PRE systems described above, solving by quadratures, at every step, a single homogeneous Riccati in the denominator variable of the transformation equations. At the last step, for example, knowing all the z_μ, $\mu = 1,...,n-1$, as functions of t and $(n-1)$ arbitrary constants, we express the y_μ in (2.5) in terms of y_n, substitute them in the $\mu = n$ equation (2.3) and solve the final Riccati

$$\dot{y}_n = y_n \left\{ \sum_{\nu=1}^{n-1} B_{n\nu} z_\nu + B_{nn} \right\} + y_n^2 \left\{ \sum_{\nu=1}^{n-1} C_\nu z_\nu + C_n \right\}, \tag{2.7}$$

by quadratures, to find the nth arbitrary constant, and hence the complete solution of the problem.

The integration method described above is especially efficient for systems of PRE's whose coefficients are independent of t. In that case, the particular solution $\alpha(t)$ of (2.1) can be chosen with all its components α_μ constants satisfying the algebraic PRE equations $a_\mu + \Sigma b_{\mu\nu}\alpha_\nu + \alpha_\mu \Sigma c_\nu \alpha_\nu = 0$. The coefficients in the dimensionally reduced PRE's (2.6) will also be constants and this will be continued all the way down to the single equation (2.7), which can be directly integrated.

III. INTEGRATION OF CONFORMAL RICCATI EQUATIONS

We now proceed to illustrate our general method of integrating systems of ODE's with superposition formulas, on the conformal Riccati equations (CRE's) (A7),

$$\dot{\mathbf{x}} = \boldsymbol{\beta} + E\mathbf{x} + a\mathbf{x} + \mathbf{x}(\boldsymbol{\gamma}^T \tilde{I}\mathbf{x}) - \tfrac{1}{2}\boldsymbol{\gamma}(\mathbf{x}^T \tilde{I}\mathbf{x}),$$
$$E\tilde{I} + \tilde{I}E^T = 0, \tag{3.1}$$

$$(\mathbf{y}^T \tilde{I}\mathbf{z}) \equiv y_1 z_1 + \cdots + y_p z_p - y_{p+1} z_{p+1} - \cdots - y_{p+q} z_{p+q},$$

by reducing this system of n $(=p+q)$ equations to $n+1$ projective Riccati's. The simplest way of doing this is to introduce the invariant $O(p,q)$ length as a new variable

$$u(\mathbf{x}) = (\mathbf{x},\mathbf{x}) = (\mathbf{x}^T \tilde{I}\mathbf{x}). \tag{3.2}$$

[Note that $O(p,q) \equiv G_L$ is the subgroup of $O(p+1,q+1)$ that acts linearly on the considered space.] As a consequence of the CRE's, the variable u satisfies

$$\dot{u} = u[(\boldsymbol{\gamma}^T \tilde{I}\mathbf{x}) + 2a] + 2(\boldsymbol{\beta}^T \tilde{I}\mathbf{x}). \tag{3.3}$$

Thus, combining (3.1)–(3.3), we immediately arrive at the desired system of $n+1$ PRE's in $\mathbf{x}$ and u:

$$\dot{\mathbf{x}} = \boldsymbol{\beta} + E\mathbf{x} + a\mathbf{x} - \tfrac{1}{2}\boldsymbol{\gamma}u + \mathbf{x}(\boldsymbol{\gamma}^T \tilde{I}\mathbf{x}),$$
$$\dot{u} = 2(\boldsymbol{\beta}^T I\mathbf{x}) + 2au + u(\boldsymbol{\gamma}^T \tilde{I}\mathbf{x}). \tag{3.4}$$

To solve this system, we can use the method of dimensional reduction, described in Sec. II. At $t = 0$ we impose (3.2) as part of the initial conditions; Eqs. (3.4) then guarantee that this condition will hold for all t, so that $\mathbf{x}$ will indeed solve the CRE's (3.1).

An alternative (but equivalent) approach, which turns out to be more convenient to use in some other cases (see Sec. IV), is to look for integrals of (3.1) of the form

$$C = u(\mathbf{x}) + (\mathbf{f}^T \tilde{I}\mathbf{x}) + f_0, \tag{3.5}$$

where $u(\mathbf{x})$ is the invariant (3.2), and $\mathbf{f} \equiv (f_1,...,f_n)$ and f_0 are functions of t to be determined as follows: Setting $dC/dt = 0$ and using Eq. (3.3) for $\dot{u}$ and (3.1) for $\dot{\mathbf{x}}$ we find first that the quadratic (in x_i) terms identically cancel out. Since the resulting equation must hold for all $\mathbf{x}$, we set the coefficients of the linear (in x_i) terms equal to zero and arrive at the following system of ODE's for $\mathbf{f}$ and f_0:

$$\dot{\mathbf{f}} = -2\boldsymbol{\beta} + a\mathbf{f} + E\mathbf{f} + f_0\boldsymbol{\gamma} - \tfrac{1}{2}\mathbf{f}(\boldsymbol{\gamma}^T \tilde{I}\mathbf{f}),$$
$$\dot{f}_0 = 2af_0 - (\boldsymbol{\beta}^T \tilde{I}\mathbf{f}) - \tfrac{1}{2}f_0(\boldsymbol{\gamma}^T \tilde{I}\mathbf{f}), \tag{3.6}$$

again a set of $(n+1)$ PRE's to be solved by the method of Sec. II.

Now the $n+1$ independent solutions $(\mathbf{f}^i, f_0^i)$ of (3.6), $i = 1,2,...,n+1$, provide $n+1$ independent integrals of (3.1) of the form (3.5), i.e.,

$$C^i = u(\mathbf{x}) + (\mathbf{f}^{iT} \tilde{I}\mathbf{x}) + f_0^i, \quad i = 1,...,n+1. \tag{3.7}$$

Subtracting the C^{n+1} integral from all the others leads to n linear inhomogeneous algebraic equations for the n components of $\mathbf{x}$

$$((\mathbf{f}^{iT} - \mathbf{f}^{(n+1)T})\tilde{I}\mathbf{x}) = C^i - C^{n+1} - f_0^i + f_0^{n+1}, \tag{3.8}$$

from which we can directly obtain the general solution $\mathbf{x}(t)$ of the CRE's (3.1).

IV. INTEGRATION OF SYMPLECTIC AND PSEUDOUNITARY SYSTEMS OF ODE'S

In this section we integrate two different types of symplectic Riccati equations with $n = 3$ introduced in the Appendix for the group $G = \mathrm{Sp}(2N,F)$ [cf. (A11)] and for the group $\mathrm{Sp}(2N,F) \otimes \mathrm{Sp}(2N,F)$. Both can be reduced to equations we have already integrated in the previous sections.

However, since for $n > 3$ this reduction can no longer be done and since their actual form is sufficiently different from CRE's and PRE's, we proceed to integrate them here independently. To do this, we apply the integration procedure described in Sec. II introducing a new variable $u(\mathbf{x})$, which is the invariant of the subgroup $G_L \subset G$ acting linearly on the corresponding space G/G_0.

We start with the SRE1 system of Eq. (A11), which, with $N = 2$ and $n = N(N+1)/2 = 3$, yields the following system of ODE's:

$$\begin{aligned}\dot{x} &= c_{11} + 2a_{11}x + 2a_{12}y + g_{11}x^2 + 2g_{12}xy + g_{22}y^2,\\ \dot{y} &= c_{12} + a_{21}x + (a_{11}+a_{22})y + a_{12}z + g_{11}xy \\ &\quad + g_{12}(xz+y^2) + g_{22}yz,\\ \dot{z} &= c_{22} + 2a_{21}y + 2a_{22}z + g_{11}y^2 + 2g_{12}yz + g_{22}z^2,\end{aligned} \tag{4.1}$$

where c_{ij}, a_{ij}, and g_{ij} are the matrix elements of C, A, and G in (A11), respectively. The subalgebra of sp(4,$\mathbb{R}$) that acts linearly on the space (x,y,z) is sl(2,$\mathbb{R}$) given by the matrix A (with $C = G = 0$). The invariant of the corresponding group SL(2,$\mathbb{R}$) to be used in the construction of integrals of the form (3.7) is the det W, i.e.,

$$u(\mathbf{x}) = -\det W = -\det\begin{pmatrix} x & y \\ y & z\end{pmatrix} = y^2 - xz. \tag{4.2}$$

Using (4.1) we find the corresponding ODE for $\dot{u}$ and combine it with (4.1) to obtain a system of four projective Riccati equations for the variables x, y, z, and u

$$\begin{aligned}\dot{x} &= c_{11} + 2a_{11}x + 2a_{22}y + g_{22}u \\ &\quad + x(g_{11}x + 2g_{12}y + g_{22}z),\\ \dot{y} &= c_{12} + 2a_{21}x + (a_{11}+a_{22})y + a_{12}z \\ &\quad - g_{12}u + y(g_{11}x + 2g_{12}y + g_{22}z),\\ \dot{z} &= c_{22} + 2a_{21}y + 2a_{22}z + g_{11}u \\ &\quad + z(g_{11}x + 2g_{12}y + g_{22}z),\\ \dot{u} &= -c_{22}x + 2c_{12}y - c_{11}z + 2(a_{11}+a_{22})u \\ &\quad + u(g_{11}x + 2g_{12}y + g_{22}z).\end{aligned} \tag{4.3}$$

This system of PRE's can again be solved by dimensional reduction as in Sec. II, with the constraint (4.2) imposed at $t = 0$ as part of the initial conditions.

The SRE2 system, on the other hand, related to the semisimple Lie algebra $\mathscr{L} = \mathrm{sp}(2N,\mathbb{R}) \oplus \mathrm{sp}(2N,\mathbb{R})$, see Eq. (A24), for $N = 1$, takes the form

$$\begin{aligned}\dot{x} &= a_1 + b_3y - b_2z + a_1(-x^2 + yz) - a_2xz - a_3xy,\\ \dot{y} &= a_2 + 2b_2x - 2b_1y - 2a_1xy + a_2x^2 - a_3y^2,\\ \dot{z} &= a_3 - 2b_3x + 2b_1z - 2a_1xz - a_2z^2 + a_3x^2.\end{aligned} \tag{4.4}$$

The subgroup of $\mathrm{Sp}(2,\mathbb{R}) \otimes \mathrm{Sp}(2,\mathbb{R})$ that acts linearly on the underlying space (x,y,z) is itself $G_L \sim \mathrm{Sp}(2,\mathbb{R}) \sim \mathrm{SL}(2,\mathbb{R})$. Its invariant is det V and hence we set for our $u(\mathbf{x})$

$$u(\mathbf{x}) = -\det V = -\det\begin{pmatrix} x & y \\ z & -x\end{pmatrix} = x^2 + yz. \tag{4.5}$$

Calculating now $\dot{u}$ with the aid of (4.4) and combining the equations for x, y, z, and u together we obtain again a special case of PRE's

$$\begin{aligned}\dot{x} &= a_1 + b_3y - b_2z + a_1u - x(2a_1x + a_3y + a_2z),\\ \dot{y} &= a_2 + 2b_2x - 2b_1y + a_2u - y(2a_1x + a_3y + a_2z),\\ \dot{z} &= a_3 - 2b_3x + 2b_1z + a_3u - z(2a_1x + a_3y + a_2z),\\ \dot{u} &= 2a_1x + a_3y + a_2z - u(2a_1x + a_3y + a_2z),\end{aligned} \tag{4.6}$$

which can be solved by dimensional reduction, with the constraint (4.5) imposed at $t = 0$.

Finally, we turn to the "pseudounitary" ODE's (A19) of the Appendix, based on the action of SU(p,1). The subgroup of SU(p,1) that acts linearly on the corresponding Grassmannian is SU($p-1$) and its invariant is precisely the "unitary length" defined in (A18) by

$$u(\xi) = -\tfrac{1}{2}(\xi^\dagger,\xi) = -\tfrac{1}{2}(\xi_1^*\xi_1 + \cdots + \xi_{p-1}^*\xi_{p-1}). \tag{4.7}$$

Here, however, the original equations of the system are already in (complex) projective Riccati form

$$\begin{aligned}\dot{\zeta} &= ic + (\alpha + \alpha^*)\zeta + (\boldsymbol{\beta}^\dagger,\xi) + \zeta[-if\zeta + (\boldsymbol{\delta}^\dagger,\xi)],\\ \dot{\xi} &= -\boldsymbol{\beta} + \boldsymbol{\delta}\zeta + E\xi + \alpha^*\xi + \xi[-if\zeta + (\boldsymbol{\delta}^\dagger,\xi)],\end{aligned} \tag{4.8}$$

where $\zeta = u + ix$, cf. (A17). We thus prefer to solve this system of PRE's directly (e.g., by the method of dimensional reduction), choosing initial conditions such that (4.7) is satisfied at $t = 0$. Since (4.7) is a group invariant, it will then be satisfied for all $t \geqslant 0$, and the complete solution of the problem will have been determined.

V. EVIDENCE OF CHAOS IN A SYSTEM OF TWO PERTURBED PRE'S

In recent publications[14,15] there have been attempts to connect the concept of integrability of dynamical systems to the Painlevé property[8,9] of their solutions in complex t. In these investigations, dynamical systems (generally non-Hamiltonian) having the Painlevé property, were always found to fall in one of three categories[14]: they either (A) possessed as many integrals as the order of the system, analytic in t and polynomial in the dependent variables; (B) could be transformed to a system of linear ODE's (with time-dependent coefficients); or (C) could be reduced to one of Painlevé's second- (or possibly higher-) order transcendental equations.[16]

Clearly, the above classification represents only a first attempt at defining integrability in generally non-Hamiltonian dynamical systems. For example, there is an obvious overlap between (B) and (A) above, since, for a system of n linear ODE's, we can always write down n integrals, as linear combinations of the dependent variables, having as coefficients the elements of the fundamental solution matrix of the system. Be that as it may, there is by now considerable analytical and numerical evidence that dynamical systems possessing the Painlevé property have globally "regular" solutions,[7–9] while infinitely branched singularities in complex time t can lead to "chaotic" or "turbulent" motions in real time.[5,17]

The systems we have analyzed in this paper belong to category (B) above, since they can all be linearized in spaces of higher dimensions.[13] In that sense, they possess, of course, the Painlevé property and are expected to be free from "strange attractors," infinite period-doubling sequences, or any other such type of chaotic phenomena.

To investigate how systems with superposition principles behave and what might happen if they are perturbed away from their precise algebraic form, we have studied the following system of ODE's:

$$\dot{x} = y + x(c_1x + c_2y),$$
$$\dot{y} = -(2 + Q\cos 2t)x + y(c_1x + c_3y). \quad (5.1)$$

Note that for $c_3 = c_2$ these two equations form a system of PRE's of the type we analyzed and explicitly solved in Sec. II, and possess the Painlevé property (generally violated for $c_3 \neq c_2$). Moreover, since all systems of ODE's with superposition principles studied in this paper can be reduced to PRE's (see Secs. III and IV), we might expect the analysis of equations such as (5.1) to have a wider applicability and a more general significance.

Before taking $c_3 \neq c_2$ to see what happens in the non-PRE case, let us first take $c_3 = c_2$ and write down the general solution of (5.1), as obtained, for example, by the methods of Sec. II:

$$x(t) = \frac{u}{(K - c_1\int^t u\, ds - c_2u)}, \quad y(t) = \frac{\dot{u}x(t)}{u}, \quad (5.2)$$

where $u(t)$ is the general solution of the Mathieu equation[18]

$$\frac{d^2u}{dt^2} + (2 + Q\cos 2t)u = 0. \quad (5.3)$$

Here K and the ratio of the two arbitrary constants in the general solution of (5.3) are the two free constants in (5.2) to be specified by the initial conditions $x(0), y(0)$.

Take, for example, the case $Q = 0$ first: If we write the solution of (5.3) as $u = A\cos\phi$ ($\phi \equiv \sqrt{2}t + \phi_0$), solve Eqs. (5.2) for $\cos\phi$, $\sin\phi$, and use $\cos^2\phi + \sin^2\phi = 1$ to eliminate ϕ, we obtain a one-parameter family of conic sections

$$x^2(K'^2 - c_2^2) + y^2(2K'^2 - c_1^2)/4 + c_1c_2xy - 2c_2x + c_1y - 1 = 0, \quad (5.4)$$

$K' \equiv K/A$. Thus, for values of $K'^2 > c_2^2 + c_1^2/2$, we find that the exact solutions trace out ellipses around the origin of the x, y plane, where the motion is oscillatory and bounded, see Fig. 1. These ellipses limit on a parabola and for $K'^2 < c_2^2 + c_1^2/2$ all solutions run away to infinity along hyperbolas.

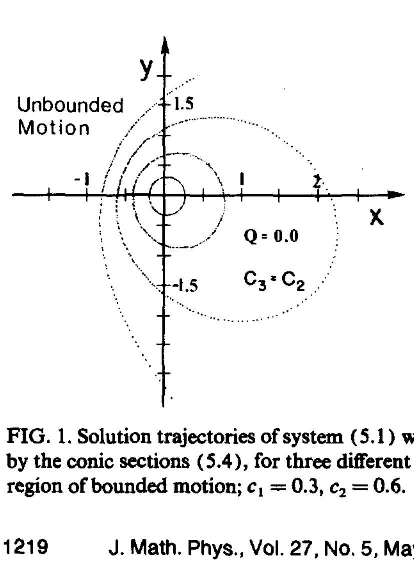

FIG. 1. Solution trajectories of system (5.1) with $c_3 = c_2$ and $Q = 0$, given by the conic sections (5.4), for three different initial conditions within the region of bounded motion; $c_1 = 0.3$, $c_2 = 0.6$.

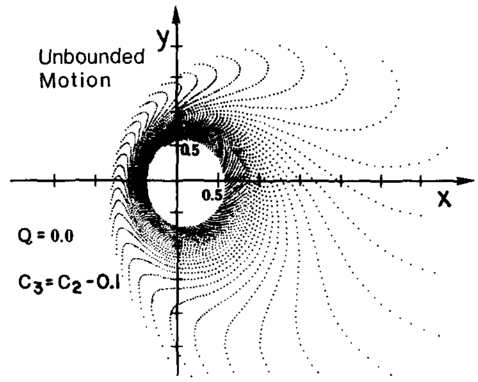

FIG. 2. One solution of (5.1) at $Q = 0$ and $c_3 = c_2 - 0.1$, $c_1 = 0.3$, $c_2 = 0.6$, starting near the $x < 0$ part of the boundary separating bounded from unbounded motion.

In the case $c_3 \neq c_2$ (and $Q = 0$), Eqs. (5.1) no longer possess nonlinear superposition rules and hence cannot be integrated by the methods of this paper. Integrating them numerically, we find that $c_3 = c_2 - 0.1$, for example, introduces an overall dissipative effect on the motion and makes the origin an equilibrium of the stable spiral type,[1] see Fig. 2.

On the other hand, for $c_3 = c_2$ and $Q \neq 0$, Eqs. (5.1) are still of the projective Riccati type and the situation is not very different from the $Q = 0$ case. The system, however, is no longer autonomous and Eqs. (5.1) are not invariant under time translations. Thus, together with the initial conditions $(x(t_0), y(t_0))$, the initial time t_0 must also be specified for a *unique* determination of each solution (since now solutions can intersect themselves on the x, y plane, without being necessarily periodic).

To take this into account, we shall study the system (5.1) in the extended phase space x, y, t making use of the periodicity of the time-dependent term $Q\cos 2t$. In other words, we shall consider Poincaré "surfaces of section"[1-4] and plot orbits (i.e., intersections of solutions) in the x, y plane at $t = k\pi$ intervals (k, integer). One such section is shown, for example, in Fig. 3 for three orbits corresponding to three different initial conditions.

Interestingly enough, the conic sections of Fig. 3 can also be obtained analytically starting with (5.2): Evaluating x, y at $t = k\pi$, we eliminate $\sin(\alpha k\pi)$ and $\cos(\alpha k\pi)$ (α is the

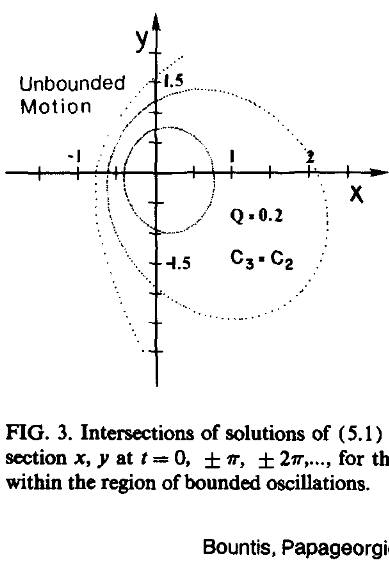

FIG. 3. Intersections of solutions of (5.1) at $c_3 = c_2$, with the surface of section x, y at $t = 0, \pm\pi, \pm 2\pi, \ldots$, for three different initial conditions within the region of bounded oscillations.

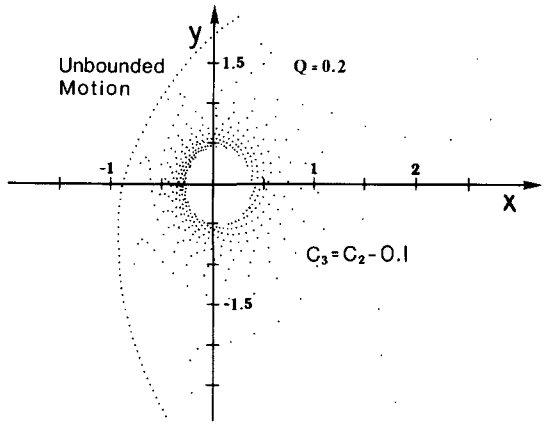

FIG. 4. Intersections of one solution of (5.1) with the Poincaré surface of section, with initial conditions very close to the $x<0$ part of the boundary; $c_1=0.3$, $c_2=0.6$, $c_3=0.5$. Note that, at $Q=0.2$, curves spiraling towards (0,0) can still be confidently traced through the points.

Floquet exponent[18]) as described above, and arrive at an equation similar to (5.4). The only difference now is that the coefficients of the terms x^2, y^2, xy, etc. in (5.4) contain sums of the Fourier coefficients of the Mathieu function solutions of (5.3).

Taking now $c_3=c_2-0.1$ and $Q\neq 0$ we plot in Figs. 4 and 5, the intersections ($t>0$) of *one* solution starting in each case with initial conditions $y(0)=0$, and an $x(0)$ very close to the smooth boundary—which still exists!—separating bounded from unbounded motion. The existence of this boundary can be verified by numerically integrating (5.1) backwards, and observing all solutions tend to it as $t\to-\infty$.

But, what is especially interesting about these two figures is the way the solutions behave near the $x<0$ part of the boundary: Observe the sudden changes in the direction of the tentaclelike "curves" that could be traced through different groups of points as they spiral towards the origin, for $Q=0.2$, in Fig. 4. In fact, as Q increases further, these "curves" change direction so sharply and accumulate near the $x<0$ part of the boundary so densely, that it becomes impossible to trace them out visually with any degree of confidence, see Fig. 5. This results in a kind of transient chaotic behavior on the Poincaré map: As long as solutions remain

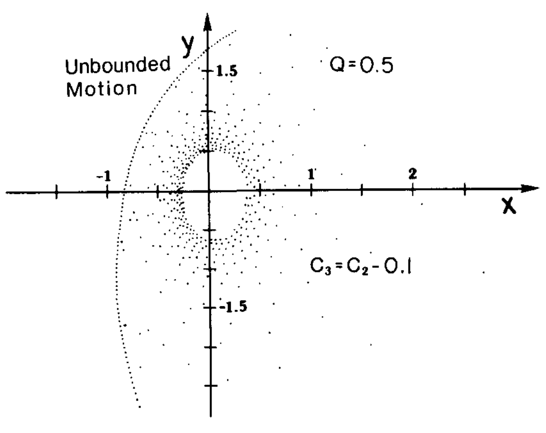

FIG. 5. Same as Fig. 4, at $Q=0.5$. Note that here the intersections of one solution wander around "chaotically" before settling in their spiraling motion towards the origin.

close to this boundary, they oscillate in a very erratic and irregular fashion and produce sequences of points that depend very sensitively on the choice of initial conditions.

This sensitive dependence on initial conditions is generally accepted as evidence of chaos, or chaotic behavior in nonintegrable dynamical systems.[1–4] It occurs most frequently near the intersections of stable and unstable invariant manifolds associated with (linearly) unstable fixed points (or unstable periodic orbits) of the Poincaré map. These intersecting manifolds can actually evolve into the so-called "strange attractors," about which so many results have appeared in the recent scientific literature.[4,19]

Of course, the chaotic behavior we have observed here has nothing to do with strange attractors, or any other of the known "scenarios" of chaos, associated with different bifurcation phenomena.[20] It does suggest, however, that chaotic behavior—in the sense of sensitive dependence on initial conditions—can also arise in certain regions of nonintegrable dynamical systems that are expected to be "sensitive," as, e.g., near a boundary separating bounded from unbounded motion.

Undoubtedly, different perturbations of systems of ODE's having superposition principles will produce different nonintegrable dynamical systems, exhibiting a variety of chaotic phenomena. What we described in this section was but one example. A more complete investigation of such perturbations is currently in progress and new results will be reported in future publications.

VI. CONCLUSIONS

Our main conclusion is that the existence of superposition formulas for equations of type (1.1) can be used as a sufficient (albeit not necessary) criterion of integrability. Moreover, for $n\leqslant 3$ we have integrated all indecomposable systems of such equations. In Sec. V we have indicated the difference in the behavior of solutions of equations allowing, or not allowing, superposition formulas, respectively.

It should be stressed that the classification of systems of ODE's obtained in the Appendix and listed in the Introduction is exhaustive for $n\leqslant 3$. Each set of equations obtained represents an equivalence class of equations. Indeed, performing an arbitrary change of coordinates on the space G/G_0 (see the Introduction), i.e., an arbitrary invertible and sufficiently smooth change of dependent variables $x_i=\phi_i(y_1,\ldots,y_n)$ ($i=1,\ldots,n$) in (1.1), we obtain a system of equations, that may look completely different, but shares all the integrability properties of the original set of equations.

It is quite easy to identify a system of equations of type (1.1) with a superposition formula, making use of a theorem due to Lie.[21] Indeed, in order to allow a superposition formula, Eqs. (1.1) must have the form

$$\frac{dx_i}{dt}=\sum_{a=1}^{r} Z_a(t)\zeta_{i,a}(x),\quad i=1,2,\ldots,n \tag{6.1}$$

(for some finite r) and the vector fields

$$X^a=\sum_{i=1}^{n}\zeta_{i,a}(x)\frac{\partial}{\partial x_i},\quad a=1,\ldots,r, \tag{6.2}$$

must generate a finite-dimensional Lie algebra

$$[X^a, X^b] = \sum_c f_{abc} X^c. \tag{6.3}$$

The vector fields (6.2) can be read off directly from the equations. If they generate a finite-dimensional Lie algebra, then the structure constants f_{abc} are given by (6.3). They in turn completely identify the algebras L and L_0 of the Introduction, and hence tell us in which equivalence class of equations we are. For further information on this subject we refer to earlier publications, in particular Ref. 13.

ACKNOWLEDGMENTS

One of the authors (T.C.B.) wishes to acknowledge partial support for this work by National Science Foundation Grant No. DMS-8314172. Another (P.W.) acknowledges support in the form of research grants from the Natural Sciences and Engineering Research Council of Canada and from the Fonds pour l'aide et le soutien a la recherche du Gouvernement du Québec.

APPENDIX: INDECOMPOSABLE LOW-DIMENSIONAL SYSTEMS OF ODE'S WITH SUPERPOSITION PRINCIPLES

In this appendix we list all transitive primitive Lie algebra pairs $\{\mathscr{L}, \mathscr{L}_0\}$, $\mathscr{L} \supset \mathscr{L}_0$, which give rise to the systems of n ODE's, $n \leqslant 3$, integrated in this paper. We make use of the known classification of all transitive primitive Lie algebras and write down the corresponding ODE's explicitly, in the low-dimensional cases analyzed in Secs. II, III, and IV. For more details on this classification of Lie algebras and its connection to systems of ODE's as well as references to the original work, see Ref. 13.

To start with, two main possibilities are distinguished: either (i) $\mathscr{L}$ is a simple Lie algebra, with $\mathscr{L}_0$ being either a maximal parabolic subalgebra or a maximal reductive subalgebra, or (ii) $\mathscr{L}$ is a semisimple (but not simple) Lie algebra. There is a third possibility: that $\mathscr{L}$ is not semisimple but $\mathscr{L} \approx \mathscr{L}_0 + V$, where V is Abelian and $\mathscr{L}_0$ acts faithfully and irreducibly on V, thus being a reductive and in particular simple or semisimple subalgebra. However, this case leads to inhomogeneous linear ODE's and will not be discussed further since our interest here is in nonlinear systems.

In many cases we find it convenient to write down the nonlinear ODE's with superposition formulas for n arbitrary, and only afterwards reduce to $n \leqslant 3$. Thus, we do not discuss here all possible families of such ODE's but only those that provide systems of $n \leqslant 3$ equations. (Certain families start with $n > 3$, while others provide systems, which for $n \leqslant 3$ are equivalent to some we have already listed.)

It should also be mentioned here that the explicit form of the systems of ODE's we are considering is coordinate dependent. Indeed, an arbitrary (invertible) change of dependent variables will lead to different (but equivalent) systems of equations. Finally, although our emphasis in this paper is on systems of $n \leqslant 3$ real equations, we also discuss for completeness complex transitive primitive Lie algebras, leading to $n \leqslant 3$ complex equations.

A. $\mathscr{L}$ is a simple Lie algebra

1. $\mathscr{L}_0$ is a maximal parabolic subalgebra

a. $\mathscr{L} = \mathrm{sl}(N,F)$, $F = \mathbb{R}$ *or* $\mathbb{C}$. We realize $\mathscr{L}$ by the $N \times N$ matrices

$$X = \begin{pmatrix} C & A \\ -D & -B \end{pmatrix}, \quad \begin{matrix} C \in F^{l\times l}, & A \in F^{l\times k}, \\ D \in F^{k\times l}, & B \in F^{k\times k}, \end{matrix} \tag{A1}$$

with $\operatorname{Tr} C = \operatorname{Tr} B$, $l + k = N$, $1 \leqslant k \leqslant l$. The subalgebra $\mathscr{L}_0 = \mathrm{aff}(l,k,F)$ corresponds to $A = 0$. Introducing affine coordinates on the homogeneous space G/G_0 as in Ref. 13 and writing the corresponding ODE's we obtain the matrix Riccati equations (MRE's)

$$\dot{W} = A + WB + CW + WDW, \quad W \in F^{l\times k}. \tag{A2}$$

The number of equations $n = lk \leqslant 3$ only if $k = 1$, whence $l = 1,2$, or 3. We thus obtain a very important special case of MRE's: the *projective* Riccati equations (PRE's)

$$\dot{\mathbf{x}} = \mathbf{a} + B\mathbf{x} + \mathbf{x}(\mathbf{c},\mathbf{x}),$$

or, in component form

$$\dot{x}_\mu = a_\mu + \sum_{\nu=1}^{n} b_{\mu\nu} x_\nu + x_\mu \sum_{\nu=1}^{n} c_\nu x_\nu, \quad \mu = 1,2,\dots,n, \tag{A3}$$

which we solved in Sec. II for a_μ, $b_{\mu\nu}$, and c_ν arbitrary functions of t.

b. $\mathscr{L} = \mathrm{o}\,(\mathrm{N}+2,\mathbb{C})$ *or* $\mathscr{L} = \mathrm{o}(p+1, q+1)$. Define the matrices J and X by

$$J = \begin{pmatrix} 0 & & I_k \\ & \tilde{I} & \\ I_k & & 0 \end{pmatrix}, \quad \begin{matrix} XJ + JX^T = 0, & 1 \leqslant k \leqslant q+1 \leqslant p+1, \\ X \in F^{(p+q)\times(p+q)}, & p + q = N, \end{matrix} \tag{A4}$$

where, if $F = \mathbb{C}$, $\tilde{I} \equiv I_{N+2-2k}$, while, if $F = \mathbb{R}$, $\tilde{I} \equiv I_{p+1-k,q+1-k}$, I_p being the $p\times p$ identity matrix and $I_{i,j}$ the $(i+j)\times(i+j)$ identity with the lower j diagonal elements replaced by -1. The algebra $\mathscr{L}$ here is realized by the matrices

$$X = \begin{pmatrix} A & B^T & C \\ D & E & -\tilde{I}B \\ G & -D^T\tilde{I} & -A^T \end{pmatrix},$$
$$A,C,G \in F^{k\times k}, \quad C + C^T = 0,$$
$$B,D \in F^{(N+2-2k)\times k}, \quad G + G^T = 0, \tag{A5}$$
$$E \in F^{(N+2-2k)\times(N+2-2k)}, \quad E\tilde{I} + \tilde{I}E^T = 0.$$

The subalgebra $\mathscr{L}_0$ is obtained by putting $B = 0$ and $C = 0$. The number of ODE's is

$$n = [(2N - 3k + 3)k/2], \quad 1 \leqslant k \leqslant q \leqslant [(N+2)/2], \tag{A6}$$

with $n \leqslant 3$ occurring in three cases.

(i) The *conformal* Riccati equations are where $k = 1$ and $p \geqslant q \geqslant 0$ [$\mathscr{L}_0 \sim \mathrm{sim}(p,q)$, the similitude algebra of the $(p+q)$-dimensional Minkowski space]

$$\dot{\mathbf{x}} = -\tilde{I}\boldsymbol{\beta} + E\mathbf{x} + a\mathbf{x} + \mathbf{x}(\boldsymbol{\gamma}^T\tilde{I}\mathbf{x}) - \tfrac{1}{2}\boldsymbol{\gamma}(\mathbf{x}^T\tilde{I}\mathbf{x}), \tag{A7}$$

$\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{x} \in F^{n\times 1}$, $E\tilde{I} + \tilde{I}E^T = 0$, $E \in F^{n\times n}$, $a \in F$,

$F = C$ leads to n complex equations (and $\tilde{I} = I_n$). For $F = \mathbb{R}$

and $n=2$ we have the o(3,1) case with $\tilde{I}=I_2$, equivalent to the case of one complex Riccati equation. For $n=3$, two inequivalent systems arise: o(4,1) with $\tilde{I}=I_3$ and o(3,2) with $\tilde{I}=I_{2,1}$ integrated together as systems CRE1 and CRE2 in Sec. III.

(ii) The orthogonal matrix Riccati equations with $k=N/2$, N even, are based on $o(N,\mathbb{C})$ or $o(k,k)$,

$$\begin{aligned} &\dot{Z}=C+AZ+ZA^T-ZGZ,\\ &Z=-Z^T\in F^{k\times k},\quad A\in F^{k\times k},\\ &C=-C^T\in F^{k\times k},\quad G=-G^T\in F^{k\times k}.\end{aligned}\tag{A8}$$

This gives a system of $n=k(k-1)/2$ equations, which for $k=3$ reduces to $n=3$ projective Riccati's (A3) as a consequence of the isomorphism between $o(6,\mathbb{C})$) and $sl(4,\mathbb{C})$ and their corresponding homogeneous spaces [respectively o(3,3) and $sl(4,\mathbb{R})$].

(iii) The $o(2k+1,\mathbb{C})$ or $o(k+1,k)$ Riccati equations with $k=(N-1)/2$, N odd give again a system of $n=k(k+1)/2$ equations, which, for $k=2$, reduces to three PRE's (A4), cf. Sec. II.

c. $\mathscr{L}=sp(2N,F)$, *a symplectic Lie algebra.* Defining the matrices $K_{\lambda\mu}$ and $X\in F^{2N\times 2N}$

$$\begin{aligned} &K_{\lambda\mu}\equiv\begin{pmatrix}0 & & I_\lambda\\ & K & \\ -I_\lambda & & 0\end{pmatrix},\quad K\equiv\begin{pmatrix}0 & I_\mu\\ -I_\mu & 0\end{pmatrix},\\ &XK_{\lambda\mu}+K_{\lambda\mu}X^T=0,\quad \lambda+\mu=N,\quad \lambda\geqslant 1,\end{aligned}\tag{A9}$$

we realize the algebra $\mathscr{L}$ by the matrices

$$\begin{aligned} &X=\begin{pmatrix}A & B^T & C\\ D & E & KB\\ -G & -D^TK & -A^T\end{pmatrix},\\ &A,C,G\in F^{\lambda\times\lambda},\quad E\in F^{2\mu\times 2\mu},\quad D\in F^{2\mu\times\lambda}\\ &C=C^T,\quad F=F^T,\quad E=KE^TK,\quad B\in F^{2\mu\times\lambda},\end{aligned}\tag{A10}$$

and obtain $\mathscr{L}_0$ by setting $B=C=0$. The number of equations here is $n=\lambda(4N-3\lambda+1)/2$, $\lambda=1,2,\ldots,N$. Thus $n=2$ does not occur, and $n=3$ occurs only for $N=2$ with $\lambda=1$, or $\lambda=N=2$.

Now the $\lambda=1$ case is easily seen to lead to a special case of $n=2N-1$ PRE's.[13] However, for $\lambda=N$ we get the $n=N(N+1)/2$ symplectic Riccati equations

$$\begin{aligned} &\dot{W}=C+AW+WA^T+WGW,\\ &C,G,W\in F^{N\times N},\quad C=C^T,\quad G=G^T,\quad W=W^T.\end{aligned}\tag{A11}$$

With $N=2$, (A11) yields the system SRE1, integrated in Sec. IV, with $W=\begin{pmatrix}x & y\\ y & z\end{pmatrix}$, cf. (4.1).

In view of the isomorphisms $sp(4,\mathbb{R})\sim o(3,2)$ and $sp(4,\mathbb{C})\sim o(5,\mathbb{C})$ both sets of $n=3$ equations based on $sp(4,F)$ can be reduced to o(3,2) or $o(5,\mathbb{C})$ equations. Those for $\lambda=N=2$ can be transformed into conformal Riccati equations. Physically, however, they are quite different and we integrate them here separately in Sec. IV. For $N>2$ the SRE's are independent of any of the other systems of equations.

d. $\mathscr{L}=su(p,q)$, *a pseudounitary Lie algebra.* Here, we set

$$\begin{aligned} &J\equiv\begin{pmatrix}0 & & I_k\\ & \tilde{I} & \\ I_k & & 0\end{pmatrix},\\ &X^\dagger J+JX=0,\quad p\geqslant q\geqslant k\geqslant 1,\\ &\tilde{I}=I_{p-k,q-k},\quad N=p+q,\end{aligned}\tag{A12}$$

where † denotes Hermitian conjugation. The algebra $\mathscr{L}$ is thus realized by the complex matrices

$$\begin{aligned} &X=\begin{pmatrix}A & B^\dagger & C\\ D & E & -\tilde{I}B\\ F & -D^\dagger\tilde{I} & -A^\dagger\end{pmatrix},\\ &A,C,F\in\mathbb{C}^{k\times k},\quad E\in\mathbb{C}^{(p+q-2k)\times(p+q-2k)},\\ &B,D\in\mathbb{C}^{(p+q-2k)\times k},\end{aligned}\tag{A13}$$

$C+C^\dagger=0, F+F^\dagger=0, E^\dagger\tilde{I}+\tilde{I}E=0$. The algebra $\mathscr{L}_0$ is obtained by putting $B=C=0$, and the number of real equations is $n=k(2N-3k)$. We see, therefore, that $n=1$ occurs for su(1,1) (yielding the ordinary real Riccati equation), that $n=2$ does *not* occur, and that $n=3$ occurs only for su(2,1) ($p=2, q=k=1$).

A linear system associated with this transitive primitive Lie algebra $(\mathscr{L},\mathscr{L}_0)$ is

$$\begin{aligned} &\begin{pmatrix}\dot{U}_1\\ \dot{U}_2\\ \dot{U}_3\end{pmatrix}=\begin{pmatrix}A & B^\dagger & C\\ D & E & -\tilde{I}B\\ F & -D^\dagger\tilde{I} & -A^\dagger\end{pmatrix}\begin{pmatrix}U_1\\ U_2\\ U_3\end{pmatrix},\\ &U_1,U_3\in\mathbb{C}^{k\times k},\quad U_2\in\mathbb{C}^{(p+q-2k)\times k}.\end{aligned}\tag{A14}$$

We remove the redundancy in the above homogeneous coordinates by introducing the usual affine coordinates as elements of the matrices

$$Z_1=U_1U_3^{-1},\quad Z_2=U_2U_3^{-1},\quad \det U_3\neq 0.$$

The transitivity of the SU(p,q) group action is imposed by restricting to the Grassmannian of isotropic planes

$$U_1^\dagger U_3+U_2^\dagger\tilde{I}U_2+U_3^\dagger U_1=0,$$

i.e.,

$$Z_1+Z_1^\dagger+Z_2^\dagger\tilde{I}Z_2=0.\tag{A15}$$

In terms of the coordinates Z_i the "pseudounitary Riccati equations" are

$$\begin{aligned} &\dot{Z}_1=C+AZ_1+Z_1A^\dagger+B^\dagger Z_2-Z_1FZ_1+Z_1D^\dagger\tilde{I}Z_2,\\ &\dot{Z}_2=-\tilde{I}B+DZ_1+EZ_2+Z_2A^\dagger-Z_2FZ_1+Z_2D^\dagger\tilde{I}Z_2,\end{aligned}\tag{A16}$$

subject to the constraint (A15). In particular, consider the case $k=q=1$ ($\tilde{I}=I$). If we put

$$Z_1\equiv\zeta=u+ix,\quad Z_2\equiv\boldsymbol{\xi},\quad u,x\in\mathbb{R},\quad \boldsymbol{\xi}\in C^{p-1},\tag{A17}$$

Eq. (A15) yields

$$u=-\tfrac{1}{2}(\boldsymbol{\xi}^\dagger,\boldsymbol{\xi}),\tag{A18}$$

while Eqs. (A16) reduce to

$$\begin{aligned} &\dot{\zeta}=ic+(\alpha+\alpha^*)\zeta+(\boldsymbol{\beta}^\dagger,\boldsymbol{\xi})+\zeta[-if\zeta+(\boldsymbol{\delta}^\dagger,\boldsymbol{\xi})],\\ &\dot{\boldsymbol{\xi}}=-\boldsymbol{\beta}+\boldsymbol{\delta}\zeta+E\boldsymbol{\xi}+\alpha^*\boldsymbol{\xi}+\boldsymbol{\xi}[-if\zeta+(\boldsymbol{\delta}^\dagger,\boldsymbol{\xi})]\end{aligned}\tag{A19}$$

(with stars denoting complex conjugation). Using (A18) to eliminate $u = \text{Re}\,\zeta$ we obtain a system of $n = 2p-1$ real equations with polynomial nonlinearities up to order 4:

$$\dot{x} = c + (\alpha + \alpha^*)x - (i/2)[(\beta^\dagger,\xi) - (\xi^\dagger,\beta)] + fx^2 + (x/2)[(\delta^\dagger,\xi) + (\xi^\dagger,\delta)] + (i/4)[(\delta^\dagger,\xi) - (\xi^\dagger,\delta)](\xi^\dagger,\xi)^2 - \tfrac{1}{4}f(\xi^\dagger,\xi)^2, \quad \text{(A20)}$$

$$\dot{\xi} = -\beta + i\delta x + E\xi + \alpha^*\xi - \tfrac{1}{2}\delta(\xi^\dagger,\xi) + \xi[fx + (\delta^\dagger,\xi)] + (i/2)f\xi(\xi^\dagger,\xi).$$

In the case of su(2,1) we have $p = 2$ so that (A20) reduces to three real equations, since ξ is just one complex variable.

2. $\mathscr{L}_0$ is a maximal reductive subalgebra

Requiring that $n = 3$ be the number of equations $n = \dim \mathscr{L} - \dim \mathscr{L}_0$, leads to only two possibilities: $\mathscr{L} = \text{o}(3,1)$ and $\mathscr{L}_0 = \text{o}(3)$ or $\text{o}(2,1)$. [Note that the corresponding complex Lie algebra o(4,C) is *not* simple.] Consider the more general case $\mathscr{L} = \text{o}(p,1)$, $\mathscr{L}_0 = \text{o}(p)$, the homogeneous space $\text{O}(p,1)/\text{O}(p)$ can be realized as the upper sheet of a two-sheeted hyperboloid $x_0^2 - x_1^2 - \cdots - x_p^2 = 1$. Using the linear equations satisfied by the homogeneous coordinates x_μ, and introducing the projective coordinates $z_i = x_i/x_0$, $x_0 \neq 0$ yields

$$\dot{z}_i = A_{i0} + A_{ik}z_k - z_iA_{0k}z_k, \quad i,k = 1,\dots,p,$$
$$A_{ik} = -A_{ki} \in \mathbb{R}^{p\times p}, \quad A_{0i} = A_{i0} \in \mathbb{R}^{p\times p}.$$

These equations are, clearly, again a special case of PRE's.

Similarly, the second possibility, viewed as a special case of $\mathscr{L} = \text{o}(p,1)$, $\mathscr{L}_0 = \text{o}(p-1,1)$, on the space $\text{O}(p,1)/\text{O}(p-1,1)$ realized as the one-sheeted hyperboloid $x_1^2 + x_2^2 + \cdots + x_p^2 - x_0^2 = 1$, leads in the projective coordinates $Z_\alpha \equiv x_\alpha/x_p$, $x_p \neq 0$, to another special case of PRE's

$$\dot{z}_\alpha = A_{\alpha p} + A_{\alpha\beta}z_\beta - z_\alpha A_{p\beta}z_\beta,$$

$\alpha, \beta = 0,1,\dots,p-1$, where $A \in \mathbb{R}^{(p+1)\times(p+1)}$ satisfies

$$AI_{p,1} + I_{p,1}A^T = 0.$$

B. $\mathscr{L}$ is a semisimple Lie algebra

As shown in Ref. 13 it is possible to construct transitive primitive Lie algebras $\{\mathscr{L},\mathscr{L}_0\}$ in which

$$\mathscr{L} = \mathscr{K} \oplus \mathscr{K}, \quad \mathscr{L}_0 = \mathscr{K}_D, \quad \text{(A21)}$$

where $\mathscr{K}$ is a simple Lie algebra and $\mathscr{K}_D \sim \mathscr{K}$ is embedded "diagonally" in $\mathscr{L}$. In this case the number of equations in the associated systems of ODE's is

$$n = \dim \mathscr{L} - \dim \mathscr{L}_0 = \dim \mathscr{K}. \quad \text{(A22)}$$

Thus, $n = 1$ or 2 are excluded, while $n = 3$ in the case of real equations, yields two inequivalent possibilities.

(i) $\mathscr{K} = \text{su}(2) \sim \text{o}(3)$. In general, taking $\mathscr{K} = \text{su}(n)$, we obtain a special case of the complex MRE's (A2). Moreover, for $n = 2$, a low-dimensional coincidence occurs and the $\text{su}(2) \oplus \text{su}(2)$ equations turn out to be a special case of the projective Riccati equations.

Similarly, taking $\mathscr{K} = \text{o}(n)$ we obtain a special case of MRE's and, in particular, for $n = 3$ the same special case of projective Riccati equations as above.

(ii) $\mathscr{K} = \text{sp}(2,\mathbb{R}) \sim \text{sl}(2,\mathbb{R}) \sim \text{o}(2,1) \sim \text{su}(1,1)$. The algebras $\text{sp}(n,F)$ and $\text{o}(p,q)$ lead to special cases of MRE's, while $\text{sl}(n,F)$ and $\text{su}(p,q)$ lead to more complicated types of equations. In the low-dimensional case of three equations, all of the above algebras $\mathscr{K}$ give equivalent results, namely a special case of the projective Riccati equations.

This is best seen by choosing $\mathscr{K} = \text{o}(p,q)$. Using the same techniques as in Ref. 13 for $\mathscr{K} = \text{o}(n,\mathbb{C})$ we obtain the $\text{O}(p,q)\times\text{O}(p,q)$ Riccati equations

$$\dot{V} = A + VB - BV - VAV, \quad A, B, V \in \mathbb{R}^{n\times n}, \quad \text{(A23)}$$
$$AI_{pq} + I_{pq}A^T = 0, \quad BI_{pq} + I_{pq}B^T = 0, \quad VI_{pq} + I_{pq}V^T = 0.$$

Setting $n = 3$ and $(p,q) = (3,0)$ or $(p,q) = (2,1)$ we put

$$V = \begin{pmatrix} 0 & x & y \\ -x & 0 & z \\ \epsilon y & \epsilon z & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ \epsilon b & \epsilon c & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & d & e \\ -d & 0 & f \\ \epsilon e & \epsilon f & 0 \end{pmatrix},$$

where $\epsilon = -1$ for $\mathscr{K} = 0(3)$ and $\epsilon = +1$ for $\mathscr{K} = \text{o}(2,1)$ and obtain

$$\begin{aligned} \dot{x} &= a + \epsilon fy - \epsilon ez + x(ax - \epsilon by - \epsilon cz), \\ \dot{y} &= b + fx - dz + y(ax - \epsilon by - \epsilon cz), \\ \dot{z} &= c - ex + dy + z(ax - \epsilon by - \epsilon cz), \end{aligned} \quad \text{(A24)}$$

clearly a special case of the PRE's.

Turning finally to the case with $\mathscr{K} = \text{sp}(2n,F)$ we obtain a system of symplectic matrix Riccati equations

$$\dot{V} = A + VB - BV - VAV, \quad \text{(A25)}$$

with

$$AK_0 + K_0A^T = 0, \quad BK_0 + K_0B^T = 0, \quad VK_0 + K_0V^T = 0,$$
$$K_0 = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

In particular, for $n = 1$ we set

$$V = \begin{pmatrix} x & y \\ z & -x \end{pmatrix}, \quad A = \begin{pmatrix} a_1 & a_2 \\ a_3 & -a_1 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & b_2 \\ b_3 & -b_1 \end{pmatrix}, \quad \text{(A26)}$$

and arrive at the SRE2 system studied in Sec. IV, cf. Eq. (4.4). While equivalent to (A24), this system is of sufficiently different form to merit a separate investigation.

Let us now summarize the results of this appendix by listing all systems of $n \leqslant 3$ real equations ($F = \mathbb{R}$) integrated in this paper.

(a) *$n = 1$:* The only single nonlinear ODE of first order with a superposition formula is the Riccati equation based on the action of $\text{sl}(2,\mathbb{R}) \sim \text{o}(2,1) \sim \text{sp}(2,\mathbb{R}) \sim \text{su}(1,1)$.

(b) *$n = 2$:* There are two types of indecomposable pairs of nonlinear ODE's with superposition formulas: (i) the projective Riccati equations (PRE's) (A3) based on $\text{sl}(3,\mathbb{R})$; and (ii) the conformal Riccati equations (CRE's) (A7) based on o(3,1).

(c) $n=3$: Four types of such indecomposable triplets exist: (i) the PRE's (A3) based on sl(4,$\mathbb{R}$); (ii) the o(4,1) CRE's (A7), with $\tilde{I}=I_3$, (iii) the o(3,2) CRE's (A7) with $\tilde{I}=I_{2,1}$; and (iv) the su(2,1) pseudounitary equations (A19) (for $p=2$) with polynomial nonlinearities of order 2,3, and 4. To these four, we add two more systems, which we integrate in Sec. IV, namely, the symplectic Riccati equation (A11) with $N=2$ and the sl(2,$\mathbb{R}$) $\oplus$ sl(2,$\mathbb{R}$) equations (4.4).

The situation is even simpler for $F=\mathbb{C}$. For $n=1$ we obtain one complex Riccati equation, whose real and imaginary parts when decoupled become two real o(3,1) CRE's. For $n=2$ the only indecomposable pair of ODE's are the SL(3,$\mathbb{C}$) PRE's [the O(4,$\mathbb{C}$) CRE's happen to be decomposable], while for $n=3$ we obtain SL(4,$\mathbb{C}$) PRE's and O(5,$\mathbb{C}$) CRE's.

[1]J. Guckenheimer and P. Holmes, "Nonlinear oscillations, dynamical systems and bifurcations of vector fields," in *Applied Mathematical Sciences* (Springer, Berlin, 1983), Vol. 42.
[2]M. Lieberman and A. Lichtenberg, "Regular and stochastic motion," in *Applied Mathematical Sciences* (Springer, Berlin, 1983), Vol. 38.
[3]V. I. Arnol'd, "Mathematical methods of classical mechanics," in *Graduate Texts in Mathematics* (Springer, Berlin, 1978), Vol. 60.
[4]R. H. G. Helleman, in *Fundamental Problems in Statistical Mechanics*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1981), Vol. 5.
[5]V. V. Kozlov, "Integrability and Nonintegrability in Hamiltonian systems," Russian Math. Surveys **38**, 1 (1983).
[6]M. A. Olshanetsky and A. M. Perelomov, Phys. Rep. **71**, 313 (1981).
[7]M. Adler and P. v. Moerbeke, Invent. Math. **67**, 297 (1982); see also Commun. Math. Phys. **83**, 85 (1982).
[8]T. Bountis, H. Segur, and F. Vivaldi, Phys. Rev. A **25**, 1257 (1982).
[9]A. Ramani, B. Grammaticos, and B. Dorizzi, J. Math. Phys. **24**, 2282, 2289 (1983).
[10]R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980).
[11]R. L. Anderson, J. Harnad, and P. Winternitz, Physica D **4**, 164 (1982).
[12]J. Harnad, P. Winternitz, and R. L. Anderson, J. Math. Phys. **24**, 1062 (1983).
[13]S. Shnider and P. Winternitz, J. Math. Phys. **25**, 3155 (1984); see also Lett. Math. Phys. **8**, 69 (1984).
[14]T. Bountis, A. Ramani, B. Grammaticos, and B. Dorizzi, Physica A **128**, 268 (1984).
[15]T. Bountis, M. Bier, and J. Hijmans, Phys. Lett. A **97**, 11 (1983).
[16]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[17]S. L. Ziglin, Trans. Moscow Math. Soc. **1**, 283 (1982).
[18]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[19]*Nonlinear Dynamics*, edited by R. H. G. Helleman (New York Academy of Sciences, New York, 1980), Vol. 357.
[20]J-P. Eckmann, Rev. Mod. Phys. **53**, 643 (1981).
[21]S. Lie and G. Scheffers, *Vorlesungen über Continuierlichen Gruppen mit geometrischen and anderen Anwendungen* (Teubner, Leipzig, 1893).

Symmetry reduction for the Kadomtsev–Petviashvili equation using a loop algebra

D. David, N. Kamran, D. Levi,[a] and P. Winternitz
Centre de Recherches Mathématiques, Université de Montréal, C.P. 6128, Succ. A, Montréal, Québec, Canada H3C 3J7

(Received 29 August 1985; accepted for publication 26 November 1985)

The Kadomtsev–Petviashvili (KP) equation $(u_t + 3uu_x/2 + \frac{1}{4}u_{xxx})_x + 3\sigma u_{yy}/4 = 0$ allows an infinite-dimensional Lie group of symmetries, i.e., a group transforming solutions amongst each other. The Lie algebra of this symmetry group depends on three arbitrary functions of time "t" and is shown to be related to a subalgebra of the loop algebra $A_4^{(1)}$. Low-dimensional subalgebras of the symmetry algebra are identified, specifically all those of dimension $n \leqslant 3$, and also a physically important six-dimensional Lie algebra containing translations, dilations, Galilei transformations, and "quasirotations." New solutions of the KP equation are obtained by symmetry reduction, using the one-dimensional subalgebras of the symmetry algebra. These solutions contain up to three arbitrary functions of t.

I. INTRODUCTION

The Kadomtsev–Petviashvili equation[1]

$$\Omega^\sigma(t,x,y;u) \equiv (u_t + \tfrac{3}{2}uu_x + \tfrac{1}{4}u_{xxx})_x + \tfrac{3}{4}\sigma u_{yy} = 0, \quad \sigma = \pm 1, \tag{1.1}$$

sometimes called the two-dimensional Korteweg–de Vries equation, is of considerable importance, both in physics and mathematics. It arises in the study of long gravity waves in a single layer, or multilayered shallow fluid, when the waves propagate predominantly in one direction with a small perturbation in the perpendicular one.[2–6] It also arises naturally in many other applications, particularly in plasma physics, gas dynamics, and elsewhere.

The mathematical interest of this equation stems from the fact that it is, in a well-defined sense, the generic member of a class of integrable partial differential equations, associated with certain infinite-dimensional Lie algebras and groups.[7,8] It is one of the few equations in more than 1 + 1 dimensions that is integrable in the sense of allowing a Lax pair, an infinity of conservation laws, soliton and multisoliton solutions, a family of analytic periodic, and quasiperiodic solutions, and having many other interesting properties.[9–11]

The purpose of this article is to study the symmetry group of the Kadomtsev–Petviashvili equation, i.e., the Lie group G of transformations acting on the independent variables (t,x,y) and the dependent variable u, such that, whenever $u(t,x,y)$ is a local solution of (1.1), then $u' = [g \circ u](t',x',y')$ is a solution for all $g \in G$ such that the function $g \circ u$ is defined.

Algorithms for calculating the invariance group of an equation or system of equations are well known.[12–17] For a good summary of this we refer to Olver[16,17] and also mention that computer programs using REDUCE,[18] MACSYMA,[19] or other symbolic manipulation systems exist that greatly facilitate the calculation of the symmetry group of a system of differential equations. This approach actually yields the Lie algebra of the symmetry group.

For the Kadomtsev–Petviashvili (KP) equation a straightforward application of the algorithm, using a REDUCE package has yielded an infinite-dimensional Lie algebra of symmetries,[20] which we will call KP symmetry algebra. Following the notations and results of Schwarz[20] (and correcting a slight misprint) we write a general element of this Lie algebra as

$$V = X(f) + Y(g) + Z(h), \tag{1.2}$$

$$X(f) = f(t)\partial_t + [\tfrac{1}{3}x\dot{f}(t) - \tfrac{2}{9}\sigma y^2 \ddot{f}(t)]\partial_x + \tfrac{2}{3}y\dot{f}(t)\partial_y - [(4\sigma/27)y^2 \dddot{f}(t) - \tfrac{2}{9}x\ddot{f}(t) + \tfrac{2}{3}u\dot{f}(t)]\partial_u, \tag{1.3a}$$

$$Y(g) = g(t)\partial_y - \tfrac{2}{3}\sigma y\dot{g}(t)\partial_x - (4\sigma/9)y\ddot{g}(t)\partial_u, \tag{1.3b}$$

$$Z(h) = h(t)\partial_x + \tfrac{2}{3}\dot{h}(t)\partial_u, \tag{1.3c}$$

where $f(t)$, $g(t)$, and $h(t)$ are arbitrary functions in $C^\infty(\mathbb{R})$, and the dots indicate derivatives with respect to t.

In Sec. II we discuss the structure of the symmetry algebra and corresponding symmetry group of the KP equation (1.1), establish a Levi decomposition, and identifying the Lie algebra with basis given by (1.3) as a subalgebra of the loop algebra $A_4^{(1)}$, a Kac–Moody type of algebra.[21,22] We also discuss the physical meaning of the finite-dimensional algebra obtained by restricting $f(t)$, $g(t)$, and $h(t)$ to first-degree polynomials.

Section III is devoted to a classification of low-dimensional subalgebras of the KP algebra, namely those of dimension $n = 1,2$, and 3, into conjugacy classes under the adjoint action of the symmetry group of the KP equation (the group of inner automorphisms of the KP algebra). This is done mainly to elucidate the structure of the considered infinite-dimensional Lie algebra and to establish the applicability of tools developed for classifying subalgebras of finite-dimensional Lie algebras. We will only use one-dimensional subalgebras to perform a symmetry reduction. Three different conjugacy classes of such subalgebras exist. A reader interested only in solutions of the KP equation can safely skip most of Sec. III.

In Sec. IV we do indeed use the one-dimensional subalgebras of the KP algebra to reduce the KP equation to par-

[a] Permanent and present address: Dipartimento di Fisica, Universitá degli Studi di Roma "La Sapienza," Piazzale A. Moro 2-00185, Roma, Italy.

 0022-2488/86/051225-13$02.50

tial differential equations in two variables. These are the Boussinesq equation, a once-differentiated Korteweg–de Vries (KdV) equation, and a linear equation, respectively. An arbitrary single solution of the Boussinesq equation will provide a family of solutions of the KP equation, depending on three arbitrary functions. A solution of the KdV equation provides a family of solutions of the KP equation, depending on two arbitrary functions. The linear equation obtained in the third case is solved explicitly, yielding another family of solutions involving three arbitrary functions of the variable t. The physical meaning of the obtained solutions is discussed as well.

Finally, in Sec. V, we state some conclusions and the future outlook.

II. THE SYMMETRY GROUP OF THE KADOMTSEV–PETVIASHVILI EQUATION

A. The symmetry Lie algebra

The Lie algebra of the symmetry group of equation (1.1) was obtained by Schwarz[20] and has been recapitulated in the Introduction. By a symmetry operator, in this context, we mean a vector field

$$V \equiv \tau(t,x,y;u)\partial_t + \chi(t,x,y;u)\partial_x + \eta(t,x,y;u)\partial_y + \varphi(t,x,y;u)\partial_u \tag{2.1}$$

such that its fourth prolongation[16,17] satisfies

$$\mathrm{pr}^{(4)}V \circ \Omega^\sigma(t,x,y;u)\big|_{\Omega^\sigma(t,x,y;u)=0} = 0, \tag{2.2}$$

where

$$\Omega^\sigma \equiv (u_t + \tfrac{3}{2}uu_x + \tfrac{1}{4}u_{xxx})_x + \tfrac{3}{4}\sigma u_{yy} \tag{2.3}$$

and the label $\sigma = \pm 1$ distinguishes between the so-called KPI and KPII equations.[5] The operator $\mathrm{pr}^{(4)}V$ has the form

$$\mathrm{pr}^{(4)}V = V + \sum_i \varphi^{x_i}\partial_{u_{x_i}} + \sum_{i\leq j} \varphi^{x_ix_j}\partial_{u_{x_ix_j}} + \sum_{i\leq j\leq k} \varphi^{x_ix_jx_k}\partial_{u_{x_ix_jx_k}} + \sum_{i\leq j\leq k\leq l} \varphi^{x_ix_jx_kx_l}\partial_{u_{x_ix_jx_kx_l}}, \tag{2.4}$$

where we have put $x_1 = t$, $x_2 = x$, $x_3 = y$. The functions φ^{x_i}, etc. depend on x_i, u, u_{x_i}, etc. in a known manner[16] through the *a priori* unknown functions τ, χ, η, and φ. These functions are determined by requiring that Eq. (2.2) should be satisfied whenever we have $\Omega^\sigma = 0$. The result of this procedure is that the most general vector field (2.1) satisfying (2.2) is given by (1.2) and (1.3) of the Introduction.

The commutation relations for this Lie algebra are easy to obtain and are

$$\begin{aligned}
&[X(f_1),X(f_2)] = X(f_1\dot f_2 - \dot f_1 f_2),\\
&[X(f),Y(g)] = Y(f\dot g - \tfrac{2}{3}\dot f g),\\
&[X(f),Z(h)] = Z(f\dot h - \tfrac{1}{3}\dot f h),\\
&[Y(g_1),Y(g_2)] = \tfrac{2}{3}\sigma Z(\dot g_1 g_2 - g_1\dot g_2),\\
&[Y(g),Z(h)] = 0,\\
&[Z(h_1),Z(h_2)] = 0.
\end{aligned} \tag{2.5}$$

Notice that we obtain a Lie algebra only if the functions $f(t)$, $g(t)$, and $h(t)$ are real functions of class C^∞ on some open subset of $\mathbb{R}$. Notice also that the algebra $L = \{X(f),Y(g),Z(h)\}$ admits a Levi decomposition[23]

$$L = S \ni\!\!\!+ N, \tag{2.6}$$

where

$$N = \{Y(g),Z(h)\} \tag{2.7}$$

is a solvable (and actually nilpotent) ideal in L, namely the nilradical (maximal nilpotent ideal) and

$$S = \{X(f)\} \tag{2.8}$$

is a simple Lie algebra. To convince ourselves that S is simple it suffices to note that it is isomorphic to the algebra

$$\mathbf{J}(\mathbb{R}) = \{f(\xi)\partial_\xi \,|\, f\in C^\infty(\mathbb{R})\}, \tag{2.9}$$

the algebra of real vector fields on $\mathbb{R}$. The algebra $\mathbf{J}(\mathbb{R})$ is simple according to a theorem of Cartan (Ref. 24, Theorem XI, p. 129). The isomorphism is given by the mapping

$$\psi:\ \mathbf{J}(\mathbb{R}) \to S,\qquad f(\xi)\partial_\xi \mapsto X[f(t)]. \tag{2.10}$$

B. The symmetry group of the KP equation

The infinitesimal symmetries $V = X(f) + Y(g) + Z(h)$ given in (1.2) and (1.3) can be integrated to yield the identity component of the group of finite transformations of the KP equation.

Let us start with the simplest case when $f = g = 0$ and $h\in C^\infty(\mathbb{R})$ is arbitrary. We have

$$\frac{dx'}{d\lambda} = h(t'),\quad \frac{dy'}{d\lambda} = 0,\quad \frac{dt'}{d\lambda} = 0,\quad \frac{du'}{d\lambda} = \frac{2}{3}\dot h(t'). \tag{2.11}$$

Integrating and requiring that for $\lambda = 0$ we obtain the identity transformation, we find

$$\begin{aligned}
&t' = t,\quad x' = x + \lambda h(t),\quad y' = y,\\
&u'(t',x',y') = u[t',x' - \lambda h(t'),y'] + \tfrac{2}{3}\lambda\dot h(t').
\end{aligned} \tag{2.12}$$

Thus, if $u(t,x,y)$ solves (1.1), then $u'(t',x',y')$ solves the same equation in the variables t',x',y'.

Now consider the case $f(t) = 0$, $g(t) \neq 0$. We have

$$\begin{aligned}
&\frac{dx'}{d\lambda} = -\frac{2}{3}\sigma y'\dot g(t') + h(t'),\quad \frac{dy'}{d\lambda} = g(t'),\\
&\frac{dt'}{d\lambda} = 0,\quad \frac{du'}{d\lambda} = -\frac{4\sigma}{9}y'\ddot g(t') + \frac{2\dot h(t')}{3}.
\end{aligned} \tag{2.13}$$

Integrating and imposing the appropriate initial condition for $\lambda = 0$, we have

$$
\begin{aligned}
&t'=t, \quad x'=x-\lambda\,[\tfrac{2}{3}\sigma\dot g(t)y-h(t)]-\tfrac{1}{3}\sigma g(t)\dot g(t)\lambda^2, \quad y'=y+\lambda g(t),\\
&u'(t',x',y')=u[t',x'+\tfrac{2}{3}\lambda\sigma\dot g(t')\,y'-\lambda h(t'),y'-\lambda g(t')]\\
&\qquad\qquad+\left[\frac{2\dot h(t')}{3}-\frac{4\sigma}{9}\ddot g(t')(y'-\lambda g(t'))\right]\lambda-\frac{2\sigma}{9}g(t')\ddot g(t')\lambda^2.
\end{aligned} \tag{2.14}
$$

Finally, consider the generic case when $f(t)\neq 0$, with $g(t)$ and $h(t)$ being arbitrary. In order to obtain the finite transformation we must integrate the equations

$$
\begin{aligned}
\frac{dt'}{f(t')}&=\frac{9dx'}{3x'\dot f(t')-2\sigma y'^2\ddot f(t')+9h(t')-6\sigma y'\dot g(t')}=\frac{3dy'}{2y'\dot f(t')+3g(t')}\\
&=\frac{-27du'}{4\sigma y'^2\dddot f(t')-6x'\ddot f(t')+18u'\dot f(t')-18\dot h(t')+12\sigma y'\ddot g(t')}=d\lambda.
\end{aligned} \tag{2.15}
$$

Introducing the notations

$$
\begin{aligned}
&\Phi(t)=\int_{t_0}^{t}\frac{ds}{f(s)}, \quad G(t',t)=[f(t)]^{2/3}\int_t^{t'}g(s)f^{-5/3}(s)ds,\\
&H(t',t)=[f(t)]^{1/3}\int_t^{t'}[\tfrac{2}{3}\sigma f^{-7/3}(s)g^2(s)+h(s)f^{-4/3}(s)]ds,
\end{aligned} \tag{2.16}
$$

we can integrate (2.15) and obtain

$$
\begin{aligned}
&t'(t)=\Phi^{-1}(\lambda+\Phi(t)),\\
&y'(t,y)=[y+G(t'(t),t)]\left[\frac{f(t'(t))}{f(t)}\right]^{2/3},\\
&x'(t,x,y)=\Bigg\{x-\frac{2\sigma y^2[\dot f(t'(t))-\dot f(t)]}{9f(t)}-\frac{2\sigma y}{9f(t)}\left[2G(t'(t),t)\dot f(t'(t))-3g(t)+3g(t'(t))\left[\frac{f(t)}{f(t'(t))}\right]^{2/3}\right]\\
&\qquad-\frac{2\sigma}{9f(t)}\left[G(t'(t),t)^2\dot f(t'(t))+3G(t'(t),t)g(t'(t))\left[\frac{f(t)}{f(t'(t))}\right]^{2/3}\right]+H(t'(t),t)\Bigg\}\left[\frac{f(t'(t))}{f(t)}\right]^{1/3},
\end{aligned} \tag{2.17a}
$$

$$
\begin{aligned}
u'(t',x',y')=\left[\frac{f(t(t'))}{f(t')}\right]^{2/3}\Bigg\{&u[t(t'),x(t',x',y'),y(t',y')]-\frac{2x'[\dot f(t(t'))-\dot f(t')]}{9f(t(t'))^{2/3}f(t')^{1/3}}\\
&-\frac{4\sigma y'}{81f(t(t'))^{4/3}f(t')^{2/3}}[2G(t',t)[3f(t(t'))\ddot f(t(t'))-\dot f(t(t'))^2]\\
&-9f(t(t'))\dot g(t(t'))+3g(t(t'))\dot f(t(t'))+3g(t')\dot f(t(t'))+9f(t')\dot g(t')-6g(t')\dot f(t')]\\
&-\frac{4\sigma(y')^2}{81f(t(t'))^{2/3}f(t')^{4/3}}[3f(t')\ddot f(t')-2\dot f(t')^2+\dot f(t')\dot f(t(t'))+\dot f(t(t'))^2-3f(t(t'))\ddot f(t(t'))]\\
&+\frac{2\dot f(t(t'))}{9f(t(t'))}H(t(t'),t')+\frac{4\sigma G^2(t(t'),t')}{81f(t')^2}[3f(t(t'))\ddot f(t(t'))-\dot f^2(t(t'))]\\
&-\frac{4\sigma G(t(t'),t')}{27f(t(t'))^2}[3f(t(t'))\dot g(t(t'))-g(t(t'))\dot f(t(t'))]\\
&+\frac{2}{3}\left[\frac{h(t')}{f(t')^{1/3}f(t(t'))^{2/3}}-\frac{h(t')}{f(t(t'))}\right]+\frac{2\sigma}{9}\left[\frac{g^2(t(t'))}{f(t')^{4/3}f(t(t'))^{2/3}}-\left[\frac{g(t(t'))}{f(t(t'))}\right]^2\right]\Bigg\},
\end{aligned}
$$

where Φ^{-1} denotes the (local) inverse function of Φ and where

$$
\begin{aligned}
&t(t')=\Phi^{-1}(-\lambda+\Phi(t')),\\
&y(t',y')=[y'+G(t(t'),t')]\left[\frac{f(t(t'))}{f(t')}\right]^{2/3},\\
&x(t',x',y')=\Bigg\{x'-\frac{2\sigma y'^2[\dot f(t(t'))-f(t')]}{9f(t')}-\frac{2\sigma y'}{9f(t')}\left[2G(t(t'),t')\dot f(t(t'))-3g(t')+3g(t(t'))\left[\frac{f(t')}{f(t(t'))}\right]^{2/3}\right]\\
&\qquad-\frac{2\sigma}{9f(t')}\left[G(t(t'),t')^2\dot f(t(t'))+3G(t(t'),t')g(t(t'))\left[\frac{f(t')}{f(t(t'))}\right]^{2/3}\right]+H(t(t'),t')\Bigg\}\left[\frac{f(t(t'))}{f(t')}\right]^{1/3}.
\end{aligned} \tag{2.17b}
$$

Formulas (2.12), (2.14), and (2.17) give us new solutions (u') from known ones (u). In particular, if we take a trivial solution $u = 0$, then (2.17) provides a family of solutions depending on $f(t)$, $g(t)$, and $h(t)$.

It should be mentioned here that the KP equation (1.1) is also invariant under certain discrete transformations, not obtained by integrating the infinitesimal transformation (1.3). These are the reflections R_y and R_{tx} defined as

$$R_y:\quad t\mapsto t,\quad x\mapsto x,\quad y\mapsto -y,\quad u\mapsto u, \tag{2.18a}$$

$$R_{tx}:\quad t\mapsto -t,\quad x\mapsto -x,\quad y\mapsto y,\quad u\mapsto u. \tag{2.18b}$$

C. A finite-dimensional subalgebra of "physical" transformations

A systematic classification of finite-dimensional subalgebras of the KP algebra will be discussed below. Here we just point out that all the "obvious" physical symmetries of the KP equation are obtained by restricting the arbitrary functions $f(t)$, $g(t)$, and $h(t)$ to be first-order polynomials in t. Indeed, in obvious notations, we have

$$\begin{aligned} &T\equiv X(1)=\partial_t,\quad D\equiv X(t)=t\,\partial_t+\tfrac{1}{3}x\,\partial_x+\tfrac{2}{3}y\,\partial_y-\tfrac{2}{3}u\,\partial_u,\\ &Y\equiv Y(1)=\partial_y\quad R\equiv Y(t)=t\,\partial_y-\tfrac{2}{3}\sigma y\,\partial_x,\\ &X\equiv Z(1)=\partial_x,\quad B\equiv Z(t)=t\,\partial_x+\tfrac{2}{3}\partial_u. \end{aligned} \tag{2.19}$$

Thus T, Y, X generate translations, D generates dilations, R has some properties of a rotation, and of a Galilei boost in the y direction, and B yields a Galilei transformation in the x direction. Thus, integrating, e.g., the vector field R, we obtain

$$\begin{aligned} &x'=x-\tfrac{2}{3}\sigma(\lambda y+\tfrac{1}{2}\lambda^2 t),\quad t'=t,\\ &y'=y+\lambda t,\quad u'=u. \end{aligned} \tag{2.20}$$

This transformation has been extensively used, e.g., by Segur and Finkel[3] to "rotate" solutions of the Korteweg–de Vries equation into solutions of the KP equation. The dilation symmetry $D = X(t)$ has been used to generate similarity solutions of the KP equation.[25,26]

The operators (2.18) form a basis of a six-dimensional solvable Lie algebra $L_p = \{D, R, B, X, Y, T\}$. It has a five-dimensional nilpotent ideal (the nilradical) $N = \{R, B, X, Y, T\}$. The commutation relations for L_p are given in Table I. It is a simple matter to classify and construct all the subalgebras of L_p, using known classification methods.[27,28] We will not present the result here since we find a classification of all the low-dimensional subalgebras of the infinite-dimensional symmetry algebra of the KP equation to be both more interesting and more useful for performing symmetry reduction (see Secs. III and IV below).

TABLE I. The commutation relations for L_p.

	D	R	B	X	Y	T
D	0	$\frac{1}{3}R$	$\frac{2}{3}B$	$-\frac{1}{3}X$	$-\frac{2}{3}Y$	$-T$
R	$-\frac{1}{3}R$	0	0	0	$\frac{2}{3}\sigma X$	$-Y$
B	$-\frac{2}{3}B$	0	0	0	0	$-X$
X	$\frac{1}{3}X$	0	0	0	0	0
Y	$\frac{2}{3}Y$	$-\frac{2}{3}\sigma X$	0	0	0	0
T	T	Y	X	0	0	0

Another finite-dimensional algebra, not contained in L_p, that is of physical interest, is obtained by restricting $f(t)$ in $X(f)$ to quadratic polynomials. We obtain $X(1) = T$, $X(t) = D$, as in (2.19), and in addition

$$\begin{aligned} X(t^2)\equiv C=t^2\,\partial_t&+\tfrac{2}{3}(tx-\tfrac{2}{3}\sigma y^2)\partial_x\\ &+\tfrac{4}{3}ty\,\partial_y+\tfrac{4}{9}(x-3tu)\partial_u. \end{aligned} \tag{2.21}$$

The commutation relations are

$$[D,T]=-T,\quad [D,C]=C,\quad [T,C]=2D, \tag{2.22}$$

so that we have obtained the algebra sl(2,$\mathbb{R}$), with C generating a type of conformal transformation:

$$\begin{aligned} &t'=t(1-\lambda t)^{-1},\\ &y'=y(1-\lambda t)^{-4/3},\\ &x'=\left[x-\frac{4\sigma\lambda y^2}{3(1-\lambda t)}\right](1-\lambda t)^{-2/3},\\ &u'=\left\{u+\frac{4\lambda x}{1-\lambda t}+\frac{8\lambda y^2}{81t(1-\lambda t)^2}[2(1-\sigma)\right.\\ &\qquad\left.-(2-\sigma)\lambda t]\right\}(1-\lambda t)^{4/3}. \end{aligned} \tag{2.23}$$

D. Relation between the KP invariance algebra and the affine Lie algebras of Kac-Moody type

An interesting feature of the KP algebra (1.3) is that it contains a subalgebra that can be embedded into an affine loop algebra. Indeed, let us consider the subalgebra L_π of the KP algebra (1.3) obtained by restricting the functions f, g, and h to be Laurent *polynomials* in t. A basis for this subalgebra is provided by

$$\begin{aligned} X(t^n)=t^n\,\partial_t&+\left[\frac{nx}{3}t^{n-1}-\frac{2\sigma}{9}n(n-1)\,y^2t^{n-2}\right]\partial_x\\ &+\frac{2n}{3}yt^{n-1}\,\partial_y-\left[\frac{4\sigma}{27}n(n-1)(n-2)\,y^2t^{n-3}\right.\\ &\left.-\frac{2n}{9}(n-1)xt^{n-2}+\frac{2n}{3}ut^{n-1}\right]\partial_u, \end{aligned} \tag{2.24}$$

$$\begin{aligned} &Y(t^n)=t^n\,\partial_y-\frac{2\sigma}{3}nyt^{n-1}\,\partial_x-\frac{4\sigma}{9}n\,(n-1)\,yt^{n-2}\,\partial_u,\\ &Z(t^n)=t^n\,\partial_x+\frac{2n}{3}t^{n-1}\,\partial_u, \end{aligned}$$

where $n\in\mathbf{Z}$.

The commutation relations of this subalgebra are [see Eq. (2.5)]

$$\begin{aligned} &[X(t^n),X(t^m)]=(m-n)X(t^{n+m-1}),\\ &[X(t^n),Y(t^m)]=(m-\tfrac{2}{3}n)Y(t^{n+m-1}),\\ &[X(t^n),Z(t^m)]=(m-\tfrac{1}{3}n)Z(t^{n+m-1}),\\ &[Y(t^n),Y(t^m)]=\tfrac{2}{3}\sigma(n-m)Z(t^{n+m-1}),\\ &[Y(t^n),Z(t^m)]=[Z(t^n),Z(t^m)]=0. \end{aligned} \tag{2.25}$$

Let us now consider the eight-dimensional Lie algebra L_0 generated by the following vector fields:

$$\Delta:=x\,\partial_x+2y\,\partial_y-2u\,\partial_u,\quad Q:=y\,\partial_x,$$

$$Y:=\partial_y,\quad X:=\partial_x,$$
$$A:=-\sigma y^2\,\partial_x+x\,\partial_u,\quad S:=y\,\partial_u,\tag{2.26}$$
$$P:=y^2\,\partial_u,\quad U:=\partial_u.$$

The Lie algebra L_0 is solvable, its nilradical is spanned by $\{Y, A, P, Q, X, S, U\}$ and it contains a five-dimensional Abelian ideal spanned by $\{P, Q, X, S, U\}$. It should be noted that L_0 is *not* a subalgebra of the KP equation.

The solvable Lie algebra L_0 can be embedded into a simple Lie algebra . The simple Lie algebra of lowest dimension that contains a five-dimensional Abelian subalgebra is A_4 (in Cartan's classification), in particular sl(5,$\mathbb{R}$) in our case. Indeed, it is easy to verify that the traceless matrices

$$\xi:=\begin{bmatrix}\delta & -a & \sigma p & s & u\\ 0 & 0 & -a & q & x\\ 0 & 0 & -\delta & -2\sigma y & 0\\ 0 & 0 & 0 & -\frac{1}{3}\delta & -y\\ 0 & 0 & 0 & 0 & \frac{1}{3}\delta\end{bmatrix}\tag{2.27}$$

provide a representation of the Lie algebra L_0 with the prescription that the matrix representing Δ is obtained by setting $\delta=1$ and all other entries equal to zero in ξ, similarly for Y, etc. We see that the Abelian subalgebra spanned by $\{P, Q, X, S, U\}$ is contained in a maximal Abelian subalgebra of sl(5,$\mathbb{R}$) with Kravchuk signature (2,0,3) (see Refs. 29–31).

Let us now establish a natural grading on L_0 by attributing the degree "n" to a monomial t^n and the degree μ $(0\leqslant\mu\leqslant4)$, equal to the distance from the diagonal in (2.27) to elements of the algebra (2.26). Thus Δ has degree 0, A and Y degree 1, P and Q degree 2, S and X degree 3, and U has degree 4 [the usual grading in the weight space of sl(5,$\mathbb{R}$)].

We now construct a loop algebra out of (2.26) following the procedure usually applied to simple Lie algebras.[8] Thus, putting

$$X(t^n)=\tfrac{1}{3}nt^{n-1}\Delta+\tfrac{2}{9}n(n-1)t^{n-2}A$$
$$\qquad-(4\sigma/27)n(n-1)(n-2)t^{n-3}P+t^n\,\partial_t,$$
$$Y(t^n)=t^nY-\tfrac{2}{3}\sigma nt^{n-1}Q-(4\sigma/9)n(n-1)t^{n-2}S,$$
$$Z(t^n)=t^nX+\tfrac{2}{3}nt^{n-1}U,\tag{2.28}$$

we see that the vector fields $X(t^n)$, $Y(t^n)$, and $Z(t^n)$ form a Lie algebra isomorphic to the subalgebra L_π of the KP symmetry algebra whose commutation relations are given by (2.25). Each element has a well-defined degree in the grading, namely $n-1$, $n+1$, and $n+3$ for $X(t^n)$, $Y(t^n)$, and $Z(t^n)$, respectively. From the embedding constructed above for L_0 into sl(5,$\mathbb{R}$) and from the representation given by (2.28) for the Lie algebra L_π, we see that L_π is a subalgebra of the affine loop algebra $A_4^{(1)}$ defined by

$$A_4^{(1)}:=\{\mathbb{R}[t,t^{-1}]\otimes\mathrm{sl}(5,\mathbb{R})\}\oplus\mathbb{R}[t,t^{-1}]\frac{d}{dt}.\tag{2.29}$$

The Levi decomposition (2.6) also holds for L_π. Indeed, from the commutation relation (2.25) we see that $N=\{Y(t^n), Z(t^n)\}$ forms a nilpotent ideal. The elements $X(t^n)$ form a Lie algebra S isomorphic to the $\mathbb{Z}$- graded algebra $\delta:=\mathbb{R}[t,t-{}^1]d/dt$. A basis for δ is given by the collection of derivations $(d_k)_{k\in\mathbb{Z}}$ defined by $d_k=t^kd/dt$ with commutation relations

$$[d_j,d_k]=(k-j)d_{j+k-1}.\tag{2.30}$$

As we restrict ourselves to Laurent polynomials, i.e., functions with only a finite number of nonzero coefficients in their Laurent expansion, it follows directly from the commutation relation (2.30) that δ is simple, i.e., it admits no nontrivial ideal.

Let us finally remark that the link between δ and the algebra of regular vector fields on S^1 has recently been investigated by Goodman and Wallach,[32] who also study the Virasoro algebra $\mathfrak{D}$, i.e., the universal central extension of δ.

III. LOW-DIMENSIONAL SUBALGEBRAS OF THE KP SYMMETRY ALGEBRA

In order to obtain solutions of the KP equation by symmetry reduction, we need to know the low-dimensional subalgebras of the KP symmetry algebra. More specifically, we need subalgebras that correspond to Lie groups having orbits of codimension 3, 2, or 1 in the four-dimensional space coordinatized by $(t,x,y;u)$. We obtain all the required subalgebras and also derive a better understanding of the structure of the KP symmetry algebra by classifying all its one-, two-, and three-dimensional subalgebras into conjugacy classes under the adjoint action of the KP symmetry group.

A. Classification of the one-dimensional subalgebras of the KP symmetry algebra under the adjoint action of the KP symmetry group

In this subsection, we show that there are three conjugacy classes of one-dimensional subalgebras of the KP symmetry algebra under the adjoint action of the KP symmetry group, with representatives spanned by $X(1)=T$, $Y(1)=Y$, and $Z(1)=X$, respectively.

The approach we take is similar to the one followed in the classification of the subalgebras of finite-dimensional Lie algebras. The difference between the finite- and infinite-dimensional cases arises in that one obtains differential conditions on the arbitrary functions labeling the group elements whose adjoint action is used to cast the generators of the subalgebras into normal forms, rather than algebraic conditions on the parameters labeling the elements of the finite-dimensional group.

We will use the explicit forms of the finite transformations of the variables (t,x,y,u) associated to the infinitesimal generators $X(F)$, $Y(G)$, and $Z(H)$ [$\exp\lambda X(F)$, $\exp\lambda Y(G)$, $\exp\lambda Z(H)$]. They are obtained, respectively, by setting $f=F$, $g=h=0$ in (2.16) and (2.17), $g=G$, $h=0$ in (2.14) and $h=H$ in (2.12).

There are three cases to be considered in the classification of the one-dimensional subalgebras of the KP symmetry algebra, generated by typical elements of the form

$$V=X(f)+Y(g)+Z(h),\tag{3.1}$$

into conjugacy classes under the adjoint action of the KP symmetry group.

Case A: $f\equiv0$, $g\equiv0$, and $h\neq0$. We claim that $V=Z(h)$ with $h\neq0$ can always be transformed into X by an element of the KP symmetry group. Actually, as suggested by the commutation relations (2.5) for the KP symmetry algebra, we can choose F so as to normalize h to 1 in $Z(h)$ by acting on it with $[\exp(\lambda X(F))]_*$ (the adjoint action). Indeed, we have

$$[\exp \lambda X(F)]_* Z(h) = h(t(t'))\left[\frac{F(t')}{F(t(t'))}\right]^{1/3}\frac{\partial}{\partial x'} + \left[\frac{2h(t(t'))[\dot F(t') - \dot F(t(t'))]}{9F(t(t'))^{1/3}F(t')^{2/3}} + \frac{2\dot h(t(t'))F(t(t'))}{3F(t')^{2/3}}\right]^{2/3}\frac{\partial}{\partial u'}. \quad (3.2a)$$

It has been shown by Neuman[33] that there exists a function F satisfying the relation

$$\left[\frac{F(t')}{F(t(t'))}\right]^{1/3} h(t(t')) = 1. \quad (3.2b)$$

Acting with $[\exp \lambda X(F)]_*$ on $Z(h)$ will normalize h to 1, as it is easily verified that (3.2b) and its differential consequences substituted into (3.2a) give, dropping primes,

$$[\exp \lambda X(F)]_* Z(h) = Z(1). \quad (3.2c)$$

The existence of a solution to (3.2b) may be argued as follows. Suppose that, for some $h \neq 0$, (3.2b) has no solution F. Then symmetry reduction by the corresponding $Z(h)$ will yield a reduced equation that cannot be equivalent under the action of any element of the KP symmetry group to the equation obtained by reducing the KP equation by $Z(1)$, namely the linear equation $u_{,yy} = 0$. But we will see in Sec. IV, where we construct all the solutions of the KP equation that are invariant under the action of a one-dimensional subgroup of the KP symmetry group having orbits of codimension 3 in the space coordinatized by (t,x,y,u), that symmetry reduction by any $V = Z(h)$ with $h \neq 0$ always gives rise to a reduced equation that is equivalent under the action of an element of the KP symmetry group to the linear equation $u_{,yy} = 0$. Thus we arrive at a contradiction, and therefore a solution to the functional equation (3.2b) must exist.

Case B: $f = 0$ and $g \neq 0$. We claim that $V = Y(g) + Z(h)$ with $g \neq 0$ always can be transformed into $Y(1)$ by an element of the KP symmetry group. Indeed, as suggested by the commutation relations (2.5) for the KP symmetry algebra, we can first choose G so as to transform h away from V by acting on it with $[\exp \lambda Y(G)]_*$. We have, dropping primes,

$$(\exp \lambda Y(G))_* V = \left[h - \frac{2\sigma}{3}y\dot g + \frac{2\sigma}{3}G\dot g\lambda - \frac{2\sigma}{3}\dot G g\lambda\right]\frac{\partial}{\partial x} + g\frac{\partial}{\partial y} + \left[\frac{2\dot h}{3} - \frac{4\sigma}{9}y\ddot g + \frac{4\sigma}{9}\lambda G\ddot g - \frac{4\sigma}{9}\lambda g\ddot G\right]\frac{\partial}{\partial u}. \quad (3.3a)$$

It is now straightforward to show that, if in (3.3a) we choose a function $G(t)$ defined by

$$G(t) := \frac{3\sigma}{2}ag\int_0^t (hg^{-2})(s)ds + cg, \quad (3.3b)$$

where $a \neq 0$ and c are arbitrary constants, as the function labeling the element $Y(G)$ of the KP symmetry algebra and $\lambda = a^{-1}$ as the value of the parameter λ along the one-parameter subgroup obtained by exponentiating $Y(G)$, we have

$$[\exp \lambda Y(G)]_* V = Y(g). \quad (3.3c)$$

Now, as again suggested by the commutation relations (2.5), we can choose F so as to normalize g to 1 in $Y(g)$ by acting on it with $[\exp \lambda X(F)]_*$. We have

$$\begin{aligned}
&[\exp \lambda X(F)]_* Y(g) \\
&= \left[\frac{-2\sigma}{3}\left[\frac{F(t(t'))}{F(t')}\right]^{1/3} y'\dot g(t(t')) + \frac{4\sigma g(t(t'))y'[\dot F(t(t')) - \dot F(t')]}{9F(t(t'))^{2/3}F(t')^{1/3}}\right]\frac{\partial}{\partial x'} \\
&\quad + g(t(t'))\left[\frac{F(t')}{F(t(t'))}\right]^{2/3}\frac{\partial}{\partial y'} \\
&\quad + \left[\frac{4\sigma F(t(t'))^{1/3}}{27F(t')^{4/3}}y'\dot g(t(t'))[\dot F(t(t')) - \dot F(t')] - \frac{8\sigma y' g(t(t'))}{81F(t(t'))^{2/3}F(t')^{4/3}}\right. \\
&\quad \times [3F(t')\ddot F(t') - 3F(t(t'))\ddot F(t(t')) + \dot F(t')\dot F(t(t')) + \dot F(t')^2 - 2\dot F(t(t'))^2] \\
&\quad \left. - \frac{4\sigma y' F(t(t'))^{4/3}\ddot g(t(t'))}{9F(t')^{4/3}}\right]\frac{\partial}{\partial u'}.
\end{aligned} \quad (3.4a)$$

Now, by Neuman's[33] result we know that, given $g(t) \neq 0$, there will always exist a function F satisfying

$$\left[\frac{F(t')}{F(t(t'))}\right]^{2/3} g(t(t')) = 1. \quad (3.4b)$$

It then follows that acting on $Y(g)$ with $[\exp \lambda X(F)]_*$ will normalize g to 1, as it is easily verified that (3.4b) and its differential consequences, when substituted into (3.4a), give

$$[\exp \lambda X(F)]_* Y(g) = Y(1). \quad (3.4c)$$

Case C: $f \neq 0$. We claim that $V = X(f) + Y(g) + Z(h)$ with $f \neq 0$ always can be transformed into $X(1)$ by an element of the KP symmetry group.

The main steps of the calculation are as follows. As suggested by the commutation relations (2.5) for the KP symmetry algebra we first can choose G so as to transform g away from V, by acting on it with $[\exp \lambda Y(G)]_*$. Indeed, it can be verified easily that if we take

$$G(t) := \left[c + a\int_0^t (gf^{-5/3})(s)ds\right] f(t)^{2/3}, \quad (3.5)$$

where $a \neq 0$ and c are arbitrary constants, as the function labeling the element $Y(G)$ of the KP symmetry algebra and $\lambda = -a^{-1}$ as the value of the parameter λ along the one-parameter subgroup obtained by exponentiating $Y(G)$, that we have

$$[\exp \lambda Y(G)]_* V = X(f) + Z(h). \quad (3.6)$$

The next step, which again is suggested by the commutation relations (2.5), is to choose the function H so as to transform h away from $X(f) + Z(h)$ by acting on it with $[\exp \lambda Z(H)]_*$. Again it is easy to check that if we take

$$H(t) := \left[c + a\int_0^t (hf^{-4/3})(s)ds\right] f(t)^{1/3}, \qquad (3.7)$$

where $a \neq 0$ and c are arbitrary constants, as the function labeling the element $Z(H)$ of the KP symmetry algebra and $\lambda = -a^{-1}$ as the value of the parameter λ along the one-parameter subgroup obtained by exponentiating $Z(H)$, we have

$$[\exp \lambda Z(H)]_*(X(f) + Z(h)) = X(f). \qquad (3.8)$$

The last step is to choose the function F so as to normalize f to 1 in $X(f)$ by acting on it with $[\exp \lambda X(F)]_*$. A tedious but otherwise straightforward calculation shows that if F is a solution of the equation

$$f(t(t'))F(t')/F(t(t')) = 1, \qquad (3.9a)$$

then

$$[\exp \lambda X(F)]_* X(f) = T. \qquad (3.9b)$$

The existence of a solution to (3.9a) for any $f \neq 0$ again follows from Neuman's result.

Our proof of the existence of three conjugacy classes of one-dimensional subalgebras of the KP symmetry algebra under the adjoint action of the KP symmetry group with representatives spanned by T, Y, and X is thus complete. To summarize, an arbitrary one-dimensional subalgebra of the KP algebra is conjugate, under the KP symmetry group, to precisely one of the following.

$$\mathfrak{L}_{1,1} = \{X(1)\}, \quad \mathfrak{L}_{1,2} = \{Y(1)\}, \quad \mathfrak{L}_{1,3} = \{Z(1)\}. \qquad (3.10)$$

B. Classification of the two-dimensional subalgebras of the KP symmetry algebra

Two types of two-dimensional Lie algebras $\{Y_1, Y_2\}$ exist, both over $\mathbb{R}$ and $\mathbb{C}$, namely Abelian algebras and solvable non-Abelian algebras, satisfying, in an appropriate basis, $[Y_1, Y_2] = Y_1$.

We will take Y_1 in one of the three possible forms, established above in Sec. III A, and let Y_2 be a general element of the KP algebra. We first impose the commutation relations, then simplify Y_2, using the isotropy group of Y_1 in the invariance group of the KP equation.

1. Abelian algebras

(1.1) $Y_1 = X(1) = T$.

We take $Y_2 = X(f) \;\; + Y(g) + Z(h)$. Requiring $[Y_1, Y_2] = 0$ and using (2.5) we find $\dot f = \dot g = \dot h = 0$. Hence $Y_2 = aX(1) + bY(1) + cZ(1)$. Replacing Y_2 by $Y_2' = Y_2 - aY_1$ we effectively set $a = 0$ in Y_2. Conjugating by $\exp[\lambda Y(t) + \mu Z(t)]$, if $b \neq 0$, we can arrange for $c \to 0$. If $b = 0$, then we put $Y_2 = Z(1)$. We thus obtain two distinct algebras $\{X(1), Y(1)\}$ and $\{X(1), Z(1)\}$.

(1.2) $Y_1 = Y(1) = Y$.

We again take $Y_2 = X(f) + Y(g) + Z(h)$. The condition $[Y_1, Y_2] = 0$ implies $\dot f = \dot g = 0$, hence $Y_2 = aX(1) + bY(1) + Z(h)$. We must put $a = 0$, or we would reobtain case (1.1). We put $b = 0$ by linear combination with Y_1 and obtain another algebra, namely $\{Y(1), Z(h)\}$. It should be noted that the remaining freedom in the KP symmetry group, namely the invariance of $\{Y(1)\}$ under dilations and time translations still could be used to give arbitrarily chosen values to any two of the Taylor coefficients of the function $h(t)$. Hereafter we choose not to lift such trivial redundancies for the equivalence classes of subalgebras labeled by arbitrary functions.

(1.3) $Y_1 = Z(1) = X$.

Requiring that Y_2 in its general form commutes with Y_1, we find $Y_2 = aX(1) + Y(g) + Z(h)$. If $a \neq 0$ we reobtain case (1.1). If $a = 0, g \neq 0$ we reobtain case (1.2). for $a = 0$, $g = 0$ we obtain a new algebra $\{Z(1), Z(h); \dot h \neq 0\}$.

2. Non-Abelian algebras

(2.1) $Y_1 = X(1)$.

The condition $[X(1), X(f) + Y(g) + Z(h)] = X(1)$ implies $\dot f = 1$, $\dot g = \dot h = 0$. Conjugating by $\exp[\lambda Y(1) + \mu Z(1)]$ we can transform $g \to 0$, $h \to 0$. We obtain a single algebra, namely $\{X(1), X(t)\}$.

(2.2) $Y_1 = Y(1)$.

Imposing the appropriate commutation relation we find $Y_2 = \frac{3}{2}X(t) + aX(1) + bY(1) + Z(h)$. We eliminate b by linear combination with Y_1 and transform a and h into 0 by conjugating by $\exp[\lambda X(1) + Z(H)]$. The new algebra is hence $\{Y(1), X(3t/2)\}$.

(2.3) $Y_1 = Z(1)$.

The commutation relation implies $Y_2 = X(3t) + aX(1) + Y(g) + Z(h)$. Conjugating by $\exp[\lambda X(1) + Y(G) + Z(H)]$ we transform $a \to 0$, $g \to 0$, $h \to 0$. The algebra that we obtain is $\{Z(1), X(3t)\}$.

Let us sum up the results. Every two-dimensional subalgebra of the KP algebra is conjugate under the invariance group of the KP equation to precisely one of the following algebras [with the reservation that any two functions $h(t)$ and $e^{\alpha}h(t - \beta)$, where α and β are constants, give equivalent algebras].

1. Abelian algebras:

$$\begin{aligned} &\mathfrak{L}_{2,1} = \{X(1), Y(1)\}, \\ &\mathfrak{L}_{2,2} = \{X(1), Z(1)\}, \\ &\mathfrak{L}_{2,3}{}^h = \{Y(1), Z(h)\}, \\ &\mathfrak{L}_{2,4}{}^h = \{Z(1), Z(h); \dot h \neq 0\}. \end{aligned} \qquad (3.11a)$$

2. Non-Abelian algebras satisfying $[Y_1, Y_2] = Y_1$:

$$\begin{aligned} &\mathfrak{L}_{2,5} = \{X(1), X(t)\}, \\ &\mathfrak{L}_{2,6} = \{Y(1), X(3t/2)\}, \\ &\mathfrak{L}_{2,7} = \{Z(1), X(3t)\}. \end{aligned} \qquad (3.11b)$$

C. Classification of the three-dimensional subalgebras of the KP symmetry algebra

A real three-dimensional Lie algebra can be either simple or solvable. We will consider these two cases separately.

1. Simple Lie algebras

Let us first allow for complex coefficients in the vector fields and construct the algebra sl(2,$\mathbb{C}$). This algebra has a two-dimensional non-Abelian subalgebra $\{Y_1, Y_2\}$. The sl(2,$\mathbb{C}$) commutation relations can be written as

$$[Y_1,Y_2]=Y_1, \quad [Y_2,Y_3]=Y_3, \quad [Y_1,Y_3]=2Y_2. \qquad (3.12)$$

We will identify $\{Y_1,Y_2\}$ with one of the algebras (3.11b), i.e., consider it to be already in standard form.

Let us start with

$$\mathfrak{L}_{2,5}: Y_1=X(1), \quad Y_2=X(t),$$
$$Y_3=X(f)+Y(g)+Z(h).$$

Imposing (3.12) we find $Y_3=X(t^2)$.

Next consider

$$\mathfrak{L}_{2,6}: Y_1=Y(1), \quad Y_2=X(3t/2),$$
$$Y_3=X(f)+Y(g)+Z(h).$$

It is easy to see that $[Y_3,Y_1]=2Y_2$ cannot be satisfied. Finally, consider

$$\mathfrak{L}_{2,7}: Y_1=Z(1), \quad Y_2=X(3t),$$
$$Y_3=X(f)+Y(g)+Z(h).$$

Again $[Y_3,Y_1]=2Y_2$ cannot be satisfied.

We thus have obtained a single class of sl(2,C) algebras, represented by $\{X(1),X(t),X(t^2)\}$ [see also (2.21) and (2.22)]. Restricting to real coefficients, we obtain sl(2,R), but not su(2).

2. Solvable Lie algebras

A solvable three-dimensional Lie algebra always will have a two-dimensional Abelian ideal (see, e.g., Refs. 34 and 35 for a classification of Lie algebras of dimension $n \leqslant 5$ into isomorphy classes). Unless the three-dimensional algebra is Abelian or nilpotent, this ideal is unique (up to conjugacy under inner automorphisms). We assume that the ideal $\{Y_1,Y_2\}$ is already in standard form ($\mathfrak{L}_{2,1},...,\mathfrak{L}_{2,4}{}^h$) and look for a third element $Y_3=X(f)=Y(g)+Z(h)$ that acts upon the ideal:

$$\begin{bmatrix}[Y_1,Y_3]\\ [Y_2,Y_3]\end{bmatrix}=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}Y_1\\ Y_2\end{bmatrix}, \quad M=\begin{bmatrix}a & b\\ c & d\end{bmatrix}. \qquad (3.13)$$

The real matrix M in (3.13) can, by change of the basis of the $\{Y_1,Y_2\}$ space, be taken into a standard form and further simplification can be achieved by rescaling Y_3. Finally, Y_3 can be simplified by transformations in the KP symmetry group that leave the algebra $\{Y_1,Y_2\}$, as a vector space, invariant.

Let us now run through this procedure for each two-dimensional Abelian subalgebra in our list (3.11a).

a. $\mathfrak{L}_{2,1}=\{X(1),Y(1)\}$

Imposing (3.13) and using the commutation relations (2.5) we obtain $d=2a/3$, $b=c=0$, and $Y_3=aX(t)+\alpha X(1)+\beta Y(1)+\gamma Z(1)$. Since $X(1)$ and $Y(1)$ are elements of the algebra we can put $\alpha=\beta=0$. If $a\neq 0$ we apply $\exp[\lambda Z(1)]$ to transform $\gamma\to 0$; we obtain a diagonal action in (3.13). If $a=0$ we take $\gamma=1$ and obtain the Abelian algebra $\{X(1),Y(1),Z(1)\}$.

b. $\mathfrak{L}_{2,2}=\{X(1),Z(1)\}$

From the commutation relations (2.5) and (3.13) we obtain $Y_3=aX(t)+bZ(t)+\gamma Y(1)$ [up to linear combinations with $X(1),Z(1)$]. If $a\neq 0$ we apply $\exp[\lambda Y(1)+\mu Z(t)]$ and transform $b\to 0$, $\gamma\to 0$; we obtain $Y_3=X(t)$ and a diagonal action in (3.13). If $a=0$ we must have $b\neq 0$ in order not to reobtain case 1 above. If $\gamma=0$ we obtain the nilpotent algebra $\{X(1),Z(1),Z(t)\}$. If $\gamma\neq 0$ we apply $\exp\lambda X(t)$ and the discrete symmetry (2.18a) to obtain another nilpotent algebra $\{X(1),Z(1),Z(t)+Y(1)\}$.

c. $\mathfrak{L}_{2,3}{}^H=\{Y(1),Z(H)\}$

The commutation relations imply $c=0, Y_3=aX(3t/2)+\alpha X(1)+Y(g)+Z(h)$, where

$$\dot g=-(3\sigma/2)bH, \quad (3at+2\alpha)\dot H+(2d-a)H=0. \qquad (3.14)$$

In this case we consider each normal form of the matrix M in (3.13) separately.

$$(3.1) \quad \begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}.$$

Then (3.14) yields the Abelian algebra $\{Y(1),Z(H),Z(h)\}$ [with $H(t)$ and $h(t)$ linearly independent] for $\alpha\neq 0$ and we reobtain $\{X(1),Y(1),Z(1)\}$ for $\alpha=0$.

$$(3.2) \quad \begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}0 & b\\ 0 & d\end{bmatrix}, \quad d\neq 0,$$

or

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}a & b\\ 0 & 0\end{bmatrix}, \quad a\neq 0.$$

In this case the action of Y_3 on the ideal is decomposable, but not Abelian. Consider first the case $a=0, d\neq 0$. Then $\alpha\neq 0$ and (3.14) implies

$$f=\alpha, \quad H=e^{-dt/\alpha}, \quad g=g_0+(3\sigma\alpha b/2d)e^{-dt/\alpha}.$$

Changing the basis in the ideal to $\{Y_1-(b/d)Y_2,Y_2\}$ we diagonalize M. Performing a conjugation by $\exp[\lambda Y(G)+\mu Z(K)]$ with appropriately chosen $G(t)$, $K(t)$, λ, and μ we obtain the decomposable algebra $\{Y(1),Z(e^{-t}),X(1)\}$. Similarly, putting $a\neq 0$, $d=0$, we obtain, after some rather tedious calculations, a further (inequivalent) decomposable algebra $\{Y(1),Z(t^{1/3}),X(3t/2)\}$.

(3.3) M nilpotent:

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}.$$

If $\alpha\neq 0$ we obtain $H=1, g=g_0-3\sigma t/2$. Performing an appropriate $\exp\lambda Z(K)$ transformation, we obtain a nilpotent algebra $\{Y(1),Z(1),X(1)-(3/2)\sigma Y(t)\}$. If $\alpha=0$, then $H=-\frac{2}{3}\sigma\dot g$ and g,h are arbitrary; we obtain a family of nilpotent algebras $\{Y(1),Z(-\frac{2}{3}\sigma\dot g),Y(g)+Z(h);\dot g\neq 0\}$. The dependence on $g(t)$ and $h(t)$ cannot be transformed away and each couple $g(t)$, $h(t)$ provides a different conjugacy class of algebras.

(3.4) M diagonalizable; $a\neq d$, $a\neq 0$, $d\neq 0$. We put $a=1$ and transform $\alpha\to 0$ (by a translation in t). Then $f=3t/2, 3t\dot H=(1-2d)H, \dot g=-3\sigma bH/2$. Conjugating appropriately by $\exp[\lambda Y(G)+\mu Z(K)]$ we obtain a one-parameter class of algebras $\{Y(1),Z[t^{(1-2d)/3}],X(3t/2)\}$.

Since $c=0$ in M, the only remaining possible form of M is a Jordan form.

$$(3.5) \quad \begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}.$$

This, after conjugating by $\exp \lambda Z(K)$, gives the algebra $\{Y(1), Z(t^{-2/3}), X(3t/2) - (9\sigma/2)Y(t^{1/3})\}$.

d. $\mathfrak{L}_{2,4}{}^{H} = \{Z(1), Z(H)\}$

The commutation relations (2.5) and (3.13) in this case imply

$$\tfrac{1}{3}\dot{f} = a + bH, \quad -f\dot{H} + \tfrac{1}{3}\dot{f}H = c + dH. \tag{3.15}$$

Moreover, we can take linear combinations of $Z(1)$ and $Z(H)$ that will take M into its standard form: $Z_1 = \alpha Z(1) + \beta Z(H)$, $Z_2 = \gamma Z(1) + \delta Z(H)$, with $\alpha\delta - \beta\gamma \neq 0$. A further conjugation by $\exp \lambda X(F)$ will then take Z_1 into $Z(1)$, Z_2 into $Z(H')$. We thus assume that M is already in standard form.

(4.1) $M = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

We obtain one new Abelian algebra, namely $\{Z(1), Z(h), Z(H)\}$ with 1, h, and H linearly independent.

(4.2) $M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

This leads to one new decomposable algebra: $\{Z(1), Z(t^{1/3}), Z(3t)\}$.

(4.3) $M = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$.

No new algebras are obtained.

(4.4) $M = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}$.

We find one class of algebras: $\{Z(1), Z(t^{(1-a)/3}), X(3t)\}$.

(4.5) $M = \begin{bmatrix} a & 1 \\ -1 & a \end{bmatrix}$.

This case corresponds to a "complex" action on the ideal (i.e., M could be diagonalized over $\mathbb{C}$ but not over $\mathbb{R}$). From (3.15) we see that we must have $f(t) \neq 0$. To solve (3.15) we depart from our usual procedure and first perform a transformation $\exp[X(F) + Y(G) + Z(K)]$ taking the elements Y_1, Y_2, and Y_3 into $X_1 = Z(h_1)$, $X_2 = Z(h_2)$, and $X_3 = Z(1)$, respectively, where h_1 and h_2 are arbitrary linearly independent functions. The commutation relations $[X_1, X_3] = aX_1 + X_2$ and $[X_2, X_3] = -X_1 + aX_2$ now imply

$$\dot{h} = -ah_1 - h_2, \quad \dot{h}_2 = h_1 - ah_2.$$

Solving and performing an appropriate time translation, we obtain the algebras $\{Z(e^{-at}\cos(t)), Z(e^{-at}\sin(t)), X(1)\}$.

(4.6) $M = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

We obtain a single class of Lie algebras, represented by $\{Z(1), Z(-\tfrac{1}{3}\ln(t)), X(3t)\}$.

To conclude this section we present in a unified manner a list of representatives of all conjugacy classes of three-dimensional subalgebras of the KP symmetry algebra, ordered by their isomorphy class. The solvable algebras are all given in the order $\{Y_1, Y_2, Y_3\}$, where $N = \{Y_1, Y_2\}$ is an Abelian ideal and the action of Y_3 on N is given in (3.13). In each case we specify the matrix M.

1. Abelian, $M = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$:

$$\begin{aligned} &\mathfrak{L}_{3,1} = \{X(1), Y(1), Z(1)\}, \\ &\mathfrak{L}_{3,2}{}^{h,H} = \{Y(1), Z[h(t)], Z[H(t)]\}, \quad h(t) \neq \lambda H(t), \\ &\mathfrak{L}_{3,3}{}^{h,H} = \{Z(1), Z[h(t)], Z[H(t)]\}, \end{aligned} \tag{3.16}$$

1, $h(t)$ and $H(t)$ linearly independent.

2. Decomposable, non-Abelian, $M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$:

$$\begin{aligned} &\mathfrak{L}_{3,4} = \{Z(e^{-t}), Y(1), X(1)\}, \\ &\mathfrak{L}_{3,5} = \{Y(1), Z(t^{1/3}), X(3t/2)\}, \\ &\mathfrak{L}_{3,6} = \{Z(1), Z(t^{1/3}), X(3t)\}. \end{aligned} \tag{3.17}$$

3. Nilpotent, $M = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$:

$$\begin{aligned} &\mathfrak{L}_{3,7} = \{Z(t), -Z(1), X(1)\}, \\ &\mathfrak{L}_{3,8} = \{Z(t) + Y(1), -Z(1), X(1)\}, \\ &\mathfrak{L}_{3,9}{}^{g,h} = \{Y(1), Z[-2\sigma\dot{g}(t)/3], Y[g(t)] \\ &\qquad + Z[h(t)]\}, \quad \dot{g}(t) \neq 0. \end{aligned} \tag{3.18}$$

4. Diagonal action on ideal, $M = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}$, $a \neq 0$:

$$\begin{aligned} &\mathfrak{L}_{3,10} = \{X(1), Y(1), X(-t)\}, \quad a = \tfrac{2}{3}, \\ &\mathfrak{L}_{3,11} = \{X(1), Z(1), X(-t)\}, \quad a = \tfrac{1}{3}, \\ &\mathfrak{L}_{3,12}{}^{a} = \{Y(1), Z[t^{(1-2a)/3}], X(3t/2)\}, \quad a \neq 0, \\ &\mathfrak{L}_{3,13}{}^{a} = \{Z(1), Z[t^{(1-a)/3}], X(3t)\}, \quad a \neq 0, \quad a \neq 1. \end{aligned} \tag{3.19}$$

5. Complex action on ideal, $M = \begin{bmatrix} a & 1 \\ -1 & a \end{bmatrix}$, $a \geqslant 0$:

$$\mathfrak{L}_{3,14}{}^{a} = \{Z[e^{-at}\cos(t)], Z[e^{-at}\sin(t)], X(1)\}, \quad a \geqslant 0. \tag{3.20}$$

6. Jordan action on ideal, $M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$:

$$\begin{aligned} &\mathfrak{L}_{3,15} = \{Y(1), Z(t^{-2/3}), X(3t/2) + Y(-9\sigma t^{1/3}/2)\}, \\ &\mathfrak{L}_{3,16} = \{Z(1), Z[-\tfrac{1}{3}\ln(t)], X(3t)\}. \end{aligned} \tag{3.21}$$

7. The simple Lie algebra sl$(2,\mathbb{R})$:

$$\mathfrak{L}_{3,17} = \{X(1), X(t), X(t^2)\} \quad [\text{see } (2.21),(2.22)]. \tag{3.22}$$

Thus all isomorphy classes of three-dimensional real Lie algebras except su(2) are represented in the list of subalgebras of the KP symmetry algebra.

It is easy to check in each case that there are no redundancies in the above list except for the possibility of "normalizing" one of the arbitrary functions in an algebra by the transformation $h(t) \mapsto e^{\alpha} h(t - \beta)$. The different Abelian algebras are mutually nonconjugate since a general element of $\mathfrak{L}_{3,1}$ can be conjugate to $X(1)$, $Y(1)$, or $Z(1)$, one of $\mathfrak{L}_{3,2}{}^{h,H}$ is conjugate to $Y(1)$ or $Z(1)$, an element of $\mathfrak{L}_{3,3}{}^{h,H}$ is always conjugate to $Z(1)$. The functions $h(t)$ and $H(t)$ cannot be changed in either case without destroying the standard form of $Y(1)$ or $Z(1)$, respectively. In the decomposable case the ideal $\{Y_1, Y_2\}$ and the two-dimensional solvable subalgebra $\{Y_1, Y_3\}$ are well defined and distinguish between the three cases. In the nilpotent case the center Y_3 is uniquely defined.

It distinguishes $\mathfrak{L}_{3,9}{}^{g,h}$ from the other two. An element of $\mathfrak{L}_{3,7}$ can be conjugate to $X(1)$ or $Z(1)$, an element of $\mathfrak{L}_{3,8}$ can be conjugate to $X(1)$, $Y(1)$, or $Z(1)$. In all other solvable cases the Abelian ideal is uniquely defined and suffices to distinguish between different mutually isomorphic cases.

IV. SOLUTIONS OF THE KADOMTSEV–PETVIASHVILI EQUATION OBTAINED BY SYMMETRY REDUCTION

In this section we will apply our knowledge of the KP symmetry group and its subgroups to construct *all* the solutions of the KP equation that are invariant under the action of a one-dimensional subgroup. In doing so we also complete the proof of the assertion in Sec. III A, namely that there exist precisely three orbits of one-dimensional subalgebras of the KP algebra. The one-dimensional subgroups have orbits of codimension 3 in the space coordinatized by (t,x,y,u). The method provides solutions that depend on three, two, or one arbitrary functions of the variable t, in addition to the arbitrary functions that may appear in the solutions of the reduced equations, which are themselves partial differential equations in two (rather than three) variables.

The method itself, called symmetry reduction, is very simple and well known.[12–16,36] It consists of taking a representative of a one-dimensional Lie algebra V and finding the scalar invariants of the corresponding one-dimensional Lie group $\exp \lambda V$. This amounts to finding a fundamental set of solutions of a first-order linear partial differential equation

$$VI(t,x,y;u)=0. \tag{4.1}$$

Solving the corresponding characteristic system we obtain two symmetry variables

$$\xi=\xi(t,x,y), \quad \eta=\eta(t,x,y) \tag{4.2}$$

and an expression for the solution

$$u(t,x,y)=\alpha(t,x,y)q(\xi,\eta)+\beta(t,x,y). \tag{4.3}$$

Here ξ, η, α, and β are explicitly known functions obtained by solving (4.1). On the other hand, $q(\xi,\eta)$ is *a priori* not known and is subject to a partial differential equation in ξ and η, obtained by substituting (4.3) back into the KP equation. The entire symmetry group then can be applied to the solution (4.3) to obtain a larger class of solutions. Two equivalent approaches can be adopted. One is to make use of the classification of one-dimensional subalgebras of the KP algebra, established in Sec. III A. We then go through the above procedure using the representatives of each conjugacy class of elements. Thus, $V=X(1)=\partial_t$ implies $\xi=x$, $\eta=y$, $\alpha=1$, $\beta=0$, and $u(x,y)$ is a solution of the Boussinesq equation. If $V=Y(1)=\partial_y$ we find $\xi=x$, $\eta=t$, $\alpha=1$, $\beta=0$, and $u(t,x)$ is a solution of the KdV equation, once differentiated with respect to x. If $V=Z(1)=\partial_x$ we obtain $\xi=t$, $\eta=y$, $\alpha=1$, $\beta=0$, and $u(t,y)$ satisfies the linear equation $u_{yy}=0$. In each of these cases we then apply the transformation (2.17) to get all solutions invariant under the action of a one-dimensional subgroup of the KP group. The solutions depend on up to three arbitrary functions.

A completely equivalent procedure is to perform the same reduction using a general element

$$V=X(f)+Y(g)+Z(h) \tag{4.4}$$

and, as usual, considering separately the three cases $f(t)\neq 0$, $f(t)=0$, and $g(t)\neq 0$, and $f(t)=g(t)=0$ but $h(t)\neq 0$. No further group transformation is necessary in this case.

We apply the second procedure, mainly because it will confirm the result established when classifying the one-dimensional subalgebras of the KP symmetry algebra; namely all the equations obtained when reducing by the generator of a one-dimensional subalgebra under a transformation of the KP symmetry group are equivalent either to the Boussinesq equation, a differentiated KdV equation, or a linear equation. Let us just list the result in each case.

Case 1. $f(t)\neq 0$:

$$\begin{aligned}
u(t,x,y)&=f^{-2/3}q(\xi,\eta)+\frac{2\dot f}{9f}x+\frac{4\sigma(2g\dot f-3f\dot g)y}{27f^2}\\
&\quad+\frac{4\sigma(2\dot f^2-3f\ddot f)y^2}{81f^2}+\frac{2\sigma g^2}{9f^2}+\frac{2h}{3f},
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
\xi&=\left[x+\frac{2\sigma g y}{3f}+\frac{2\sigma\dot f y^2}{9f}\right]f^{-1/3}\\
&\quad-\int_0^t\left[\frac{2\sigma}{3}g^2(s)f^{-7/3}(s)+h(s)f^{-4/3}(s)\right]ds,\\
\eta&=yf^{-2/3}-\int_0^t g(s)f^{-5/3}(s)ds.
\end{aligned}$$

Here $u(t,x,y)$ is a solution of the KP equation for any sufficiently smooth functions $f(t)\neq 0$, $g(t)$, and $h(t)$, if and only if $q(\xi,\eta)$ satisfies the Boussinesq equation

$$\sigma q_{\eta\eta}+(q^2)_{\xi\xi}+\tfrac{1}{3}q_{\xi\xi\xi\xi}=0. \tag{4.6}$$

Case 2. $f(t)=0$, $g(t)\neq 0$:

$$\begin{aligned}
u(t,x,y)&=g^{-1/2}q(\xi,\eta)+\frac{\dot g x}{3g}\\
&\quad+\frac{(2\dot h g-\dot g h)y}{3g^2}+\frac{\sigma(\dot g^2-2g\ddot g)y^2}{9g^2}-\frac{\sigma h^2}{2g^2},
\end{aligned} \tag{4.7}$$

$$\xi=g^{-1/2}\left[x-\frac{h}{g}y+\frac{\sigma\dot g}{3g}y^2\right],$$

$$\eta=\int_0^t g^{-3/2}(s)ds.$$

In this case $u(t,x,y)$ is a solution of the KP equation if and only if $q(\xi,\eta)$ satisfies the once-differentiated KdV equation

$$[q_\eta+\tfrac{3}{2}qq_\xi+\tfrac{1}{4}q_{\xi\xi\xi}]_\xi=0. \tag{4.8}$$

Case 3. $f(t)=g(t)=0$, $h(t)\neq 0$:

$$u(t,x,y)=q(t,y)+\frac{2\dot h}{3h}x-\frac{4\sigma\ddot h}{9h}y^2. \tag{4.9}$$

Here (4.9) solves the KP equation for any sufficiently smooth $h(t)$ if and only if q solves the linear equation

$$q_{yy}=0. \tag{4.10}$$

Integrating (4.10) we obtain a family of solutions of the KP equation depending on three arbitrary functions of t:

$$u(t,x,y)=\frac{2\dot h}{3h}x-\frac{4\sigma\ddot h}{9h}y^2+K(t)y+L(t). \tag{4.11}$$

The classes of solutions of the KP equation presented in

(4.5), (4.7), and (4.9) depend on three, two, or one arbitrary functions of t, in addition to the arbitrary functions possibly appearing in the solutions of the reduced equations. In general, they will diverge at infinity unless the functions f, g, and h are appropriately restricted.

We now proceed to list some special cases of the above solutions that are of physical interest and illustrate how f, g, and h ought to be chosen so as to preserve decay at infinity.

Boiti and Pempinelli[37] have shown that the similarity solutions of the potential Boussinesq equation

$$\sigma w_{\eta\eta} + (w^2_{\xi})_{\xi} + \tfrac{1}{3} w_{\xi\xi\xi\xi} = 0, \tag{4.12}$$

obtained by setting $q = w_{\xi}$ in the Boussinesq equation (4.6), are of the form

$$w(\xi,\eta) = \varphi\,[\xi - \sqrt{-\sigma/3}\,\eta] - (\sigma/6) g_1 \eta^2 + \sqrt{-\sigma/3}\,h_0 \eta + \xi/6 + k, \tag{4.13}$$

where g_1, h_0, and k are arbitrary constants. The equation satisfied by w reduces to an ordinary differential equation for φ, which is equivalent to the first Painlevé transcendent equation if $g_1 \neq 0$ and to the equation for the Weierstrass $\wp$-functions if $g_1 = 0$. It follows from (4.13) that

$$q = w_{\xi} = \tfrac{1}{6} + \varphi'[\xi - \sqrt{-\sigma/3}\,\eta] \tag{4.14}$$

satisfies the Boussinesq equation (4.6) and therefore that u, as given by (4.5), is a solution of the KP equation. From (4.5) we see that if we start with a solution of the Boussinesq equation, which is bounded at infinity, the corresponding solution of the KP equation given by (4.5) will have the same property only if f and g are constants, say $f_0 \neq 0$ and g_0, while h is arbitrary. We thus obtain two classes of solutions of the KP equation depending on one arbitrary function $h(t)$ by performing a "Galilei-like" transformation, involving an arbitrary function $h(t)$ on the solutions of the Boussinesq equation arising from similarity solutions of the potential Boussinesq equation. They are given by

$$\begin{aligned} u(t,x,y) &= -2(-g_1/2)^{2/5} f_0^{-2/3} P_I(\xi) + \frac{2\sigma g_0^2}{9 f_0^2} + \frac{2h(t)}{3 f_0} + \frac{1}{6}, \\ \xi &= \left(\frac{-g_1}{2}\right)^{1/5}\Bigg[x f_0^{-1/3} + \frac{2}{3}\sigma g_0 f_0^{-4/3} y - f_0^{-4/3}\int_0^t h(s)\,ds - \frac{2}{3}\sigma g_0^2 f_0^{-7/3} t \\ &\quad - \sqrt{\frac{-\sigma}{3}}(f_0^{-2/3} y - g_0 f_0^{-5/3} t) + \frac{g_2}{g_1}\Bigg], \end{aligned} \tag{4.15}$$

and

$$\begin{aligned} u(t,x,y) &= -2 f_0^{-2/3} \wp(\chi, g_2, g_3) + \frac{2\sigma g_0^2}{g f_0^2} + \frac{2h(t)}{3 f_0} + \frac{1}{6}, \\ \chi &= x f_0^{-1/3} + \frac{2}{3}\sigma g_0 f_0^{-4/3} y - f_0^{-4/3}\int_0^t h(s)\,ds - \frac{2}{3}\sigma g_0^2 f_0^{-7/3} t \\ &\quad - \sqrt{\frac{-\sigma}{3}}(f_0^{-2/3} y - g_0 f_0^{-5/3} t). \end{aligned} \tag{4.16}$$

[$P_I(\xi)$ is the first Painlevé transcendent, $\wp(\chi,g_2,g_3)$ the Weierstrass elliptic function.] Notice that certain restrictions must be imposed upon the constants f_0, g_0, g_1, g_2, and g_3 in order to obtain real solutions of the KP equation. In particular the above solutions only can be real for the KPII equation, i.e., when $\sigma = -1$. In addition, "lump"-type solutions of the KPII are obtained from (4.16) when $g_2 = g_3 = 0$, since we have $\wp(\chi,0,0) = \chi^{-2}$.

From (4.7) we see that if we start from a solution q of the differentiated KdV equation (4.8) that is bounded, the corresponding solution of the KP equation will share the same property if and only if g and h are constants, say $g_0 \neq 0$ and h_0. This solution is given by

$$\begin{aligned} u(t,x,y) &= g_0^{-1/2} q(\xi,\eta) - \tfrac{1}{3}\sigma h_0^2 g_0^{-2}, \\ \xi &= g_0^{-1/2}(x - h_0 y/g_0), \quad \eta = g_0^{-3/2} t. \end{aligned} \tag{4.17}$$

Solutions of the differentiated KdV equation (4.8) can thus be "rotated" into solutions of the KP equation. This property has been extensively used by Segur[2,3] in his construction of KP solutions of genus 1, which he obtains by rotating cnoidal wave solutions of the KdV equation according to (4.17). Of course, soliton and similarity solutions of the KdV equation also may be transformed into solutions of the KP equation, having physical significance. For example, the similarity solution of the KdV[38] equation

$$q(\xi,\eta) = -(\tfrac{3}{4}t)^{-2/3}[V_z(z,\mu) + V^2(z,\mu)], \tag{4.18}$$

where $z = 2^{2/3}\xi(3\eta)^{-1/3}$, and $V(z,\mu)$ is the second Painlevé transcendent satisfying

$$V_{zz}(z,\mu) = 2V^3(z,\mu) + zV(z,\mu) + \mu - \tfrac{1}{2}, \tag{4.19}$$

give rise to solutions of the KP equation via (4.17).

In view of the form of transformations (4.5) and (4.7) it is quite possible to obtain bounded solutions of the KP equation from solutions of the Boussinesq or KdV equations that diverge asymptotically. One way of obtaining such solutions is to perform a different choice of symmetry variables than the ones described above and thus to reduce to a different partial differential equation in two variables. We have proven that any choice is equivalent under the action of the KP symmetry group to one of the three choices discussed above. It is, however, possible for the group to transform bounded solutions of an equation into unbounded solutions of an equivalent one.

For example, let us choose the following symmetry variables:

$$\begin{aligned} \xi &= [g_1/3(g_0 + g_1 t)]^{1/3} \times\left[x - \frac{3\sigma h_0^2}{4 g_1(g_0 + g_1 t)} - \frac{h_0 y}{g_0 + g_1 t} + \frac{\sigma g_1 y^2}{3(g_0 + g_1 t)}\right], \\ \tau &= \tfrac{1}{3}\ln[(g_0 + g_1 t)/g_0], \end{aligned} \tag{4.20}$$

where g_0, g_1, and h_0 are constants [this choice is different from, but equivalent to, (4.7)]. It is easy to show that

$$u(t,x,y) = -[g_1/3(g_0 + g_1 t)]^{2/3}\vartheta(\xi,\tau) \tag{4.21}$$

will satisfy the KP equation if and only if $\vartheta(\xi,\tau)$ satisfies the following nonlinear evolution equation:

$$\vartheta_\tau + \vartheta_{\xi\xi\xi} - 6\vartheta\vartheta_\xi - 4\xi\vartheta_\xi - 2\vartheta = 0. \qquad (4.22)$$

Bounded solutions of this equation have been obtained by Calogero and Degasperis[11] using the inverse scattering method. They are given by

$$\vartheta(\xi,\tau) = \vartheta^*(\xi - z^*(\tau), \rho^*(\tau)), \qquad (4.23)$$

where

$$\vartheta^*(p,q) = 2q[2\,\mathrm{Ai}'(p)\mathrm{Ai}(p) + q\{\mathrm{Ai}(p)\}^4 G(p,q)]G(p,q), \qquad (4.24)$$

$$G(p,q) = (1 + q[\mathrm{Ai}'(p)]^2 - qp[\mathrm{Ai}(p)]^2)^{-1}, \qquad (4.25)$$

$$z^*(\tau) = z_0^* e^{-4\tau}, \quad \rho^*(\tau) = \rho_0^* e^{-4\tau}, \qquad (4.26)$$

and $\mathrm{Ai}(q)$ denotes an Airy function. The solutions of the KP equation defined by (4.20)–(4.26) contain the solutions obtained by Nakamura[26] as a special case.

We have shown that the use of one-dimensional subalgebras of the KP algebra makes it possible to generate large classes of solutions of the KP equation. For this particular equation the higher-dimensional subalgebras are of less use. Indeed, consider the two-dimensional subalgebras, all of which are listed in (3.11). Performing symmetry reduction with any of these we obtain a system of two linear first-order partial differential equations:

$$Y_1 I(t,x,y,u) = 0, \quad Y_2 I(t,x,y,u) = 0. \qquad (4.27)$$

Typically this system yields one symmetry variable ξ and an expression for the solution of the KP equation:

$$u(t,x,y) = \alpha(t,x,y)q(\xi) + \beta(t,x,y), \qquad (4.28)$$

where α, β, and $\xi(t,x,y)$ are explicitly known. Substituting (4.28) into the KP equation we obtain an ordinary differential equation for $q(\xi)$. The solution (4.28) then can be transformed by a general transformation of the KP symmetry group into a more general solution. However, one of the two operators in (4.27), say Y_1, will always coincide with one of those used above to reduce the KP equation to the Boussinesq equation, the KdV equation, or a linear equation. The other operator Y_2 then provides a further reduction. In other words, we do not obtain new solutions but particular cases of those discussed above.

V. CONCLUSIONS

The method of symmetry reduction for solving partial differential equations is certainly not new: it lies at the origin of Lie group theory itself.[39] Several new factors have recently emerged that should give new life to this old method.

The first is that many of the important nonlinear partial differential equations of modern physics turn out to have infinite-dimensional symmetry groups, the Lie algebras of which involve arbitrary functions. Such is the case of the Kadomtsev–Petviashvili equation treated in this article, but also of other integrable partial differential equations in 2 + 1 dimensions. Thus, the "modified potential Kadomtsev–Petviashvili" equation

$$[u_{xxx} - 2(u_x)^3 - 4u_t]_x - 6u_{xx}u_y + 3u_{yy} = 0, \qquad (5.1)$$

introduced by Jimbo and Miwa,[7] turns out to have such a symmetry group.[19] The symmetry group of the Davey–Stewartson equation[40] [the (2 + 1)-dimensional Schrödinger equation] is also infinite dimensional.[41]

The second factor is that methods have been developed for classifying subgroups of finite-dimensional Lie groups.[27,28,30,31] These methods can be generalized to infinite-dimensional Lie groups and we have seen in this article that it was not difficult to find representatives of all conjugacy classes of low-dimensional subalgebras of the KP symmetry algebra. A knowledge of the subgroups of the symmetry groups is essential for implementing the program of symmetry reduction.

The third new factor is the recent development of the theory of infinite-dimensional Lie groups and Lie algebras, in particular, Kac–Moody algebras and the realization of the important role they play in the study of integrable dynamical systems. It is of interest to note, as pointed out in Sec. II, that the symmetry algebra of the KP equation contains a loop algebra structure.

Our future plans include a treatment of other (2 + 1)-dimensional integrable equations and also an extension of the methods of this article to include more general transformations (contact transformations and Lie–Bäcklund transformations[15]). We also plan to consider the possibility of combining Lie symmetries and Bäcklund transformations for the KP equation. Finally, a more complete classification of the subgroups of the KP group would allow us to approach, in a systematic manner, the question of symmetry breaking for the KP equation, i.e., the construction of related equations, invariant under subgroups of the KP group, rather than under the entire group.

ACKNOWLEDGMENTS

Research on this project was partially supported by research grants from the Natural Sciences and Engineering Research Council of Canada and the "Fonds FCAR pour l'aide et le soutien à la recherche du Gouvernement du Québec." Among the authors, D. D. acknowledges an FCAR doctoral fellowship, N.K. an NSERC postdoctoral fellowship, D. L. thanks the CRM of Université de Montréal for its hospitality in Montréal, and P. W. thanks the Physics Department of the University of Rome "La Sapienza" for its hospitality in Rome.

We have benefitted from discussions with R. V. Moody, J. Patera, and Y. Saint-Aubin. Particular thanks are due to J. Aczel who posed the problem of proving the existence of solutions to the functional equations (3.2b), (3.4b), and (3.9a) at the 23rd International Symposium on Functional Equations held in Gragnano, Italy, in June 1985.

[1] V. V. Kadomtsev and V. I. Petviashvili, Sov. Phys. Dokl. **15**, 539 (1970).
[2] H. Segur, "Some physical applications of solitons," in *Nonlinear Phenomena, Lecture Notes in Physics*, Vol. 189 (Springer, New York, 1983).
[3] H. Segur and A. Finkel, ARAP technical memo 84-12, 1984.

[4]D. David, D. Levi, and P. Winternitz, preprint CRM-1280, Montréal, 1985.
[5]M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[6]R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons and Nonlinear Waves* (Academic, New York, 1982).
[7]M. Jimbo and T. Miwa, Publ. RIMS Kyoto **19**, 943 (1983).
[8]V. Kac, *Infinite Dimensional Lie Algebras* (Birkhäuser, Boston, 1984).
[9]M. J. Ablowitz and A. S. Fokas, "The Inverse Scattering Transform for Multidimensional $(2+1)$ Problems," in Ref. 2.
[10]S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons—the Inverse Method* (Plenum, New York, 1984).
[11]F. Calogero and A. Degasperis, *Spectral Transform and Solitons I* (North-Holland, Amsterdam, 1982).
[12]L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[13]G. W. Gluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York 1974).
[14]W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1972), Vol. II.
[15]R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[16]P. J. Olver, in D. H. Sattinger, *Group Theoretic Methods in Bifurcation Theory* (Springer, New York, 1979); P. J. Olver, J. Diff. Geom. **14**, 497 (1979).
[17]P. Winternitz, "Lie Groups and Solutions Nonlinear Differential Equations," in Ref. 2.
[18]F. Schwarz, Comput. Phys. Commun. **27**, 179 (1982); F. Schwarz, Computing **34**, 91 (1985).
[19]B. Champagne and P. Winternitz, preprint CRM-1278, Montréal, 1985.
[20]F. Schwarz, J. Phys. Soc. Jpn. **51**, 2387 (1982).
[21]R. V. Moody, J. Algebra **10**, 211 (1968).
[22]V. Kac, Izv. Akad. Nauk. SSSR Ser. Mat. **32**, 1923 (1968) [Math. USSR-Izv. **2**, 1271 (1968)].
[23]N. Jacobson, *Lie Algebras* (Wiley-Interscience, New York, 1962).
[24]E. Cartan, Ann. Ec. Norm. **26**, 93 (1909) [or see *Oeuvres complètes* (Editions du CNRS, Paris, 1984), Vol. II, p. 893].
[25]L. G. Redekopp, Stud. Appl. Math. **63**, (1980).
[26]A. Nakamura, J. Phys. Soc. Jpn. **51**, 19 (1982).
[27]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597, 1615 (1975).
[28]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).
[29]D. A. Suprunenko and R. I. Tyshkevich, *Commutative Matrices* (Academic, New York, 1968).
[30]P. Winternitz and H. Zassenhaus, preprint CRMA 1199, Montréal, 1983.
[31]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **24**, 1973 (1983).
[32]R. Goodman and N. R. Wallach, J. Reine Angew. Math. **273**, 178 (1984).
[33]F. Neuman, in *23rd International Symposium on Functional Equations*, Gragnano, Italy (Univ. Waterloo, Centre for Information Theory, Waterloo, to be published).
[34]G. M.Mubarakzyanov, Izv. Vyss. Ucebn. Zaved. Mat. **32**, 114 (1963); **34**, 99 (1963), **35**, 104 (1963).
[35]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[36]M. Grundland, J. Harnad, and P. Winternitz, J. Math. Phys. **25**, 791 (1984).
[37]M. Boiti and F. Pempinelli, Nuovo Cimento B **56**, 148 (1980).
[38]M. Boiti and F. Pempinelli, Nuovo Cimento B **51**, 70 (1979).
[39]S. Lie, Math. Ann. **25**, 71 (1885).
[40]A. Davey and K. Stewartson, Proc. R. Soc. London Ser. A **338**, 101 (1974).
[41]B. Champagne and P. Winternitz, "On the infinite dimensional symmetry groups of the Davy–Stewartson and modified Kadomtsev–Petviashvili equations," preprint CRM, Montréal, 1985 (to be published).

Solutions to a generalized spheroidal wave equation: Teukolsky's equations in general relativity, and the two-center problem in molecular quantum mechanics

E. W. Leaver
Department of Physics and the College of Science Computer, University of Utah, Salt Lake City, Utah 84112

(Received 9 July 1985; accepted for publication 3 January 1986)

The differential equation, $x(x-x_0)(d^2y/dx^2)+(B_1+B_2x)(dy/dx)+[\omega^2x(x-x_0)-[2\eta\omega(x-x_0)+B_3]y=0$, arises both in the quantum scattering theory of nonrelativistic electrons from polar molecules and ions, and, in the guise of Teukolsky's equations, in the theory of radiation processes involving black holes. This article discusses analytic representations of solutions to this equation. Previous results of Hylleraas [E. Hylleraas, Z. Phys. **71**, 739 (1931)], Jaffé [G. Jaffé, Z. Phys. **87**, 535 (1934)], Baber and Hassé [W. G. Baber and H. R. Hassé, Proc. Cambridge Philos. Soc. **25**, 564 (1935)], and Chu and Stratton [L. J. Chu and J. A. Stratton, J. Math. Phys. (Cambridge, Mass.) **20**, 3 (1941)] are reviewed, and a rigorous proof is given for the convergence of Stratton's spherical Bessel function expansion for the ordinary spheroidal wave functions. An integral is derived that relates the eigensolutions of Hylleraas to those of Jaffé. The integral relation is shown to give an integral equation for the scalar field quasinormal modes of black holes, and to lead to irregular second solutions to the equation. New representations of the general solutions are presented as series of Coulomb wave functions and confluent hypergeometric functions. The Coulomb wave-function expansion may be regarded as a generalization of Stratton's representation for ordinary spheroidal wave functions, and has been fully implemented and tested on a digital computer. Both solutions given by the new algorithms are analytic in the variable x and the parameters B_1, B_2, B_3, ω, x_0, and η, and are uniformly convergent on any interval bounded away from x_0. They are the first representations for generalized spheroidal wave functions that allow the direct evaluation of asymptotic magnitude and phase.

I. INTRODUCTION

Generalized spheroidal wave equations have been the topic of much applied mathematical research. They are usually characterized as being second-order linear differential equations having two regular singular points and one confluently irregular singular point. In this article the problem of generating general solutions to the specific equation

$$x(x-x_0)\frac{d^2y}{dx^2}+(B_1+B_2x)\frac{dy}{dx}+[\omega^2x(x-x_0)-2\eta\omega(x-x_0)+B_3]y=0 \qquad (1)$$

(where B_1, B_2, B_3, ω, η, and x_0 are constants) is approached from the point of view of the computational physicist. Equation (1) will hereafter be referred to as "the generalized spheroidal wave equation." The intervals of physical interest are both $[0\leqslant x\leqslant x_0]$ and $[x_0\leqslant x<\infty)$. Representations for solutions on the bounded interval $[0\leqslant x\leqslant x_0]$ are well understood, and are reviewed here only to illustrate properties of three-term recurrence relations. The purpose of this paper is to present new representations for solutions on the semi-infinite interval $[x_0\leqslant x<\infty)$.

The differential equation (1) arises in two specific physical contexts: the separation of the one-particle Schrödinger equation in prolate spheroidal coordinates, and the separation of linearized perturbation equations on the backgrounds of Schwarzschild and Kerr black holes. (Teukolsky's equations governing perturbations of the Kerr metric are generalized spheroidal wave equations.) This paper is an exposition on neither quantum mechanics nor general relativity, and the physics underlying these equations will be mentioned only in the context of boundary conditions relevant to the solutions.

Researchers in both astrophysics and molecular physics have long recognized the frequent inadequacy of numerical integration techniques in supplying satisfactory solutions to generalized spheroidal wave equations.[1-3] The original goal of this study was the development of analytic representations for solutions to Eq. (1) on the interval $[x_0\leqslant x<\infty)$ that would be useful in the investigation of resonance phenomena in low-energy molecular scattering processes. For that end, we sought a representation that was both analytic in the independent variable x and the parameters B_1, B_2, B_3, ω, x_0, and η, and from which the analytic behavior of the solutions as $x\to\infty$ could readily be inferred. The power of the resulting Coulomb wave-function expansion is demonstrated in an article on the spectral decomposition of the perturbation response of Schwarzschild black holes.[4] The present paper presents the new algorithm, and how we arrived at it.

In the process it reviews earlier work of Hylleraas, Jaffé, Baber and Hassé, Chu and Stratton, and Morse. These authors' results form a natural starting point for this study, which may be considered to be a continuation of their previous efforts, and are a seemingly forgotten topic in themselves. Review of their work is particularly worthwhile in view of

 0022-2488/86/051238-28$02.50

enduring misconceptions concerning the convergence properties of some of their representations.

Lastly, we have discussion of two representations that we have not yet used in computational problems, nor verified numerically. They are the second solutions of Jaffé's type presented in Sec. IV C, and the confluent hypergeometric function expansions of Sec. VII. The first of these (if it is correct) may eventually be of considerable computational utility. The second is more difficult to evaluate. The representations of which we have made extensive computational use are the regular Jaffé series discussed in Sec. IV A, and the Coulomb wave-function expansion presented in Sec. VI. The present (July 1985) computer implementation of these algorithms is discussed briefly in the Summary. The paper is outlined as follows.

Section II shows the equivalence of the separated parts of the one-particle Schrödinger equation in prolate spheroidal coordinates to the Teukolsky equations that describe the perturbations of the Weyl tensor for Kerr black holes. The angular and radial parts of both sets of equations are cast in the common form of Eq. (1), and solutions at the singular points $x = 0$, $x = x_0$, and $x = \infty$ are discussed.

Section III briefly reviews the theory of three-term recurrence relations and illustrates the usefulness thereof in generating spheroidal harmonics and in obtaining the eigenvalues of the angular differential equation on the interval $[0 \leqslant x \leqslant x_0]$. The origins of the method are lost in antiquity, and most of the material in this section is stolen from more recent articles by Baber and Hassé,[5] and Gautschi.[6]

Section IV turns to the study of solutions on the interval $[x_0 \leqslant x < \infty)$, and starts with a review of the eigensolutions of Hylleraas[7] and Jaffé.[8] Convergence properties of both representations are discussed in detail, and an integral relating the two is derived. Jaffé's solution is of critical importance, since it can be generalized to all values of the frequency ω, and provides solutions that are regular and analytic as $x \to x_0$. Section IV C contains a rather lengthy digression on the possibility of generating second solutions to the differential equation by means of a confluent hypergeometric function expansion related to the Laguerre polynomial expansion of Hylleraas. The resulting expressions have yet to be verified numerically.

Section V reviews Stratton's classic solution to the ordinary spheroidal wave equation, and generalizes Stratton's solution to the case of the Schrödinger's equation for an electron in the field of a finite dipole. Rigorous proofs of the convergence of the resulting spherical Bessel function expansions are discussed in detail, and form the basis for the full generalization in terms of Coulomb wave functions presented in Sec. VI. The discussion in Sec. V is important because it shows for the first time how analytic solutions may be constructed for a spheroidal wave equation in a space with a nonzero potential.

Section VI presents the ultimate result of this study: the expansion of solutions to the fully generalized spheroidal wave equation (1) in convergent series of Coulomb wave functions. The solutions provided by this representation are both irregular as $x \to x_0$, but are analytic in the operational sense that they allow the asymptotic (large x) behavior of any solution to the generalized spheroidal wave equation to be computed directly from the value of the solution and its derivative at any finite x greater than x_0. The algorithm has seen full computational implementation, and has been used to characterize the nature of the perturbation response of the Schwarzschild black hole to an appreciably greater extent than has previously been possible.[4] Sections V and VI may be read independently from Sec. IV.

Section VII presents another expansion for generalized spheroidal wave functions as series of confluent hypergeometric functions.

Section VIII looks at what happens to the generalized spheroidal wave equation and its Coulomb wave-function solutions (Sec. VI) in the confluence as $x_0 \to 0$. This happens at the extreme Kerr limit of black hole rotation, and concludes the present analysis of generalized spheroidal wave functions.

Section IX is a summary and contains a brief description of the computer programs that generate the Jaffé solutions and the Coulomb wave-function expansions.

Lastly, it has not been possible for the present paper to reference all the literature pertaining to spheroidal wave functions, much of which is due to the efforts of Miexner *et al.* The interested reader will find a comprehensive bibliography in their recent monograph.[9]

II. ORIGINS OF THE EQUATION AND ASYMPTOTIC SOLUTIONS

Generalized spheroidal wave equations are ordinary differential equations with two regular singular points and one confluently irregular singular point. Although the Helmholtz equation separates in spheroidal coordinates into particular, and special, examples of such equations (ordinary spheroidal wave equations),[10] the earliest physical context of a *generalized* spheroidal wave equation arose in the consideration of the quantum mechanics of hydrogen moleculelike ions. Early investigations into this subject are reviewed by Baber and Hassé,[5] and much of the discussion in this and the following two sections is excerpted from their article. Generalized spheroidal wave equations also result from the separation of linearized covariant wave equations on black hole background metrics, and the quasinormal modes of the perturbations of these geometries may be found by the same techniques used to determine the bound-state eigenfunctions of the hydrogen molecule ion.[11] This section explores the similarity of the differential equations in the astrophysical problem to corresponding differential equations in the molecular ion problem, and reduces them both to the form of Eq. (1).

A. Schrödinger equation for hydrogen moleculelike ions

If N_1 and N_2 are the charges on two fixed nuclei A and B, $2a$ is the distance AB between them, and r_1 and r_2 are the distances of an electron from A and B, respectively, then the prolate spheroidal coordinates λ and μ are defined by $\lambda = (r_1 + r_2)/2a$ and $\mu = (r_1 - r_2)/2a$. At large values of r_1 and r_2, λ becomes a simple measure of the distance from the molecule or ion, and is referred to as the "radial coordinate." Under the same conditions, μ reduces to the cosine of the usual polar angle θ, and μ is termed the "angular coordinate." The time-independent Schrödinger equation $\nabla^2\psi + (E - V)\psi = 0$ separates if $\psi = \Psi(\lambda)\Phi(\mu) \times \exp(im\phi)$, where ϕ is the azimuthal angle about the axis AB. A description of this separation is given in Eyring *et al.*[12] The resulting ordinary differential equations for Ψ and Φ are

$$\frac{d}{d\lambda}\left[(\lambda^2 - 1)\frac{d\Psi}{d\lambda}\right] + [\omega^2\lambda^2 + 2a(N_1 + N_2)\lambda - A_{lm} - m^2/(\lambda^2 - 1)]\Psi = 0, \tag{2}$$

and

$$\frac{d}{d\mu}\left[(1 - \mu^2)\frac{d\Phi}{d\mu}\right] + [-\omega^2\mu^2 - 2a(N_1 - N_2)\mu + A_{lm} - m^2/(1 - \mu^2)]\Phi = 0, \tag{3}$$

where $\omega^2 = 2a^2E$ in atomic units.

Equations (2) and (3) are generalized spheroidal wave equations. If we write $\Psi = (\lambda^2 - 1)^{m/2}f(\lambda)$ and $\Phi = (1 - \mu^2)^{m/2}g(\mu)$, the differential equations for f and g are

$$(\lambda^2 - 1)f_{,\lambda\lambda} + 2(m + 1)\lambda f_{,\lambda} + [\omega^2\lambda^2 + 2a(N_1 + N_2)\lambda + m(m + 1) - A_{lm}]f = 0, \tag{4}$$

and

$$(1 - \mu^2)g_{,\mu\mu} - 2(m + 1)\mu g_{,\mu} - [\omega^2\mu^2 + 2a(N_1 - N_2)\mu + m(m + 1) - A_{lm}]g = 0. \tag{5}$$

The form (1) is obtained if we let $x = \lambda + 1$ in Eq. (4), and $x = \mu + 1$ in Eq. (5):

$$x(x - 2)f_{,xx} + 2(m + 1)(x - 1)f_{,x} + [\omega^2x(x - 2) + 2a(N_1 + N_2)(x - 2) + \omega^2 + 2a(N_1 + N_2) + m(m + 1) - A_{lm}]f = 0, \tag{6}$$

$$x(x - 2)g_{,xx} + 2(m + 1)(x - 1)g_{,x} + [\omega^2x(x - 2) + 2a(N_1 - N_2)(x - 2) + \omega^2 + 2a(N_1 - N_2) + m(m + 1) - A_{lm}]g = 0. \tag{7}$$

Generalized spheroidal wave equations are characterized by two regular and one confluently irregular singular points. These occur at $x = 0$, $x = x_0$, and at $x = \infty$, respectively. For Eqs. (6) and (7) the regular singularities correspond to the physical locations of the two nuclei, which are at the foci of the coordinate system, $\lambda = 1$ and $\mu = \pm 1$. If

$$\lim_{x\to 0} y \sim x^{k_1}, \quad \text{and} \quad \lim_{x\to x_0} y \sim (x - x_0)^{k_2},$$

then the indices k_1 and k_2 take the values

$$k_1 = 0, \quad 1 + B_1/x_0, \quad \text{and} \quad k_2 = 0, \quad 1 - B_2 - B_1/x_0.$$

For Eqs. (6) and (7) these values are 0, $-m$ both for k_1 and for k_2.

B. Covariant wave equation on Schwarzschild and Kerr backgrounds

A separable linearized partial differential wave equation obeyed by components of weak electromagnetic and gravitational fields on the background geometry of the Schwarzschild black hole was derived through the efforts of Wheeler,[13] Regge and Wheeler,[14] Zerilli,[15] Chandrasekhar,[16] and Chandrasekhar and Detweiler.[17] Analysis of wave equations on the Kerr geometry of rotating black holes was provided by Teukolsky.[18] Generalized spheroidal wave equations result in each case.

1. Schwarzschild geometry

The Schwarzschild geometry is spherically symmetric, and the partial differential equation for the field components separates in polar spatial coordinates r, θ, and ϕ, and Schwarzschild's time coordinate t. These are the Schwarzschild coordinates.

Denote either a massless scalar field or a component of the electromagnetic or gravitational fields by a generic field function $\Phi(t,r,\theta,\phi)$. Fourier analyze and expand Φ in spherical harmonics as

$$\Phi(t,r,\theta,\phi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t} \times\left(\sum_l \frac{1}{r}\psi_l(r,\omega)Y_{l0}(\theta,\phi)\right)d\omega. \tag{8}$$

The homogeneous differential equation obeyed by the Fourier component $\psi_l(r)$ is

$$r(r - 1)\psi_{l,rr} + \psi_{l,r} + \left[\frac{\omega^2r^3}{r - 1} - l(l + 1) + \frac{s^2 - 1}{r}\right]\psi_l = 0, \tag{9}$$

where the coordinates t and r have been scaled so that the horizon, which usually appears at $r = 2M$, is now at $r = 1$. The parameter s is the spin of the field, and takes the values 0, 1, or 2 depending on whether Φ is, respectively, a component of the massless scalar, electromagnetic, or gravitational field.

The history of the derivation of these perturbation equations is long and rich. The derivation of the radial component of the scalar wave equation [$s = 0$ in Eq. (9)] on the Schwarzschild background is a straightforward exercise in perturbation theory.[19,20] The $s = 1$ equation for electromagnetic perturbations was derived by Wheeler in 1955,[13] and the $s = 2$ equation for odd parity gravitational perturbations by Regge and Wheeler a few years later.[14] A very similar equation obeyed by even parity gravitational perturbations was obtained by Zerilli in 1970,[15] and the equivalence of Zerilli's even parity equation to Regge and Wheeler's odd parity equation [Eq. (9) with $s = 2$] was demonstrated by Chandrasekhar[16] and Chandrasekhar and Detweiler[17] in 1975. This 20 years of effort has been summarized by Professor Chandrasekhar in Chap. 4 of his recent book.[21]

Equation (9) may be put in the form of Eq. (1) by means of the substitution

$$\psi_l(r,\omega) = r^{1+s}(r - 1)^{-i\omega}y(r,\omega). \tag{10}$$

The differential equation for y is

$$r(r - 1)y_{,rr} + [2(s + 1 - i\omega)r - (2s + 1)]y_{,r} + [\omega^2r(r - 1) + 2\omega^2(r - 1) + 2\omega^2 - l(l + 1) + s(s + 1) - (2s + 1)i\omega]y = 0, \tag{11}$$

and the indicial structure at the regular singular points $r=0$ and $r=1$ is given by

$$y \overset{r\to 0}{\to} r^{k_1}, \quad y \overset{r\to 1}{\to} (r-1)^{k_2},$$

where $k_1 = 0, -2s$ and $k_2 = 0, 2i\omega$. With the signs for ω chosen in Eqs. (8) and (10), the exterior (i.e., $1<r<\infty$) solution y that is regular at $r=1$ corresponds, for $\mathrm{Re}(\omega)>0$, to a field function that radiates into the horizon. This is the physically meaningful case, but a second exterior solution may be found simply by replacing ω by $-\omega$ in Eqs. (10) and (11).

2. Kerr geometry

The geometry of the rotating black hole has oblate spheroidal nature, and the wave equation for the components of the massless fields can be separated in the oblate spheroidal spatial coordinates λ, μ, and ϕ, and a timelike coordinate t. The coordinates λ and μ may be defined as in Sec. II A for prolate spheroids, but the axis of oblate rotation is the semiminor axis of the family of ellipses parametrized by constant values of λ. The oblate spheroidal coordinate ϕ measures the azimuthal angle about the semiminor axis. The singularities of the coordinate system, which are the fixed locations of the two foci for prolate spheroids, become a singular ring of radius a when the foci rotate about the semiminor axis.

Kerr's spatial coordinates r and θ are simply related to the oblate spheroidal coordinates λ and μ by[22]

$$r = a(\lambda^2 - 1)^{1/2} \quad \text{and} \quad \theta = \sin^{-1}\mu.$$

Simplification of the Kerr metric is obtained by the introduction of the Boyer–Lindquist azimuthal coordinate $\bar{\phi}$, which is related to the azimuthal angle ϕ and the radial coordinate r by

$$d\bar{\phi} = d\phi + a(r^2 - 2Mr + a^2)^{-1}dr.$$

We will follow the usual convention of dropping the "$-$" from $\bar{\phi}$ and denote the Boyer–Lindquist coordinates simply by t, r, θ, and ϕ. These coordinates reduce to Schwarzschild's coordinates as $a\to 0$. The Kerr metric in Boyer–Lindquist coordinates is

$$\begin{aligned} ds^2 = {} & (1 - 2Mr/\Sigma)dt^2 + (4Mar\sin^2(\theta)/\Sigma)dt\,d\phi \\ & - (\Sigma/\Delta)dr^2 - \Sigma\, d\theta^2 \\ & - \sin^2(\theta)(r^2 + a^2 - 2Ma^2 r\sin^2(\theta)/\Sigma)d\phi^2, \end{aligned} \tag{12}$$

where

$$\Sigma = r^2 + a^2\cos^2\theta$$

and

$$\Delta = r^2 - 2Mr + a^2.$$

It is convenient to define one last angular coordinate $u = \cos(\theta) = \pm(1-\mu^2)^{1/2}$. The field function $\Phi(t,r,u,\phi)$ can then be expanded as

$$\begin{aligned} \Phi(t,r,u,\phi) = {} & \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t} \sum_{l=|s|}^{\infty} \sum_{m=-l}^{l} e^{im\phi} S_{lm}(u) \\ & \times R_{lm}(r)d\omega \end{aligned} \tag{13}$$

and, after a rescaling of t and r so that $2M=1$, the following differential equations are obtained[1,18] for the angular function $S(u)$ and the radial function $R(r)$:

$$\begin{aligned} & [(1-u^2)S_{lm,u}]_{,u} + [a^2\omega^2u^2 - 2a\omega su + s + A_{lm} \\ & \quad - (m+su)^2/(1-u^2)]S_{lm} = 0, \end{aligned} \tag{14}$$

and

$$\Delta R_{lm,rr} + (s+1)(2r-1)R_{lm,r} + V(r)R_{lm} = 0, \tag{15}$$

where

$$\begin{aligned} V(r) = {} & [(r^2+a^2)^2\omega^2 - 2am\omega r + a^2m^2 \\ & + is(am(2r-1) - \omega(r^2-a^2))]\Delta^{-1} \\ & + [2is\omega r - a^2\omega^2 - A_{lm}]. \end{aligned}$$

Equations (14) and (15) are the Kerr geometry linearized wave equation analogs to the Schrödinger prolate spheroidal equations (2) and (3). The functions R_{lm} and S_{lm} are referred to as "Teukolsky's functions,"[21] and we will now show that they are, in fact, generalized spheroidal wave functions.

Define an auxiliary rotation parameter b by $b = (1-4a^2)^{1/2}$, and define r_+ and r_- to be the zeros of Δ, so that $\Delta = (r-r_-)(r-r_+)$. Then $r_\pm = (1\pm b)/2$, and $r = r_+$ corresponds to the event horizon. The solutions of Eqs. (15) and (14) at the regular singularities $u=\pm 1$ and $r = r_\pm$ can be found in the usual way: if

$$\lim_{u\to -1} S_{lm} \sim (1+u)^{k_1},$$

and

$$\lim_{u\to +1} S_{lm} \sim (1-u)^{k_2}, \tag{16}$$

then

$$k_1 = \pm\tfrac{1}{2}(m-s), \quad \text{and} \quad k_2 = \pm\tfrac{1}{2}(m+s).$$

Similarly if

$$\lim_{r\to r_-} R_{lm} \sim (r-r_-)^{k_-},$$

and

$$\lim_{r\to r_+} R_{lm} \sim (r-r_+)^{k_+},$$

then

$$k_- = -(i/b)(\omega r_- - am), \quad -s + (i/b)(\omega r_- - am),$$

and

$$k_+ = (i/b)(\omega r_+ - am), \quad -s - (i/b)(\omega r_+ - am).$$

The physically meaningful solutions to the angular equation (14) are regular at the axis ($u=\pm 1$), so the usual choices for k_1 and k_2 are $k_1 = |m-s|/2$ and $k_2 = |m+s|/2$. Similarly, the usual exterior solutions to the radial equation are those that correspond to fields radiating into the event horizon at $r = r_+$. This corresponds to $k_+ = -s - i(\omega r_+ - am)/b$. Boyer–Lindquist coordinates are not well suited to analysis of the physics of the interior problem, but the choice $k_- = -s + i(\omega r_- - am)/b$ turns out to be convenient for the present study restricted to just the differential equation. Letting

$$R_{lm} = (r - r_-)^{-s + (i/b)(\omega r_- - am)} \times (r - r_+)^{-s - (i/b)(\omega r_+ - am)} y(r - r_-) \quad (17)$$

and

$$S_{lm} = (1 + u)^{(1/2)|m - s|} (1 - u)^{(1/2)|m + s|} g(u), \quad (18)$$

then the differential equations for y and g are

$$x(x - b) y_{,xx} + [2(1 - s - i\omega)x + (s - 1 + 2i\omega)b - 2i(\omega r_+ - am)] y_{,x} + [\omega^2 x(x - b) + 2(\omega + is)\omega(x - b) + (1 + b - a^2)\omega^2 + (2s - 1)i\omega + is\omega b - 2s - A_{lm}] y = 0, \quad (19)$$

where $x = r - r_-$, and

$$z(z - 2) g_{,zz} + [2(k_1 + k_2 + 1)z - 2(2k_1 + 1)] g_{,z} + [-a^2\omega^2 z(z - 2) + 2a\omega s z - s - A_{lm} - (s + a\omega)^2 + (k_1 + k_2)(k_1 + k_2 + 1)] g = 0, \quad (20)$$

where $z = u + 1$. These are generalized spheroidal wave equations of form (1).

C. Exponents of the solutions near the regular singular points $x = 0$ and $x = x_0$, and asymptotic solutions

We may now take Eq. (1) to be our standard form for the generalized spheroidal wave equation:

$$x(x - x_0) \frac{d^2 y}{dx^2} + (B_1 + B_2 x) \frac{dy}{dx} + [\omega^2 x(x - x_0) - 2\eta\omega(x - x_0) + B_3] y = 0.$$

We recapitulate the solution forms at the regular singular points. If

$$\lim_{x \to 0} y \sim x^{k_1}, \quad \text{and} \quad \lim_{x \to x_0} y \sim (x - x_0)^{k_2}, \quad (21)$$

then

$$k_1 = 0, 1 + B_1/x_0, \quad \text{and} \quad k_2 = 0, 1 - B_2 - B_1/x_0.$$

Asymptotic solutions are found through the substitution

$$y(x) = x^{B_1/2x_0} (x - x_0)^{-(1/2)(B_2 + B_1/x_0)} v(x).$$

The differential equation for $v(x)$ as $x \to \infty$ is approximately

$$\frac{d^2 v}{dx^2} + \left[\omega^2 - \frac{2\eta\omega}{x} - \frac{\frac{1}{2}B_2(\frac{1}{2}B_2 - 1) - B_3}{x^2} + O(x^{-3})\right] v = 0,$$

so that two independent asymptotic solutions for Eq. (1) are

$$\lim_{x \to \infty} y_+(x) = x^{B_1/2x_0} (x - x_0)^{-(1/2)(B_2 + B_1/x_0)} \times [G_{\nu_a}(\eta, \omega x) + iF_{\nu_a}(\eta, \omega x)] \times [1 + O(x^{-3})], \quad (22)$$

and

$$\lim_{x \to \infty} y_-(x) = x^{B_1/(2x_0)} (x - x_0)^{-(1/2)(B_2 + B_1/x_0)} \times [G_{\nu_a}(\eta, \omega x) - iF_{\nu_a}(\eta, \omega x)] \times [1 + O(x^{-3})], \quad (23)$$

where $F_{\nu_a}(\eta, \omega x)$ and $G_{\nu_a}(\eta, \omega x)$ are the Coulomb wave functions of (usually complex) order

$$\nu_a = \tfrac{1}{2}[-1 \pm (1 + B_2(B_2 - 2) - 4B_3)^{1/2}].$$

To lower order, the asymptotic approximations simplify to

$$\lim_{x \to \infty} y_\pm(x) \sim x^{-((1/2)B_2 \pm i\eta)} e^{\pm i\omega x} [1 + O(1/x)]. \quad (24)$$

Coulomb wave functions will be discussed further in Sec. VI.

III. THREE-TERM RECURRENCE RELATIONS AND THE ANGULAR EIGENVALUE PROBLEM

Every representation of generalized spheroidal wave functions discussed in this paper will involve either a power series expansion, or an expansion in a series of special functions. Since the expansion coefficients in each case will be defined by a three-term recurrence relation, a review of some properties of such relations is in order. The discussion here will be quite brief, and is excerpted primarily from the first few sections of the excellent article on three-term recurrence relations by Gautschi.[6] The theory is illustrated by a simple and relevant example of a sequence determined by a three-term recurrence relation: the coefficients for a power series solution to Eq. (1) about the regular singular point $x = 0$.

A. Power series solutions on $[0 \leqslant x < x_0]$

Equation (1) was

$$x(x - x_0) \frac{d^2 y}{dx^2} + (B_1 + B_2 x) \frac{dy}{dx} + [\omega^2 x(x - x_0) - 2\eta\omega(x - x_0) + B_3] y = 0.$$

Following Baber and Hassé,[5] a power series solution about $x = 0$ may be obtained by letting

$$y(x) = e^{i\omega x} \sum_{n=0}^{\infty} a_n^\theta x^n. \quad (25)$$

We use the superscript θ in this solution to denote its usual association with the angular equations (3) and (14). The sequence of expansion coefficients $\{a_n^\theta: n = 1,2,...\}$ is defined by the three-term recurrence relation

$$\begin{aligned} \alpha_0^\theta a_1^\theta + \beta_0^\theta a_0^\theta &= 0, \\ \alpha_n^\theta a_{n+1}^\theta + \beta_n^\theta a_n^\theta + \gamma_n^\theta a_{n-1}^\theta &= 0, \quad n = 1,2,..., \end{aligned} \quad (26)$$

where

$$\begin{aligned} \alpha_n^\theta &= -x_0 n^2 + (B_1 - x_0)n + B_1, \\ \beta_n^\theta &= n^2 + (B_2 - 2i\omega x_0 - 1)n + 2\eta\omega x_0 + i\omega B_1 + B_3, \\ \gamma_n^\theta &= 2i\omega n + i\omega(B_2 - 2) - 2\eta\omega. \end{aligned} \quad (27)$$

Equations (26) and (27) are equivalent to Baber and Hassé's Eq. (10). We will take Eq. (26) to be the standard form for a three-term recurrence relation. In Secs. V and VI we will also discuss double-ended sequences in which the index n runs from $-\infty$ to $+\infty$, as opposed to the single-ended variety considered here.

Three-term recurrence relations, like second-order differential equations, possess two independent solution sequences $\{A_n: n = 1,2,...\}$ and $\{B_n: n = 1,2,...\}$. The two sequences frequently have the property that $\lim_{n\to\infty} A_n/B_n = 0$. The sequence $\{A_n: n = 1,2,...\}$ is then referred to as the "solution sequence minimal as $n\to\infty$," or briefly, as *minimal*. Any nonminimal solution sequence $\{B_n: n = 1,2,...\}$ is referred to as *dominant* (Gautschi, p. 25). Dominant sequences are not unique, as any multiple of the minimal solution may be added to them without destroying their dominant property. We typically denote either type of sequence by the general sequence $\{a_n: n = 1,2,...\}$. Whether the a_n are minimal or dominant will be seen to depend on the ratio a_1/a_0.

The large n behavior of the expansion coefficients $\{a_n^\theta: n = 1,2,...\}$ may be analyzed by writing Eq. (26) as

$$\alpha_n^\theta \frac{a_{n+1}^\theta}{a_n^\theta} + \beta_n^\theta + \gamma_n^\theta \frac{a_{n-1}^\theta}{a_n^\theta} = 0, \tag{28}$$

dividing by n^2, and keeping only the leading order terms in the result,

$$-x_0 \frac{a_{n+1}^\theta}{a_n^\theta} + 1 + \frac{2i\omega}{n}\frac{a_{n-1}^\theta}{a_n^\theta} \approx 0. \tag{29}$$

We then see that the a_n^θ are elements of the minimal solution sequence if

$$\lim_{n\to\infty} \frac{a_n^\theta}{a_{n-1}^\theta} \sim -\frac{2i\omega}{n}, \tag{30}$$

and the a_n^θ are dominant if

$$\lim_{n\to\infty} \frac{a_n^\theta}{a_{n-1}^\theta} = \frac{1}{x_0}. \tag{31}$$

The ratio of successive elements of the minimal solution sequence to the recurrence relation (26) is given by the continued fraction[6]

$$\frac{a_{n+1}^\theta}{a_n^\theta} = \frac{-\gamma_{n+1}^\theta}{\beta_{n+1}^\theta -}\frac{\alpha_{n+1}^\theta\gamma_{n+2}^\theta}{\beta_{n+2}^\theta -}\frac{\alpha_{n+2}^\theta\gamma_{n+3}^\theta}{\beta_{n+3}^\theta -}\cdots, \tag{32}$$

which for $n = 0$ gives

$$\frac{a_1^\theta}{a_0^\theta} = \frac{-\gamma_1^\theta}{\beta_1^\theta -}\frac{\alpha_1^\theta\gamma_2^\theta}{\beta_2^\theta -}\frac{\alpha_2^\theta\gamma_3^\theta}{\beta_3^\theta -}\cdots. \tag{33}$$

However, for single-ended sequences such as arise out of power series expansions, the first of Eqs. (26) requires that

$$a_1^\theta/a_0^\theta = -\beta_0^\theta/\alpha_0^\theta. \tag{34}$$

Equations (33) and (34) cannot both be satisfied for arbitrary values of the recurrence coefficients $\alpha_n^\theta, \beta_n^\theta$, and γ_n^θ, so that the general solution sequence to Eq. (26) is a dominant one and can usually be generated by simple forward recursion from a chosen value of a_0^θ. The resulting power series (25) will converge for all x of magnitude less than the magnitude of x_0, but will diverge when $|x| \geqslant |x_0|$.

A power series solution for Eq. (1) about the singular point $x = x_0$ may be obtained simply by letting $z = x - x_0$. Then Eq. (1) in this new variable becomes

$$z(z + x_0)y_{,zz} + (B_1 + B_2x_0 + B_2z)y_{,z} + [\omega^2 z(z + x_0) - 2\eta\omega z + B_3]y = 0, \tag{35}$$

which is of the same form as Eq. (1), and a power series solution about $z = 0$ can be generated in the same manner as before. Such a solution could be useful in obtaining the behavior near $x = x_0$ of solutions on the exterior interval $[x_0 \leqslant x < \infty)$. However, the radius of convergence of this series expansion if just $|x_0|$. It is no more useful in obtaining eigensolutions on $[0 \leqslant x \leqslant x_0]$ as is the series (25), and is vastly inferior to Jaffé's solution on $[x_0 \leqslant x < \infty)$. Second power-series solutions, in the cases when $1 + B_1/x_0$ or $1 - B_2 - B_1/x_0$ are integers, may be found by the method of Frobenius.

B. The angular eigenvalue problem

The prolate angular coordinate $\mu = (r_1 - r_2)/2a$ of Eq. (3) and the oblate angular coordinate $u = \pm(1 + \mu^2)^{1/2}$ of Eq. (14) play the same role in their respective wave functions, and the physically meaningful solutions to either of Eqs. (7) or (20) are those that are finite both at $x = 0$ and at $x = x_0$ (i.e., μ or u equal ± 1). These solutions are simple Sturmian eigensolutions, and are obtained for a given value of ω if the angular separation constant A_{lm}, which appears as part of the equation parameter B_3 in the β_n^θ, can be adjusted so that Eqs. (33) and (34) are both satisfied. If so, the resulting solution sequence $\{a_n^\theta: n = 1,2,...\}$ will be purely minimal and the power series (25) will converge at $x = x_0$. Equating the right-hand sides of Eqs. (33) and (34) yields an implicit continued fraction equation for the angular separation constant A_{lm}:

$$0 = \beta_0^\theta - \frac{\alpha_0^\theta\gamma_1^\theta}{\beta_1^\theta -}\frac{\alpha_1^\theta\gamma_2^\theta}{\beta_2^\theta -}\frac{\alpha_2^\theta\gamma_3^\theta}{\beta_3^\theta -}\cdots. \tag{36}$$

The α, β, and γ are defined as explicit functions of B_3 and the other parameters of the differential equation in Eqs. (27), and Eq. (36) may be solved for A_{lm} (that is, B_3) by standard nonlinear root-search techniques. The expansion coefficients a_n^θ are then generated by downward recursion on (26), starting from ratios given by (32) at a suitably large value of n.

Fackerell and Crossman[23] have obtained a continued fraction equation for the eigenvalues of the spin-weighted angular spheroidal equation (14) by expanding $S_{lm}(u)$ in a series of Jacobi polynomials, and discuss the normalization properties of these functions (see also Breuer *et al.*[24]). There is probably an integral relating Fackerell and Crossman's Jacobi polynomial solution with the power series solution reviewed here. Hunter and Guerrieri[25] have done a detailed Wentzel–Kramer–Brillouin–Jeffreys (WKBJ) analysis of the angular equation for large values of A_{lm}, which has provided analytic insight into branch points associated with these eigenvalues. Their work might complement Ferrari and Mashoon's[26] WKBJ analysis of the Schwarzschild quasinormal frequencies to provide useful insight into the large l behavior of the Kerr quasinormal frequencies. It is interesting that none of these recent studies of the angular equation reference the early results of Wilson,[27] or of Baber and Hassé.[5] Fackerell and Crossman's expansions (19) and (20), for instance, apparently are independently derived generalizations of Baber and Hassé's expansions (30) and

(33). The power series expansion we have given here [cf. Eq. (25)] is equivalent to Eq. (34) of Baber and Hassé.

IV. THE SOLUTIONS OF HYLLERAAS AND JAFFÉ, INTEGRAL RELATIONS, AND SECOND SOLUTIONS

Although Hylleraas is generally given credit for the first solution to the bound state problem of the hydrogen molecule ion in 1931,[7] the solution to Eq. (2) derived by Jaffé in 1934[8] was the first to contain a proof of convergence. Such proof did not exist for Hylleraas's representation until Baber and Hassé provided one in 1935.[5] (Baber and Hassé apparently also made independent discovery of Jaffé's solution.) This section will discuss the eigensolutions of Hylleraas and Jaffé, and their convergence properties. In particular, Jaffé's representation will be shown to be simply convergent for noneigenfunction solutions to Eq. (1), in addition to being uniformly convergent for eigenfunctions. An integral equation for Sturmian eigenfunctions is derived and used to illuminate the relationship between the representations of Hylleraas and Jaffé, and to express the solution to Eq. (1) that is regular as $x\to\infty$ in terms of the solution that is regular at $x=x_0$.

A. The solutions of Hylleraas and Jaffé on $[x_0<x<\infty)$

Equation (1) was

$$x(x-x_0)\frac{d^2y}{dx^2}+(B_1+B_2x)\frac{dy}{dx}+[\omega^2x(x-x_0)-2\eta\omega(x-x_0)+B_3]y=0.$$

Hylleraas, using hydrogen atom eigenfunctions as *Ansätze*, expanded the solution $y(x)$ that is regular at $x=x_0$ in a series of Laguerre polynomials:

$$y=e^{i\omega x}\sum_{n=0}^{\infty}\frac{n!a_n^r}{\Gamma(\frac{1}{2}B_2+i\eta+B_1/x_0+1+n)}\times L_n^{B_2+B_1/x_0-1}(-2i\omega(x-x_0)). \quad (37)$$

[The superscript (r) on the expansion coefficients a_n^r indicates that they are related to solutions of "radial" equations, such as (2), (9), and (15).]

Jaffé took a more rigorous approach and reasoned that since a power series expansion of solutions to a differential equation about one regular singular point generally has a radius of convergence equal to the distance from the point of expansion to the next nearest singular point, and that since the singular point at $x=0$ obstructs the convergence of a power series between x_0 and ∞, the obvious solution to the power series convergence problem was to rearrange the singular points so that the point $x=x_0$ was moved to 0, the point at ∞ was moved to 1, and the bothersome singular point at 0 was shuffled off to oblivion. Jaffé effected this rearrangement with the variable change $u=(x-x_0)/x$ and then let

$$y(x)=e^{i\omega x}x^{-(1/2)B_2-i\eta}f(u).$$

The differential equation for f in terms of the variable u is

$$u(1-u)^2f_{,uu}+(c_1+c_2u+c_3u^2)f_{,u}+(c_4+c_5u)f=0, \quad (38)$$

where

$$c_1=B_2+B_1/x_0,\quad c_2=-2[c_1+1+i(\eta-\omega x_0)],$$
$$c_3=c_1+2(1+i\eta),$$
$$c_5=(\tfrac{1}{2}B_2+i\eta)(\tfrac{1}{2}B_2+i\eta+1+B_1/x_0),$$
$$c_4=-c_5-\tfrac{1}{2}B_2(\tfrac{1}{2}B_2-1)+\eta(i-\eta)+i\omega x_0c_1+B_3.$$

The function $f(u)$ can then be expanded in a power series in $u, f(u)=\Sigma_{n=0}^{\infty}a_nu^n$, and Jaffé's solution to the generalized spheroidal wave equation is

$$y_1(x)=e^{+i\omega x}x^{-(1/2)B_2-i\eta}\sum_{n=0}^{\infty}a_n^r\left(\frac{x-x_0}{x}\right)^n. \quad (39)$$

With the Laguerre polynomials defined in Appendix A, the coefficients a_n^r in the Hylleraas expansion (37) and the coefficients a_n^r in the Jaffé expansion (39) have the amusing property of being identical. They obey the same three-term recurrence relation

$$\begin{aligned}&\alpha_0^ra_1^r+\beta_0^ra_0^r=0,\\ &\alpha_n^ra_{n+1}^r+\beta_n^ra_n^r+\gamma_n^ra_{n-1}^r=0,\quad n=1,2,\ldots,\end{aligned} \quad (40)$$

where

$$\alpha_n^r=(n+1)(n+B_2+B_1/x_0),$$
$$\beta_n^r=-2n^2-2[B_2+i(\eta-\omega x_0)+B_1/x_0]n-(\tfrac{1}{2}B_2+i\eta)(B_2+B_1/x_0)+i\omega(B_1+B_2x_0)+B_3, \quad (41)$$
$$\gamma_n^r=(n-1+\tfrac{1}{2}B_2+i\eta)(n+\tfrac{1}{2}B_2+i\eta+B_1/x_0).$$

The normalization of the Laguerre polynomials is important. The convention here is that used by Slater,[28] and by Gradshteyn and Ryzhik.[29] Relevant recurrence and differential properties, as well as alternate normalizations, will be found in Appendix A.

Convergence of the Hylleraas and Jaffé expansions may be analyzed by determining the behavior of the expansion coefficients at large n and applying the ratio test to successive terms in the series. To this end divide recurrence relation (40) by $n^2a_n^r$, retain terms to $O(1/n)$, and expand

$$\lim_{n\to\infty}\frac{a_{n+1}^r}{a_n^r}=1+\frac{a}{\sqrt{n}}+\frac{b}{n}+\cdots. \quad (42)$$

The resulting approximate recurrence relation can be written

$$\left(1+\frac{u}{n}\right)\left(1+\frac{a}{\sqrt{n}}+\frac{b}{n}\right)-\left(2+\frac{v}{n}\right)+\left(1+\frac{w}{n}\right)\times\left(1-\frac{a}{\sqrt{n}}+\frac{a^2-b}{n}+\frac{2ab-a/2-a^3}{n^{3/2}}\right)\approx 0, \quad (43)$$

where u, v, and w are constants given by

$$u=B_2+B_1/x_0+1,$$
$$v=2[B_2+B_1/x_0+i(\eta-\omega x_0)],$$

and

$$w=B_2+B_1/x_0+2i\eta-1.$$

Retaining terms to $O(n^{-3/2})$ and solving for a and b we find $a^2=v-u-w$ and $b=v/2-u$, or

$$a^2=-2i\omega x_0,\quad b=i(\eta-\omega x_0)-\tfrac{3}{4}. \quad (44)$$

The large n behavior of the a_n^r may then be deduced by writing (42) as

$$\lim_{n\to\infty} \frac{a^r_{n+1} - a^r_n}{a^r_n} = \frac{a}{\sqrt{n}} + \frac{b}{n}, \quad (45)$$

and integrating with respect to n. The result[5] is

$$\lim_{n\to\infty} a^r_n \approx n^b e^{2a\sqrt{n}} = n^{i(\eta - \omega x_0) - 3/4} \exp(\pm 2\sqrt{-2i\omega x_0 n}). \quad (46)$$

The two signs ($\pm$) in the exponent indicate the asymptotic behavior of the two independent solution sequences to the recurrence relation. It is apparent that one solution sequence will be dominant and the other minimal for all ωx_0 that are not pure negative imaginary.

The Laguerre polynomials $L_n(z)$ are a dominant solution to the recurrence relation

$$(n+1)L^\alpha_{n+1}(z) - (2n+\alpha+1-z)L^\alpha_n(z) + (n+\alpha)L^\alpha_{n-1}(z) = 0. \quad (47)$$

Repeating the procedure that found $\lim a^r_{n+1}/a^r_n$, we find

$$\lim_{n\to\infty} \frac{L^\alpha_{n+1}(z)}{L^\alpha_n(z)} = 1 + \sqrt{-\frac{z}{n}} + \frac{z+1-\alpha}{2n},$$

where $z = -2i\omega(x - x_0)$. The limiting form of the ratio of successive terms of the series (37) is

$$\lim_{n\to\infty} \frac{\Gamma(\frac{1}{2}B_2 + B_1/x_0 + i\eta + n + 1)}{\Gamma(\frac{1}{2}B_2 + B_1/x_0 + i\eta + n + 2)} \times \frac{(n+1)!a^r_{n+1}L^\alpha_{n+1}(z)}{n!a^r_n L^\alpha_n(z)}$$

$$= 1 + \frac{\sqrt{2i\omega(x - x_0)} \pm \sqrt{-2i\omega x_0}}{\sqrt{n}} + O\left(\frac{1}{n}\right). \quad (48)$$

The ($\pm$) arises from the ratio a^r_{n+1}/a^r_n, and is ($-$) only for sequences a^r_n that are minimal. Hence the only condition under which Hylleraas's expansion (37) can converge is if *both* (i) $2i\omega(x - x_0)$ is purely negative real, *and* (ii) the sequence $\{a^r_n : n = 0,1,2,...\}$ is minimal. [We will not consider cases in which $2i\omega(x - x_0)$ and $-2i\omega x_0$ are both purely negative real. Analysis of that condition hinges on the $O(1/n)$ terms, and in light of the much stronger convergence properties of Jaffé's expansion, is not terribly relevant.] In the context of the quantum mechanics of hydrogen molecule ion condition (i) is automatically satisfied for any negative real energy $E = -\rho^2/2a$ (where $\rho = -i\omega$ in the usual notation), and the fulfillment of condition (ii) becomes the quantization condition on ω. [The continued fraction equation (53) must be satisfied for the recurrence coefficients given in (41).] Hence the Hylleraas expansion successfully represents the eigenfunctions of the bound states of hydrogen moleculelike ions, but very little else.

Jaffé's expansion, on the other hand, is absolutely convergent on $[x_0 < x < \infty)$, and is uniformly convergent there if Σa^r_n is finite (usually only if the a^r_n are minimal). Proof of absolute convergence is trivial: Choose an x from the half plane in which $|(x - x_0)/x| < 1$. Then

$$\lim_{n\to\infty} \left| \frac{a^r_{n+1}[(x - x_0)/x]^{n+1}}{a^r_n[(x - x_0)/x]^n} \right| = \left| \frac{x - x_0}{x} \right| < 1,$$

and convergence at any finite x is assured.

The condition for uniform convergence is similarly demonstrated:

$$\lim_{n\to\infty} \left[\lim_{x\to\infty} \left| \frac{a^r_{n+1}[(x - x_0)/x]^{n+1}}{a^r_n[(x - x_0)/x]^n} \right| \right] = \lim_{n\to\infty} \frac{a^r_{n+1}}{a^r_n} = 1 \pm \frac{\sqrt{-2i\omega x_0}}{\sqrt{n}} - \frac{1 - i(\eta - \omega x_0)}{n}.$$

Convergence is guaranteed if (i) the ($-$) sign is obtained, which is the case if the sequence $\{a^r_n : n = 1,2,...\}$ is minimal, or (ii) if $\mathrm{Re}(\sqrt{2i\omega x_0}) = 0$ and $\mathrm{Im}(\eta - \omega x_0) > 0$. The first case again defines the quantization condition for the hydrogen moleculelike ions, and has also been used to characterize the quasinormal modes of black holes—a problem for which $2i\omega x$ is complex and the Hylleraas expansion is useless. The second case can arise in the consideration of hydrogen moleculelike ion wavefunctions for negative noneigenenergies if we define ρ by $E = -\rho^2/2a$ (i.e., $\rho = -i\omega$) as before, and expand the solution $y(x)$ as

$$y_2(x) = e^{-i\omega x} x^{-(1/2)B_2 + i\eta} \sum_{n=0}^{\infty} b^r_n \left(\frac{x - x_0}{x}\right)^n. \quad (49)$$

The expansion coefficients b^r_n are generated by a three-term recurrence relation

$$\begin{aligned} \tilde{\alpha}_0 b^r_1 + \tilde{\beta}_0 b^r_0 &= 0, \\ \tilde{\alpha}_n b^r_{n+1} + \tilde{\beta}_n b^r_n + \tilde{\gamma}_n b^r_{n-1} &= 0, \quad n = 1,2,..., \end{aligned} \quad (50)$$

where now the recurrence coefficients $\tilde{\alpha}_n$, $\tilde{\beta}_n$, and $\tilde{\gamma}_n$ are given by

$$\begin{aligned} \tilde{\alpha}_n &= (n+1)(n + B_2 + B_1/x_0), \\ \tilde{\beta}_n &= -2n^2 - 2[B_2 - i(\eta - \omega x_0) + B_1/x_0]n \\ &\quad - (\tfrac{1}{2}B_2 - i\eta)(B_2 + B_1/x_0) - i\omega(B_1 + B_2 x_0) + B_3, \\ \tilde{\gamma}_n &= (n - 1 + \tfrac{1}{2}B_2 - i\eta)(n + \tfrac{1}{2}B_2 - i\eta + B_1/x_0). \end{aligned} \quad (51)$$

The $\tilde{\alpha}_n$, $\tilde{\beta}_n$, and $\tilde{\gamma}_n$ of Eq. (51) are the complex conjugates of the α_n, β_n, and γ_n of Eq. (41) only if the parameters B_1, B_2, B_3, ω, x_0, and η are purely real. When $\omega = i\rho$ lies on the positive imaginary axis the independent solutions sequences to recurrence relation (50) are neither minimal nor dominant, so this expression is not well suited to determine the exact hydrogen moleculelike ion eigenfunctions—but it does generate the general negative energy solutions in a stable manner, and was useful in the numerical verification of the integral relationships to be discussed forthwith (Sec. IV C).

B. The radial eigenvalue problem

The eigensolutions of the generalized spheroidal wave equation (1) on the interval $[0 < x < \infty)$ are those functions $y_1(x)$ or $y_2(x)$ of Eqs. (39) and (49) for which Σa^r_n or Σb^r_n converge. The function $y_1(x)$ then describes an eigenfunction that is regular at $x = x_0$ and has purely $\exp[+i(\omega x - \eta \ln x)]$ behavior as $x\to\infty$, and $y_2(x)$ describes an eigenfunction that is regular at $x = x_0$ and has purely $\exp[-i(\omega x - \eta \ln x)]$ behavior as $x\to\infty$. The sums over a^r_n or b^r_n will usually converge iff the a^r_n or b^r_n are minimal solutions to their respective recurrence relations

(40) and (50), and this will happen only for certain characteristic values of the frequency ω. (The values of ω for which the a_n^r are minimal will not be the same as the values of ω for which the b_n^r are minimal.) As in our previous discussion of the angular eigenvalue problem, the coefficients a_n^r will be minimal iff they satisfy the continued fraction equation

$$\frac{a_{n+1}^r}{a_n^r} = \frac{-\gamma_{n+1}^r}{\beta_{n+1}^r -} \frac{\alpha_{n+1}^r \gamma_{n+2}^r}{\beta_{n+2}^r -} \frac{\alpha_{n+2}^r \gamma_{n+3}^r}{\beta_{n+3}^r -} \cdots, \quad (52)$$

which in turn will require that ω be a root of

$$0 = \beta_0^r - \frac{\alpha_0^r \gamma_1^r}{\beta_1^r -} \frac{\alpha_1^r \gamma_2^r}{\beta_2^r -} \frac{\alpha_2^r \gamma_3^r}{\beta_3^r -} \cdots. \quad (53)$$

Here the α_n^r, β_n^r, and γ_n^r are defined as functions of ω in Eqs. (41). Analogous equations can be written concerning the b_n^r and the $\tilde{\alpha}_n$, $\tilde{\beta}_n$, and $\tilde{\gamma}_n$ in the instances when eigensolutions of the type y_2 are desired.

In most physical situations both the α_n^θ, β_n^θ, and γ_n^θ for the angular eigenvalue equation (36) and the α_n^r, β_n^r, and γ_n^r for the radial eigenvalue equation (53) are functions of both the angular separation constant A_{lm} and of the frequency ω. This will then require the simultaneous solution of Eqs. (36) and (53), which usually is not difficult numerically. Such solutions were demonstrated for the electronic spectra of the hydrogen molecule ion by Hylleraas,[7] Jaffé,[8] and Baber and Hassé.[5] Analogous solutions for the quasinormal modes of black holes are given by Leaver.[11] With use of eigensolutions of type y_2 a similar approach can be taken to the "algebraically special" black hole perturbations discussed by Chandrasekhar.[30]

C. Second solutions by way of an integral transform

If we express the solutions to Eq. (1) near the singular point $x = x_0$ as

$$\lim_{x \to x_0} y(x) = (x - x_0)^{k_2}, \quad (54)$$

then the exponent k_2 takes the values 0 and $1 - B_2 - B_1/x_0$. If $B_2 + B_1/x_0$ is not an integer, a second solution to Eq. (1) may be found through the substitution $y(x) = (x - x_0)^{1 - B_2 - B_1/x_0} g(x)$. The differential equation for g will be

$$\begin{aligned} &x(x - x_0) g_{,xx} + [B_1 + (2 - B_2 - 2B_1/x_0)x] g_{,x} \\ &\quad + \{\omega^2 x(x - x_0) - [2\eta\omega + (1 - B_2 - B_1/x_0)B_1/x_0] \\ &\quad \times (x - x_0) + B_3\} g = 0, \end{aligned} \quad (55)$$

which is of the same form as Eq. (1), and a regular solution for f may be found by the method of Jaffé. If $B_2 + B_1/x_0$ is an integer, then a second solution to Eq. (38) may be found by the method of Frobenius. The expansion coefficients for the resulting second solution will obey an inhomogeneous three-term recurrence relation, and contain a free parameter that may be empirically adjusted to vary the amount of the first solution that appears in the second. This property is interesting, but the procedure is tedious and will not be dealt with here (see Rabenstein[31] for a discussion of Frobenius's method).

A more entertaining approach to the second solutions is open to those who remain curious about the equality of the Hylleraas and Jaffé expansion coefficients. Wilson[27] speculated that "the solution of (a generalized spheroidal wave equation) is probably expressible as a homogeneous integral equation." One such integral had already been given by Ince[32] for the particular parameters choice $\eta = \pm i(B_1 + B_2)/2$, and although the contour used by Ince was $[-1,1]$, his expressions can be made valid on $[1,\infty)$. Another integral relation for a different, though still specific, choice $\eta = \pm i(B_2/2 - 1)$ is arrived at through consideration of the equality of the Hylleraas and Jaffé expansion coefficients, and leads directly to a representation for a second solution to the differential equation as a series of irregular confluent hypergeometric functions. The new representation is valid for arbitrary η. The argument goes as follows: Start with Eq. (1)

$$\begin{aligned} &x(x - x_0) \frac{d^2y}{dx^2} + (B_1 + B_2 x) \frac{dy}{dx} \\ &\quad + [\omega^2 x(x - x_0) - 2\eta\omega(x - x_0) + B_3] y = 0, \end{aligned}$$

and make the substitution $y = e^{i\omega x} f(x)$. The differential equation for $f(x)$ is

$$\begin{aligned} &x(x - x_0) f_{,xx} + [B_1 + B_2 x + 2i\omega x(x - x_0)] f_{,x} \\ &\quad + [(B_2 + 2i\eta) i\omega x + 2\eta\omega x_0 + i\omega B_1 + B_3] f = 0, \end{aligned} \quad (56)$$

and f admits to the expansion

$$\begin{aligned} f(x) = &\sum_{n=0}^{\infty} \frac{n! a_n^r}{\Gamma(\frac{1}{2}B_2 + i\eta + B_1/x_0 + 1 + n)} \\ &\times L_n^{B_2 + B_1/x_0 - 1}(-2i\omega(x - x_0)) \end{aligned} \quad (57)$$

(Hylleraas), and

$$\bar{f}(x) = x^{-(1/2)B_2 - i\eta} \sum_{n=0}^{\infty} a_n^r \left(\frac{x - x_0}{x}\right)^n \quad (58)$$

(Jaffé). The coefficients a_n^r are the same for each expansion and $\bar{f}(x)$ is proportional to $f(x)$ when both are eigenfunctions such that $\Sigma\, a_n^r$ converges. Specializing to the case $i\eta = B_2/2 - 1$, these expressions, respectively, become

$$\begin{aligned} f(x) = &\sum_{n=0}^{\infty} \frac{n! a_n^r}{\Gamma(B_2 + B_1/x_0 + n)} \\ &\times L_n^{B_2 + B_1/x_0 - 1}(-2i\omega(x - x_0)), \end{aligned} \quad (59)$$

and

$$\bar{f}(x) = x^{1 - B_2} \sum_{n=0}^{\infty} a_n^r \left(\frac{x - x_0}{x}\right)^n. \quad (60)$$

Perusal of standard integral tables[33] reveals

$$\int_0^\infty e^{-st} t^\alpha L_n^\alpha(t) dt = \frac{\Gamma(\alpha + n + 1)}{n!} s^{-\alpha - 1} \left(\frac{s - 1}{s}\right)^n, \quad (61)$$

so that with the associations $\alpha = B_2 + B_1/x_0 - 1$, $t = -2i\omega(x - x_0)$, and $s = x/x_0$, we conclude

$$\begin{aligned} \bar{f}(x) = x^{1 + B_1/x_0} &\int_c e^{2i\omega x(t - x_0)/x_0} \\ &\times (t - x_0)^{B_2 + B_1/x_0 - 1} f(t) dt, \end{aligned} \quad (62)$$

for some contour c that includes x_0 and ∞. Multiplicative constants have been omitted from the integration. This result is verified via the theory of integral transforms in Appendix B. The important result of that derivation is the procurement of the bilinear concomitant

$$P(x,t) = t(t-x_0)\left[f(t)\frac{d}{dt}K(x,t) - K(x,t)\frac{d}{dt}f(t)\right] + (2i\omega t^2 + (B_2 + 2B_1/x_0 - 2i\omega x_0)t - B_1 - x_0)K(x,t)f(t), \qquad (63)$$

where the kernel $K(x,t)$ is given by

$$K(x,t) = \exp[2i\omega x(t-x_0)/x_0](t-x_0)^{-s_2}$$

and

$$s_2 = 1 - B_2 - B_1/x_0.$$

The exponent s_2 takes the second of the allowed values of k_2 of Eq. (54). The bilinear concomitant must vanish at each end of the integration contour.

On such an integration contour, Eq. (62) is an integral relation among solutions $f(x)$ and $\bar{f}(x)$ to Eq. (56). This does not necessarily mean that f and $\bar{f}$ are the same solution to the differential equation, however. Equation (62) is an integral equation only for functions $f(x)$ that have the decreasing exponential behavior at $x=\infty$. If such an $f(x)$ should also happen to be regular at $x=x_0$, then $f(x)$ is an eigenfunction of Eq. (56) and one end point of the contour c can be taken directly to $t=x_0$. In this case $f(x)$ and $\bar{f}(x)$ are proportional, and Eq. (62) becomes an integral equation for eigenfunctions. It may be noted that the quasinormal modes of black holes can be described by this kind of eigenfunction, although the requirement $i\eta = B_2/2 - 1$ restricts the applicability of (62) to consideration only of scalar fields [$s=0$ in Eqs. (11) and (19)].

The integration contour c is determined by the requirement that the bilinear concomitant $P(x,t)$ vanish at its end points. If $f(x)\to(x-x_0)^{s_i}$ as $x\to x_0$, then the allowed values for the exponent s_i are $s_1 = 0$ and $s_2 = 1 - B_2 - B_1/x_0$. We consider two general cases.

(1) $f(x) \overset{x\to x_0}{\longrightarrow}$ (constant) and either $\text{Re}(s_2) < 0$ or $s_2 = 0$.

In this case $P(x,t)$ vanishes at $t = x_0$ and the contour c may be taken to be that shown in Fig. 1(a). The approach angle θ is chosen such that $\text{Re}(2i\omega xt/x_0) < 0$. The kernel $K(x,t)$ is then an exponentially decreasing function of x, and Eq. (62) expresses the solution $\bar{f}(x)$ regular as $x\to\infty$ in terms of the solution $f(t)$ that is regular as $t\to x_0$. If $f(t)$ is also regular as $t\to\infty$, then ω is an eigenfrequency, $\bar{f}$ is proportional to f, and Eq. (62) becomes an integral equation for the eigenfunctions. We can see how this works by substituting the Jaffé expansion for $f(t)$ into Eq. (62):

$$\bar{f}(x) = x^{1+B_1/x_0}\int_{x_0}^{\infty}\exp[2i\omega x(t-x_0)/x_0] \times (t-x_0)^{-s_2}t^{1-B_2}\left[\sum_{n=0}^{\infty} a_n^r\left(\frac{t-x_0}{t}\right)^n\right]dt. \qquad (64)$$

The behavior of $\bar{f}(x)$ near $x=x_0$ is determined by the large t behavior of the integrand. If ω is an eigenfrequency, the series $\Sigma\, a_n^r\,[(t-x_0)/t\,]^n$ is uniformly convergent as $t\to\infty$ and the integral for large t, $x\to x_0$ looks like

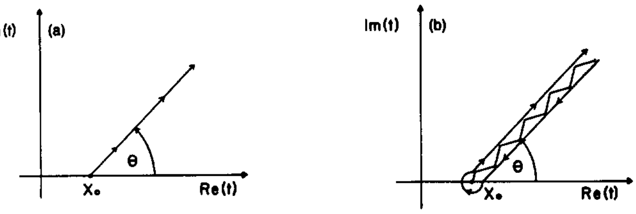

FIG. 1. Contours for use with integral relation (62).

$$\bar{f}(x) \overset{x\to x_0}{\longrightarrow} x^{1+B_1/x_0}\int^{\infty}\exp[2i\omega x(t-x_0)/x_0]t^{B_1/x_0}\,dt$$

(since $s_2 = 1 - B_2 - B_1/x_0$) and is always finite with the aforementioned choice of approach angle θ. Hence $\bar{f}(x)$ is finite as $x\to x_0$. If ω is not an eigenfrequency, then

$$\Sigma\, a_n^r\,[(t-x_0)/t\,]^n \overset{t\to\infty}{\longrightarrow} t^{B_2-2}e^{-2i\omega t},$$

the integral looks like

$$\bar{f}(x) \overset{x\to x_0}{\longrightarrow} x^{1+B_1/x_0}\int^{\infty}\exp[2i\omega(x-x_0)t/x_0]t^{-s_2-1}\,dt,$$

and

$$\lim_{x\to x_0}\bar{f}(x) \sim (x-x_0)^{s_2}$$

as required for the independent second solution. Note that while in physical contexts the variable x is a spatial coordinate and is positive and real, the flexibility afforded in the choice of the approach angle allows Eq. (62) to describe functions f for which $\text{Im}(\omega)\leqslant 0$.

(2) $f(x) \overset{x\to x_0}{\longrightarrow} (x-x_0)^{s_2}$ or $\text{Re}(s_2) > 0$. Note that the restriction $\text{Re}(s_2)>0$ is artificial, since one can always obtain $\text{Re}(s_2)<0$ for a function $g(x)$ by substituting $f(x) = (x-x_0)^{s_2}g(x)$ in Eq. (56). Either way $P(x,t)$ is not zero at $t=x_0$ and the contour c is chosen to be that illustrated in Fig. 1(b). This contour has the appearance of being all-purpose and do-everything, but we shall see that if one actually had the information necessary to use it, one would also have the information to convert the problem to that considered in case (1) above, and would end up using the contour of Fig. 1(a).

The branch cut arises from the factor $(t-x_0)^{-s_2}$ in $K(x,t)$ if s_2 is not an integer, and from the logarithmic term inherent in f if s_2 is an integer. The function $\bar{f}(x)$ is regular as $x\to\infty$ regardless of the behavior of $f(t)$, so that given any solution $f(x)$ to Eq. (56), Eq. (62) will always give the solution $\bar{f}(x)$ that is regular as $x\to\infty$ for the chosen contour c. On this contour, Eq. (62) is an integral equation for all functions f that are regular at $t=\infty$, but is of limited computational utility as an integral equation for noneigenfunctions because a solution that is irregular at $x=x_0$ and regular at $x=\infty$ must be a weighted sum of two component functions, one regular and the other irregular at $x=x_0$, and both irregular at $x=\infty$. Detailed knowledge of the weighting factors in the sum is necessary, since the product of only one of the component functions and the kernel $K(x,t)$ will contribute to the integral. The product of the other component function and the kernel will have the same value on each side of the branch cut, and will give no contribution. To see how this works, first consider s_2 not an integer. Then $f(t)$ can be written

$$f(t) = t^{1-B_2}\left[\sum_{n=0}^{\infty} a_n^r \left(\frac{t-x_0}{t}\right)^n + (t-x_0)^{s_2} \sum_{n=0}^{\infty} b_n^r \left(\frac{t-x_0}{t}\right)^n\right], \quad (65)$$

which is just the sum of two independent Jaffé solutions. Only the product of $K(x,t)$ and the function corresponding to the first sum will contribute to the integral, and if that function were known (i.e., if we knew the value of a_0^r), we could use case (1) above. Similarly, if s_2 is an integer, then any solution irregular at x_0 is expressible as

$$f(t) = t^{1-B_2}\left[\log\left(\frac{t-x_0}{t}\right) \sum_{n=0}^{\infty} a_n^r \left(\frac{t-x_0}{t}\right)^n + (t-x_0)^{s_2} \sum_{n=0}^{\infty} b_n^r \left(\frac{t-x_0}{t}\right)^n\right], \quad (66)$$

which is the form of the second Jaffé solution as obtained by the method of Frobenius. Here again only the product of $K(x,t)$ and the term containing the first sum will contribute to the integral. If the product of the logarithm and the first solution were known, we again would revert to case (1) since the difference of the logarithm across the branch cut is just the constant $2\pi i$, and the integrand becomes effectively integrable at x_0. Either way we are required to know the function that is regular at $x = x_0$ in order to evaluate the difference across the branch cut, and if that solution is known (such as by Jaffe's method), then the problem reduces to the one considered in case (1).

As noted previously, Hylleraas's expansion converges only when ω is a purely imaginary eigenfrequency. We have shown how in that case the relation of the Hylleraas expansion coefficients to the Jaffé coefficients leads to an integral equation for eigenfunctions (at least when $i\eta = B_2/2 - 1$) and how, when ω is not an eigenfrequency, the same integral will transform the first solution that is regular at $x = x_0$ into a second independent solution that is regular at $x = \infty$. Jaffé's method always gives a convergent expansion for the regular first solution, and it is interesting to examine the result of transforming Jaffé's expansion term by term.

We interchange the summation and the integration to explicitly evaluate the right-hand side of Eq. (64):

$$\begin{aligned}\bar{f}(x) &= x^{1+B_1/x_0} \sum_{n=0}^{\infty} a_n^r \left\{\int_{x_0}^{\infty} \exp[2i\omega x(t-x_0)/x_0]\, t^{1-B_2-n}(t-x_0)^{B_2+B_1/x_0+n-1}\, dt\right\} \\ &= x^{1+B_1/x_0} \sum_{n=0}^{\infty} a_n^r\, \Gamma(B_2+B_1/x_0+n)\, U(B_2+B_1/x_0+n, 2+B_1/x_0, -2i\omega x) \\ &= \sum_{n=0}^{\infty} a_n^r\, \Gamma(B_2+B_1/x_0+n)\, U(B_2-1+n, -B_1/x_0, -2i\omega x). \end{aligned} \quad (67)$$

Here $U(a,b,z)$ is the irregular confluent hypergeometric function defined by the integral representation

$$\Gamma(a)U(a,b,z) = \int_0^{\infty} e^{-zt} t^{a-1}(t+1)^{b-a-1}\, dt, \quad (68)$$

and obeys the Kummer relation[28]

$$U(a,b,z) = z^{1-b}U(1+a-b, 2-b, z). \quad (69)$$

The normalization in $\bar{f}(x)$ is not important here, and the constant multiplying factors were dropped during the integration.

The last of Eqs. (67) may also be arrived at by the usual eigenfunction expansion method of solving ordinary differential equations (see Appendix C), which produces a result that holds for arbitrary η:

$$\bar{f}(x) = \sum_{n=0}^{\infty} a_n^r\, \Gamma(B_2+B_1/x_0+n) \times U(B_2/2+i\eta+n, -B_1/x_0, -2i\omega x). \quad (70)$$

The expansion coefficients a_n^r are the same as Jaffé's [Eq. (41)], and since $y(x) = e^{i\omega x}\bar{f}(x)$, we now have a second independent solution to the generalized spheroidal wave equation (1). Expansion (70) is absolutely convergent on any interval bounded away from x_0, is uniformly convergent as $x \to \infty$, diverges at $x = x_0$ when ω is not an eigenfrequency, and is uniformly convergent as $x \to x_0$ when ω is an eigenfrequency.

The derivations for the second solutions may again be repeated with the substitutions $y(x) = e^{-i\omega x}f(x)$. We then have our first four convergent representations for solutions to the generalized spheroidal wave equation:

$$y_1(x) = e^{+i\omega x} x^{-B_2/2-i\eta} \sum_{n=0}^{\infty} a_n^r \left(\frac{x-x_0}{x}\right)^n, \quad (71)$$

$$y_2(x) = e^{-i\omega x} x^{-B_2/2+i\eta} \sum_{n=0}^{\infty} b_n^r \left(\frac{x-x_0}{x}\right)^n, \quad (72)$$

$$y_3(x) = e^{+i\omega x} \sum_{n=0}^{\infty} a_n^r\, (B_2+B_1/x_0)_n \times U(\tfrac{1}{2}B_2+i\eta+n, -B_1/x_0, -2i\omega x), \quad (73)$$

$$y_4(x) = e^{-i\omega x} \sum_{n=0}^{\infty} b_n^r\, (B_2+B_1/x_0)_n \times U(\tfrac{1}{2}B_2-i\eta+n, -B_1/x_0, +2i\omega x). \quad (74)$$

Here y_3 and y_4 have been normalized by a factor $1/\Gamma(B_2+B_1/x_0)$, and $(z)_n \equiv \Gamma(z+n)/\Gamma(z)$ is Pochhammer's symbol. The a_n^r are defined by Eqs. (40) and (41), and the b_n^r by Eqs. (50) and (51). The solutions $y_1(x)$ and $y_2(x)$ are both regular as $x \to x_0$, and are proportional by the factor $e^{2i\omega x_0} a_0^r/b_0^r$. However, convergence properties and growth behavior of individual terms in the series will differ markedly if $\text{Im}(\omega)$ is not zero. Solutions $y_3(x)$ and $y_4(x)$ are independent and are both irregular as $x \to x_0$ when ω is not an eigenfrequency. When ω is an eigenfrequency then one or the other of $y_3(x)$ and $y_4(x)$ will be regular at $x = x_0$ (see Appendix C). Solutions $y_3(x)$ and $y_4(x)$ should have the limiting forms

$$\lim_{|x| \to \infty} y_3(x) = a_0^r\, x^{-(1/2)B_2} \exp[+i(\omega x - \eta \ln x)] \tag{75}$$

and

$$\lim_{|x| \to \infty} y_4(x) = b_0^r\, x^{-(1/2)B_2} \exp[-i(\omega x - \eta \ln x)] \tag{76}$$

(see Slater,[28] Eq. 13.5.2). Although we have made extensive computational use of Eqs. (71) and (72)—they are the basis of our standard algorithm for generating regular solutions to the generalized spheroidal wave equation near $x = x_0$—we have not yet (as of July 1985) been able to verify expansions (73) and (74) and their asymptotic forms (75) and (76) with a computer. But they do look as if they might be useful.

D. Computational limitations of the Jaffé solutions

As might be expected, the absolute convergence property of Jaffé's expansion makes (39) an extremely useful expression for the numerical evaluation of the generalized spheroidal wave function that is regular at $x = x_0$, and for those eigenfunctions for which convergence is uniform it provides the algorithm of choice. However, for arbitrary ω the a_n^r are dominant, and it behooves one to graph the behavior of the sequence $\{a_n^r[(x-x_0)/x]^n: n = 0,1,2,...\}$ as a function of n before concluding one really can sum its terms. Assume that the a_n are dominant. If the sequence is normalized such that $a_0^r = 1$, the sum

$$\sum_{n=0}^{\infty} a_n^r \left(\frac{x - x_0}{x}\right)^n$$

will typically have magnitude of $O(1 + |\omega^{-1}|)$. For a rough estimate, ignore the n^b term in Eq. (46). Then for large x,

$$\lim_{n \to \infty} \left| a_n^r \left(\frac{x - x_0}{x}\right)^n \right| \approx \left|\frac{x - x_0}{x}\right|^n |e^{2\rho\sqrt{n}}| \tag{77}$$

(where $\rho = \sqrt{-2i\omega x_0}$), and $|a_n^r[(x-x_0)/x]^n|$ has a maximum at $n_{\max} \approx (\rho x/x_0)^2$. To give an idea of the numerical problems lurking in wait of the unwary, consider the not unreasonable case of $\rho = 1$, $x/x_0 = 5$. Then $n_{\max} \approx 25$ and

$$a_{n_{\max}}^r \left(\frac{x - x_0}{x}\right)^{n_{\max}} \approx 80,$$

which is only two orders of magnitude greater than the sum of all the terms. But if $x/x_0 = 20$,

$$a_{n_{\max}}^r \left(\frac{x - x_0}{x}\right)^{n_{\max}} \approx 3 \times 10^8,$$

and rounding considerations dictate the use of extended precision if the series (39) is to be summed with any accuracy.

V. THE STRATTON SOLUTIONS TO THE ORDINARY SPHEROIDAL WAVE EQUATION, AND A PRELIMINARY GENERALIZATION

The separated parts of the one-particle Schrödinger equation simplify in free space where there is no potential and $N_1 = N_2 = 0$. Equations (2) and (3) are then the same as those resulting from the separation of the Helmholtz equation in spheroidal coordinates, and Stratton's[34] representations of the ordinary spheroidal wave functions are a natural starting point for investigation of solutions to more general forms of the equation. We were originally attracted to Stratton's spherical Bessel function expansion for two reasons. First, the asymptotic magnitude and phase of any convergent series of spherical Hankel functions $\Sigma a_n h_n^{(1)}(z)$ or $\Sigma a_n h_n^{(2)}(z)$ can be readily calculated. Second, numerical algorithms to generate Bessel functions for a variety of orders and a wide range of magnitudes of the argument are reasonably well understood. The first property will be dealt with in full generality in Sec. VI. The second will be touched upon in the Summary. The present section reviews the Stratton representation for ordinary spheroidal wave functions, and generalizes it to the case of the equations that arise for an electron in the field of a finite dipole $N_1 = -N_2 \neq 0$. A detailed discussion of convergence properties is given, which will serve as a model for the convergence proofs of the general Coulomb wave function expansions presented in Sec. VI, and which (it is hoped) will dispel misconceptions concerning the convergence of Stratton's solutions to the ordinary spheroidal wave equation.

A. The ordinary spheroidal wave equation

The ordinary spheroidal wave equation results from the separation of the Helmholtz and free-particle Schrödinger equations in spheroidal coordinates. It is a special case of Eqs. (2) and (3) for which N_1 and N_2 are both zero, and for which Eqs. (2) and (3) become the same. The angular equation (3) simplifies to

$$\frac{d}{d\mu}\left[(1-\mu^2)\frac{d\Phi}{d\mu}\right] + \left[-\omega^2\mu^2 + A_{lm} - \frac{m^2}{1-\mu^2}\right]\Phi = 0, \tag{78}$$

and the solution function $\Phi(\mu)$ can be expanded in a series of Gegenbauer polynomials[35]:

$$\Phi(\mu) = (1-\mu^2)^{m/2} \sum_{n=0,1}^{\infty}{}' \, d_n T_n^m(\mu). \tag{79}$$

The " ′ " indicates the sum is to be taken over even values of n if l is even and over odd values of n if l is odd. The Gegenbauer polynomials are generated[36] by

$$\sum_{n=0}^{\infty} h^n T_n^m(z) = \frac{2^m \Gamma(m + \frac{1}{2})}{\sqrt{\pi}(1 + h^2 - 2hz)^{m+(1/2)}} \quad (|h| < 1),$$

and are related to the regular Gauss hypergeometric series ${}_2F_1$ by

$$T_n^m(z) = \frac{(n+2m)!}{2^m n! m!}\, {}_2F_1\left(n + 2m + 1, -n; m+1; \frac{1-z}{2}\right).$$

The expansion coefficients $\{d_n: n = 0,2,4,... \text{ or } n = 1,3,5,...\}$ obey the recurrence relation[37]

$$\begin{aligned}
&\alpha_0 d_2 &&+ \beta_0 d_0 &&= 0, \\
&\alpha_1 d_3 &&+ \beta_1 d_1 &&= 0, \\
&\alpha_n d_{n+2} &&+ \beta_n d_n + \gamma_n d_{n-2} &&= 0, \quad n = 2,3,4,...,
\end{aligned}$$

where

$$\alpha_n = \omega^2 \frac{(n+2m+1)(n+2m+2)}{(2n+2m+3)(2n+2m+5)},$$

$$\beta_n = \omega^2 \frac{2[(n+m)(n+m+1)-m^2]-1}{(2n+2m+3)(2n+2m-1)} + (n+m)(n+m+1) - A_{lm}, \quad (80)$$

$$\gamma_n = \omega^2 \frac{n(n-1)}{(2n+2m-1)(2n+2m-3)}.$$

In order for the series to converge at $\mu = \pm 1$, the separation constant A_{lm} must be chosen such that the d_n are minimal and the continued fraction equation

$$\beta_0 = \frac{\alpha_0\gamma_2}{\beta_2 -} \frac{\alpha_2\gamma_4}{\beta_4 -} \frac{\alpha_4\gamma_6}{\beta_6 -} \cdots \quad (81)$$

is satisfied.

The simplified spheroidal radial equation (2) becomes

$$\frac{d}{d\lambda}\left[(\lambda^2-1)\frac{d\Psi}{d\lambda}\right] + \left[\omega^2\lambda^2 - A_{lm} - \frac{m^2}{\lambda^2-1}\right]\Psi = 0, \quad (82)$$

which is the same as (78) but in the coordinate λ instead of μ. Next, if $\Phi(\mu)$ is a solution to (78), then

$$\Psi(\lambda) = (\lambda^2-1)^{m/2} \int_{-1}^{+1} e^{i\omega\lambda\mu}(1-\mu^2)^{m/2}\Phi(\mu)\, d\mu \quad (83)$$

is a solution also,[32] but in the variable λ. Integrating the series (79) for Φ term by term and using the relation

$$\int_{-1}^{+1} e^{i\omega z t}(1-t^2)^m T_n^m(t)\, dt = i^n \frac{2(n+2m)!}{n!(\omega z)^m} \times j_{n+m}(\omega z)$$

(Morse and Feshbach,[10] p. 643—the j_{n+m} are spherical Bessel functions), we obtain the final result

$$\Psi(\lambda) = \left(\frac{\lambda^2-1}{\lambda^2}\right)^{m/2} \sum_{n=0,1}^{\infty}{}' \, i^n d_n \frac{(n+2m)!}{n!} \times j_{n+m}(\omega\lambda). \quad (84)$$

(See Morse and Feshbach, Eq. 11.3.91. We have left off the normalization factors.) The d_n are the same as in the expansion (79) for Φ and satisfy recurrence relation (80). A second solution to Eq. (82) is obtained by substituting the irregular spherical Bessel functions $y_n(\omega\lambda)$ in place of the $j_n(\omega\lambda)$ in expression (84).

The convergence properties of both solutions will be discussed in Sec. V C. The important point for now is that the series (84) converges only if the d_n form a minimal solution to the recurrence relation (80), which can happen only for specific values of the parameter A_{lm}.

B. Preliminary generalization: Schrödinger's equation for an electron in the field of a finite dipole

The simplest generalization of the ordinary spheroidal wave equation is the removal of the freedom to choose A_{lm}. The physical context wherein this complexity arises is the separation of the Schrödinger equation for an electron in the dipole field of two fixed but oppositely charged nuclei. In this consideration $N_2 = -N_1$ in Eqs. (2) and (3), so that Eq. (2) still simplifies to Eq. (82). However, Eq. (3) becomes more complicated, and while it is still readily solvable by the power series method described in Sec. III, the resulting separation constant A_{lm} is now dependent on the dipole moment $2aN_1$ in addition to ω. If we again try to expand the solution to Eq. (82) as

$$\Psi(\lambda) = \left(\frac{\lambda^2-1}{\lambda^2}\right)^{m/2} \sum_{n=0,1}^{\infty}{}' \, i^n d_n \frac{(n+2m)!}{n!} \times j_{n+m}(\omega\lambda) \quad (85)$$

we find that although the d_n still satisfy the recurrence relation (80), they will in general form a dominant solution sequence since A_{lm} can no longer have a value that will force them to be minimal. The series (85) will then diverge.

The convergence problem can be solved if we can find some other parameter in the recurrence relation that may be adjusted so that the expansion coefficients d_n form a minimal solution sequence. There are no free parameters left in the differential equation (82) itself, so a new parameter must be introduced in the representation of the solution. Consideration of the physical problem of the electron in the dipole field leads to the suspicion that the natural choice for such a parameter will have something to do with the asymptotic phase of the solution function $\Psi(\lambda)$, since the asymptotic phase of the Schrödinger wave function will be shifted as the dipole moment $2aN_1$ of the source potential increases away from zero. Specifically, a solution to Eq. (82) for arbitrary values of A_{lm} and ω may be expressed as a generalized Neumann expansion[38]

$$\Psi_1(\lambda) = \left(\frac{\lambda^2-1}{\lambda^2}\right)^{m/2} \sum_{L=-\infty}^{\infty}{}' \, a_L j_{L+\nu}(\omega\lambda). \quad (86)$$

The $j_{L+\nu}$ are again spherical Bessel functions. A second solution may be obtained by substituting the irregular spherical Bessel functions $y_{L+\nu}$ for the $j_{L+\nu}$:

$$\Psi_2(\lambda) = \left(\frac{\lambda^2-1}{\lambda^2}\right)^{m/2} \sum_{L=-\infty}^{\infty}{}' \, a_L y_{L+\nu}(\omega\lambda). \quad (87)$$

The phase (or order) parameter ν in expansions (86) and (87) is free to be adjusted to make the a_L minimal, and thus to obtain convergence of the series. The recurrence relation obeyed by the a_L is

$$\alpha_L a_{L+2} + \beta_L a_L + \gamma_L a_{L-2} = 0, \quad (88)$$

where

$$\alpha_L = -\omega^2 \frac{(L+\nu-m+1)(L+\nu-m+2)}{(2L+2\nu+3)(2L+2\nu+5)},$$

$$\beta_L = +\omega^2\left[\frac{2[(L+\nu)(L+\nu+1)-m^2]-1}{(2L+2\nu-1)(2L+2\nu+3)}\right] + (L+\nu)(L+\nu+1) - A_{lm},$$

$$\gamma_L = -\omega^2 \frac{(L+\nu+m)(L+\nu+m-1)}{(2L+2\nu-1)(2L+2\nu-3)}.$$

If ν should equal m, then this recurrence relation is exactly the same as the recurrence relation (80) obeyed by the d_n if we make the substitution $a_L = i^L[(L+2m)!/L!]d_L$. In this case the series (86) must be started at $L=0$ or $L=1$ instead of $L=-\infty$, and the solution representation (86) reduces to representation (84) for the ordinary spheroidal wave functions. Convergence properties of the two representations are therefore the same.

In the more general case when $N_1=-N_2\neq 0$, the values of the parameter ν for which the series (86) and (87) converge will not be integers, and the sequence $\{a_L: L=..., -2, -1, 0, +1, +2,...\}$ must be made minimal both as $L\to\infty$ and as $L\to-\infty$. The ratios of successive a_L must then satisfy both

$$\frac{a_L}{a_{L-2}}=\frac{-\gamma_L}{\beta_L-}\frac{\alpha_L\gamma_{L+2}}{\beta_{L+2}-}\frac{\alpha_{L+2}\gamma_{L+4}}{\beta_{L+4}-}...,\tag{89}$$

for $L=+2, +4, +6,...$ and

$$\frac{a_L}{a_{L+2}}=\frac{-\alpha_L}{\beta_L-}\frac{\alpha_{L-2}\gamma_L}{\beta_{L-2}-}\frac{\alpha_{L-4}\gamma_{L-2}}{\beta_{L-4}-}...,\tag{90}$$

for $L=-2, -4, -6,...$. The recursion equation at $L=0$ requires

$$\beta_0=-\alpha_0\frac{a_2}{a_0}-\gamma_0\frac{a_{-2}}{a_0}.\tag{91}$$

Substituting the right-hand sides of Eqs. (89) and (90) into (91), we obtain an implicit characteristic equation for ν that must be satisfied if the series (86) is to converge[38]:

$$\beta_0=\frac{\alpha_{-2}\gamma_0}{\beta_{-2}-}\frac{\alpha_{-4}\gamma_{-2}}{\beta_{-4}-}\frac{\alpha_{-6}\gamma_{-4}}{\beta_{-6}-}...+\frac{\alpha_0\gamma_2}{\beta_2-}\frac{\alpha_2\gamma_4}{\beta_4-}\frac{\alpha_4\gamma_6}{\beta_6-}....\tag{92}$$

The α_L, β_L and γ_L are given by Eqs. (88).

Existence and uniqueness: We do not have a formal proof that Eq. (92) actually has solutions in the parameter ν. However, solutions can be found numerically, so they must exist. Expansions (86) and (87) must reduce to the ordinary solutions (84) in the limit as the dipole moment $2aN_1$ goes to zero, so $\nu=m$ must be a solution to (92) at that limit. Knowing the approximate location of a root to a nonlinear equation is the first step towards finding it, and it is not difficult to start at a given ω with $\nu=m$, $N_1=-N_2=0$, A_{lm} an eigenvalue of the ordinary prolate spheroidal wave equation, and then track the values of ν that solve Eq. (92) as the dipole moment is gradually increased.

The solutions ν are not unique, but rather are periodic with period 1: if ν is a solution to Eq. (92), then so is $\nu\pm n$, where n is any integer. The correct choice of ν depends on how the coefficients a_L are generated. The a_L are to be minimal as $L\to\pm\infty$, and if the a_L are generated by forward recursion from $L=-\infty$ upward to $L=0$ and by backward recursion from $L=+\infty$ downward to $L=0$, then the largest a_L will be $a_{\max}=a_0$. As $\omega\to 0$ this $a_{\max}=a_0$ becomes the only coefficient that contributes to the series, and *with this choice* of $a_{\max}=a_0$ the correct limiting value of ν as the dipole moment $2aN_1$ reduces to zero is $\nu=l+m$. Solutions ν that do not reduce to $l+m$ as $2aN_1\to 0$ are spurious; they will enable expansions (86) and (87) to converge, but the resulting expressions will not solve the differential equation. That $\nu=l+m$ is the correct limit as $\omega\to 0$ or as $2aN_1\to 0$ will be demonstrated at the end of the section. This value may appear inconsistent with expansion (84), but it is not. For $|\omega|<1$, the d_n of (84) will have a maximum at $d_{\max}=d_l$ or d_{l-1}, and for $2aN_1=0$ and $\omega\neq 0$ the a_L of (86) will become zero only for $L<-l$ or $-l+1$. The two expansions are the same; it is only the indexing that is different.

C. Convergence properties

To establish the convergence of the series (86) and (87) it is necessary to examine the ratios

$$\lim_{L\to\pm\infty}\frac{a_L f_{L+\nu}(\omega\lambda)}{a_{L-2}f_{L+\nu-2}(\omega\lambda)},$$

where $f_{L+\nu}$ is either $j_{L+\nu}$ or $y_{L+\nu}$. We assume that ν has been chosen to satisfy Eq. (92), so that the a_L are minimal as $L\to\pm\infty$ and have the limiting behavior

$$\lim_{L\to+\infty}\frac{a_L}{a_{L-2}}=\frac{\omega^2}{4L^2}\left[1-\frac{2\nu+1}{L}+O\left(\frac{1}{L^2}\right)\right],$$

and

$$\lim_{L\to-\infty}\frac{a_L}{a_{L+2}}=\frac{\omega^2}{4L^2}\left[1-\frac{2\nu+1}{L}+O\left(\frac{1}{L^2}\right)\right].\tag{93}$$

The $j_{L+\nu}$ and the $y_{L+\nu}$ are both solutions to the recurrence relation

$$\frac{1}{2L+2\nu+3}f_{L+\nu+2}+\frac{1}{2L+2\nu-1}f_{L+\nu-2}+\left[\frac{4L+4\nu+2}{(2L+2\nu+3)(2L+2\nu-1)}-\frac{2L+2\nu+1}{\omega^2\lambda^2}\right]f_{L+\nu}=0,\tag{94}$$

the $j_{L+\nu}$ being the particular solution sequence that is minimal as $L\to+\infty$. In general, the $j_{L+\nu}$ and the $y_{L+\nu}$ will both be dominant as $L\to-\infty$ (the exception being the inconsequential case when ν is an odd multiple of $\frac{1}{2}$), and we consider the following three cases.

(1) $f_{L+\nu}(\omega\lambda)=j_{L+\nu}(\omega\lambda)$ and $L\to+\infty$. From Eq. (94) the ratio $j_{L+\nu}/j_{L+\nu-2}$ has the limiting forms

$$\lim_{L\to\infty}\frac{j_{L+\nu}(\omega\lambda)}{j_{L+\nu-2}(\omega\lambda)}=\begin{cases}\frac{\omega^2\lambda^2}{4L^2}\left[1-\frac{2\nu}{L}+O\left(\frac{1}{L^2}\right)\right] & (L\gg|\omega\lambda|),\\ -1 & (L\ll|\omega\lambda|).\end{cases}$$

In either case

$$\lim_{L\to\infty}\left|\frac{a_L j_{L+\nu}(\omega\lambda)}{a_{L-2}j_{L+\nu-2}(\omega\lambda)}\right|<\left|\frac{\omega^2}{4L^2}\right|,\tag{95}$$

and this part of the series is absolutely convergent for all λ. If ν is an integer, the negative L part of the series (86) truncates, and (95) describes the convergence of the regular solution (84) to the ordinary spheroidal wave equation.

(2) $f_{L+\nu}(\omega\lambda)=y_{L+\nu}(\omega\lambda)$ and $L\to+\infty$. Again from Eq. (94) we find

$$\lim_{L\to\infty} \frac{y_{L+\nu}(\omega\lambda)}{y_{L+\nu-2}(\omega\lambda)} = \begin{cases} \frac{4L^2}{\omega^2\lambda^2}\left[1+\frac{2\nu}{L}+O\left(\frac{1}{L^2}\right)\right] & (L\gg|\omega\lambda|), \\ -1 & (L\ll|\omega\lambda|). \end{cases}$$

Hence

$$\lim_{L\to\infty} \frac{a_L y_{L+\nu}(\omega\lambda)}{a_{L-2} y_{L+\nu-2}(\omega\lambda)} = \begin{cases} \frac{1}{\lambda^2}\left[1-\frac{1}{L}+O\left(\frac{1}{L^2}\right)\right] & (L\gg|\omega\lambda|), \\ -\frac{\omega^2}{4L^2} & (L\ll|\omega\lambda|), \end{cases}$$

and the series (87) converges rapidly for large λ and mediocre ω, but diverges when $|\lambda| = 1$. Explicitly, for every λ such that $|\lambda| > 1$ and for every $\epsilon > 0$ there exits an $N(\omega,\lambda,\epsilon)$ such that

$$\sum_{L=N}^{\infty}{}' a_L y_{L+\nu}(\omega\lambda) < \epsilon.$$

For each $\epsilon > 0$, N will increase without bound as $\lambda \to 1$.

Contrary assertions[39] notwithstanding, $\Sigma' a_L y_{L+\lambda}(\omega\lambda)$ is absolutely convergent for all $|\lambda| > 1$, and is in no sense asymptotic: N does not go to zero as $\lambda \to \infty$. This convergence was demonstrated numerically by Sinha and MacPhie,[40] but to my knowledge the present analysis constitutes the first rigorous proof. Jen and Hu[41] have recently derived accurate approximations for the ordinary spheroidal wave functions (both regular and irregular) that are rapidly convergent for large values of ω, where the convergence of Stratton's expansion is rather slow.

(3) $f_{L+\nu}(\omega\lambda) = j_{L+\nu}(\omega\lambda)$ or $f_{L+\nu}(\omega\lambda) = y_{L+\nu}(\omega\lambda)$ and $L \to -\infty$. This case is similar to the one just discussed. Equation (94) once more yields

$$\lim_{L\to-\infty} \frac{f_{L+\nu}(\omega\lambda)}{f_{L+\nu+2}(\omega\lambda)} = \begin{cases} \frac{4L^2}{\omega^2\lambda^2}\left[1+\frac{2\nu+4}{L}+O\left(\frac{1}{L^2}\right)\right] & (|L|\gg|\omega\lambda|), \\ -1 & (|L|\ll|\omega\lambda|), \end{cases}$$

which together with Eq. (93) gives

$$\lim_{L\to-\infty} \frac{a_L f_{L+\nu}(\omega\lambda)}{a_{L+2} f_{L+\nu+2}(\omega\lambda)} = \begin{cases} \frac{1}{\lambda^2}\left[1+\frac{3}{L}+O\left(\frac{1}{L^2}\right)\right] & (|L|\gg|\omega\lambda|) \\ -\frac{\omega^2}{4L^2} & (|L|\ll|\omega\lambda|). \end{cases}$$

Therefore the negative L part of either series (86) or (87) behaves like the positive L part of the series (87) discussed previously, and series (86) and (87) are two independent and convergent irregular solutions to our preliminary generalization of the ordinary spheroidal wave equation.

D. Solutions as $\omega \to 0$

(1) $N_1 = -N_2 = 0$. When $N_1 = -N_2 = 0$ and ω is very small, Eq. (78) reduces to

$$\frac{d}{d\mu}\left[(1-\mu^2)\frac{d\Phi}{d\mu}\right] + \left[A_{lm} - \frac{m^2}{1-\mu^2}\right]\Phi = 0, \tag{96}$$

and with the substitution $\Phi(\mu) = (1-\mu^2)^{m/2} g(\mu)$ the differential equation for g becomes

$$(1-\mu^2) g_{,\mu\mu} - 2(m+1)\mu g_{,\mu} - [m(m+1) - A_{lm}] g = 0. \tag{97}$$

Hence

$$\lim_{\omega\to 0} \Phi(\mu) = (1-\mu^2)^{m/2} T_l^m(\mu),$$

and

$$\lim_{\omega\to 0} A_{lm} = (l+m)(l+m+1).$$

This limiting value for A_{lm} may also be obtained by setting $\omega = 0$ and $B_2 = 2m+2$ in Eq. (27), then finding the B_3 that truncates the series (25) by making $\beta_l^\theta = 0$. Equation (83) then gives

$$\lim_{\omega\to 0} \Psi(\lambda) = \left(\frac{\lambda^2 - 1}{\lambda^2}\right)^{m/2} j_{l+m}(\omega\lambda),$$

and we must have $\nu \to l + m$ as $\omega \to 0$ when N_1 and N_2 are zero if we are to keep a_0 the maximum term in series (86).

(2) $N_1 = -N_2 \neq 0$. By Eqs. (89) and (90) we see that if we fix the largest expansion coefficient to be $a_{\max} = a_0 = 1$ in (86), then all the other a_L must become zero as $\omega \to 0$ and the series will reduce to the single term

$$\lim_{\omega\to 0} \Psi_1(\lambda) = \left(\frac{\lambda^2 - 1}{\lambda^2}\right)^{m/2} a_0 j_{\nu_0}(\omega\lambda). \tag{98}$$

This single term must also suffice as $\omega\lambda \to \infty$ ($|\omega|$still$\ll 1$), and ν_0 may be determined by looking at this limit. Asymptotic solutions to Eq. (82) are

$$\lim_{\lambda\to\infty} \Psi_1(\lambda) = \left(\frac{\lambda+1}{\lambda-1}\right)^{1/2} j_{\nu_a}(\omega(\lambda+1))[1 + O(\lambda^{-3})], \tag{99}$$

and

$$\lim_{\lambda\to\infty} \Psi_2(\lambda) = \left(\frac{\lambda+1}{\lambda-1}\right)^{1/2} y_{\nu_a}(\omega(\lambda+1))[1 + O(\lambda^{-3})], \tag{100}$$

where ν_a is a function of ω and takes the value $(-1 + \sqrt{1 + 4A_{lm}})/2$ when $\omega = 0$ [see Eqs. (6) and (22) with $zj_\nu(z) = F_\nu(0,z)$]. The $\omega \to 0$ limit ν_0 must equal this ν_a, which is consistent with our previous determination that ν_0 should equal $l + m$ in the special case when $N_1 = -N_2 = 0$ and $A_{lm} = (l+m)(l+m+1)$. That ν_0 must equal ν_a may be recognized by considering λ large enough that $[(\lambda+1)/(\lambda-1)]^{1/2} \approx [1 - \lambda^{-2}]$, and ω small enough that $\omega\lambda \ll 1$. The asymptotic solutions (99) and (100) are valid in this region, but can represent the regular solution (98) only if the orders ν_a and ν_0 are the same. This is because $j_\nu(x) \propto x^\nu$ for small x. We should not be surprised to find that ν_0 will become complex when the dipole moment $2aN_1$ is greater than the critical value that makes $A_{lm} < -\frac{1}{4}$. Only those values of ν that are solutions to Eq. (92) and are contiguous with ν_a as $\omega \to 0$ can be used to generate a sequence a_L that allows expansions (86) and (87) to give true solutions to the differential equation (82). In other words, if one wants to find a ν with which to generate solutions to the differential equation (82) via expansions (86) and (87) for some nonzero (and perhaps even complex) ω, one should first solve Eq. (91) for an ω very near zero using the ν_a given at Eq. (23) as a starting point, then move ω toward the desired value in small increments, re-solving (91) at each step to make certain the final value of ν obtained is on the correct branch.

VI. SOLUTIONS BY EXPANSION IN SERIES OF COULOMB WAVE FUNCTIONS

Heartened by success with the preliminary generalization discussed in Sec. V, one is immediately tempted to try an expansion of the same sort as (86) and (87) for the general case of Eq. (2) when $2a(N_1+N_2)\neq 0$. In terms of the problem of the two-center Schrödinger equation, this full generalization implies a net Coulomb charge on the nuclei, hence the expansion must be in terms of Coulomb wave functions rather than spherical Bessel functions. This is not a great conceptual complication, since the Coulomb wave function $F_L(\eta,\rho)$ and the spherical Bessel function $j_L(\rho)$ are related when $\eta=0$ by $F_L(0,\rho)=\rho j_L(\rho)$. Unfortunately, if the expansion

$$\Psi(\lambda)=\frac{1}{\lambda}\left(\frac{\lambda^2-1}{\lambda^2}\right)^{(m/2)}\sum_{L=-\infty}^{\infty}a_L F_{L+\nu}(\eta,\omega\lambda)$$

[with $\eta=-a(N_1+N_2)/\omega$] is substituted into the differential equation (2), the resulting recurrence relation amongst the a_L will have five terms instead of three, being of the form

$$-\frac{\omega^2}{4}\alpha_L a_{L+2}-\frac{\omega^2}{L^2}\eta m\,\alpha'_L\,a_{L+1}+L^2\beta_L a_L+\frac{\omega^2}{L^2}\eta m\,\gamma'_L\,a_{L-1}-\frac{\omega^2}{4}\gamma_L a_{L-2}=0,$$

where the α_L, α'_L, β_L, γ'_L, and γ_L are functions of L, ν, and the parameters of the differential equation and are normalized such that they each approach 1 as $L\to\infty$. Exact expressions for these recurrence coefficients are nearly as ghastly to derive as they are to contemplate once written down, and as we know of no reasonable computational method for dealing with five-term recurrence relations, we will spare the reader the agony of their further consideration and turn instead to the presentation of a more elegant representation for the generalized spheroidal wave functions.

A. The Coulomb wave-function expansion

Equation (1) was

$$x(x-x_0)\frac{d^2y}{dx^2}+(B_1+B_2x)\frac{dy}{dx}+[\omega^2x(x-x_0)-2\eta\omega(x-x_0)+B_3]y=0.$$

With the substitutions $y(x)=x^{-B_2/2}h(x)$ and $z=\omega x$ the differential equation becomes

$$z(z-\omega x_0)[h_{,zz}+(1-2\eta/z)h]+C_1\omega h_{,z}+(C_2+C_3\omega/z)h=0, \tag{101}$$

where

$$C_1=B_1+B_2x_0,$$
$$C_2=B_3-\tfrac{1}{2}B_2(\tfrac{1}{2}B_2-1), \tag{102}$$
$$C_3=-\tfrac{1}{2}B_2[x_0(\tfrac{1}{2}B_2+1)+B_1].$$

The function $h(z)$ can then be expanded in a series of Coulomb wave functions:

$$h(z)=\sum_{L=-\infty}^{\infty}a_L u_{L+\nu}(z), \tag{103}$$

where $u_{L+\nu}(z)$ is any combination of the Coulomb wave functions $F_{L+\nu}(\eta,z)$ and $G_{L+\nu}(\eta,z)$. The $\eta=0$, $\nu=l+m$ ordinary spheroidal wave-function limit of this expansion can no doubt be obtained by an integral transformation of Eq. (30) or (33) of Baber and Hassé, but there is little to be gained by further consideration of such special cases. The Coulomb wave functions satisfy the recurrence relation

$$\frac{1}{2L+2\nu+1}R_{L+1}u_{L+\nu+1}-\left(\frac{1}{z}+Q_L\right)u_{L+\nu}+\frac{1}{2L+2\nu+1}R_L u_{L+\nu-1}=0, \tag{104}$$

and the differential relation

$$\frac{d}{dz}u_{L+\nu}=-\frac{L+\nu}{2L+2\nu+1}R_{L+1}u_{L+\nu+1}-Q_L u_{L+\nu}+\frac{L+\nu+1}{2L+2\nu+1}R_L u_{L+\nu-1}, \tag{105}$$

where

$$Q_L=\eta/[(L+\nu)(L+\nu+1)]$$

and

$$R_L=[(L+\nu)^2+\eta^2]^{1/2}/(L+\nu), \tag{106}$$

and are solutions of the differential equation

$$\frac{d^2}{dz^2}u_{L+\nu}+\left[1-\frac{2\eta}{z}-\frac{(L+\nu)(L+\nu+1)}{z^2}\right]u_{L+\nu}=0. \tag{107}$$

The $F_{L+\nu}(z)$ form a solution sequence to recurrence relation (104) that is minimal as $L\to+\infty$, and are the solution to the Coulomb wave equation (107) that is proportional to $z^{L+\nu+1}$ as $z\to 0$. The $G_{L+\nu}(z)$ are irregular solutions to the Coulomb wave equation, are proportional to $z^{-L-\nu}$ as $z\to 0$, and form a dominant solution sequence to recurrence relation (104). The $F_{L+\nu}(z)$ and $G_{L+\nu}(z)$ are normalized such that the Wronskian

$$F_{L+\nu,z}G_{L+\nu}-G_{L+\nu,z}F_{L+\nu}=1, \tag{108}$$

and have the asymptotic form

$$G_{L+\nu}(\eta,z)\pm iF_{L+\nu}(\eta,z)\underset{z\to\infty}{\to}\exp[\pm i(z-\eta\ln 2z-(L+\nu)(\pi/2)+\sigma_L)], \tag{109}$$

where

$$\sigma_L=-\frac{i}{2}\ln\left[\frac{\Gamma(L+\nu+1+i\eta)}{\Gamma(L+\nu+1-i\eta)}\right]. \tag{110}$$

Coulomb wave functions are defined by the integral representations

$$G_{L+\nu} \pm iF_{L+\nu} = \frac{e^{\pi\eta/2} e^{\pm iz} (2z)^{-L-\nu}}{[\Gamma(L+\nu+1+i\eta)\Gamma(L+\nu+1-i\eta)]^{1/2}} \times \int_0^\infty e^{-t} t^{L+\nu\pm i\eta} (t \mp 2iz)^{L+\nu\mp i\eta} dt, \quad (111)$$

and afford an alternate way of expressing the confluent hypergeometric functions. [The Coulomb wave functions have usually been defined only for non-negative integer orders and real charge parameter η. Equations (110) and (111) were obtained from the discussion of Coulomb wave functions and confluent hypergeometric functions given by Morse and Feshbach, who define Coulomb wave functions in a completely analytic manner.]

The expansion coefficients a_L in series (103) are defined by the recurrence relation

$$\alpha_L a_{L+1} + \beta_L a_L + \gamma_L a_{L-1} = 0, \quad (112)$$

where

$$\alpha_L = -\omega R_{L+1}/(2L+2\nu+3) \times [(L+\nu+1)(L+\nu+2)x_0 - (L+\nu+2)C_1 - C_3],$$

$$\beta_L = (L+\nu)(L+\nu+1) + C_2 + \omega Q_L [(L+\nu)(L+\nu+1)x_0 - C_1 - C_3],$$

$$\gamma_L = -\omega R_L/(2L+2\nu-1) \times [(L+\nu)(L+\nu-1)x_0 + (L+\nu-1)C_1 - C_3].$$

The C_1, C_2, and C_3 are given in terms of the B_1, B_2, and B_3 in Eqs. (102). The a_L will be minimal as $L \to \pm\infty$ if ν is a solution of the implicit equation

$$\beta_0 = \frac{\alpha_{-1}\gamma_0}{\beta_{-1}-} \frac{\alpha_{-2}\gamma_{-1}}{\beta_{-2}-} \frac{\alpha_{-3}\gamma_{-2}}{\beta_{-3}-} \cdots + \frac{\alpha_0\gamma_1}{\beta_1-} \frac{\alpha_1\gamma_2}{\beta_2-} \frac{\alpha_2\gamma_3}{\beta_3-} \cdots. \quad (113)$$

Solutions to (113) exist and are periodic with period 1. Roots ν that are integer multiples of $\frac{1}{2}$ are usually spurious (there are some special exceptions). The only solutions ν for which the series (103) will actually solve the differential equation are those that map to the correct asymptotic values as $\omega \to 0$ or as $\omega x \to \infty$. See Eq. (121).

B. Convergence properties

Convergence of the Coulomb wave-function series solutions (103) to the generalized spheroidal wave equation (1) is similar to that of the Neumann series solutions (86) to the finite dipole wave equation (82). When ν is a solution to the continued fraction equation (113) the sequence of expansion coefficients $\{a_L: L = ..., -2, -1,0,1,2,...\}$ is minimal as $L \to \pm\infty$, but in the usual case that ν is not an integer, the negative L part of the series cannot be truncated and both sequences of Coulomb wave functions $\{F_{L+\nu}: L = ..., -2, -1,0,1,2,...\}$ and $\{G_{L+\nu}: L = ..., -2, -1,0,1,2,...\}$ will be dominant either as $L \to +\infty$ or as $L \to -\infty$, or both. The solutions given by expansion (103), though independent, will both be seen to be irregular as $x \to x_0$.

Convergence properties of the solutions are illustrated by analysis of the limiting behavior of $a_L G_{L+\nu}(\eta,\omega x)$ as $L \to +\infty$. This behavior will be shared by the $L \to -\infty$ part of both series, and the positive L part of $\Sigma_L a_L F_{L+\nu}$ is obviously convergent. We will here consider only the case when $L^2 \gg |\eta\omega|$. From the recurrence relation (112) for the a_L, we obtain the limiting ratios

$$\lim_{L\to+\infty} \frac{a_L}{a_{L-1}} = \frac{\omega R_L}{2L^2} [x_0 L + C_1], \quad (114)$$

and

$$\lim_{L\to-\infty} \frac{a_L}{a_{L+1}} = \frac{\omega R_L}{2L^2} [x_0 L - C_1] \quad (115)$$

(the a_L being minimal as $L \to \pm\infty$), and from the recurrence relation (104) for the Coulomb wave functions

$$\lim_{L\to\infty} \frac{G_{L+\nu}(\eta,\omega x)}{G_{L+\nu-1}(\eta,\omega x)} \sim \frac{2L}{R_L}\left(\frac{1}{\omega x} + Q_{L-1}\right) \quad (L \gg |\omega x|),$$

and

$$\lim_{L\to\infty} \frac{G_{L+\nu+1}(\eta,\omega x)}{G_{L+\nu-1}(\eta,\omega x)} = -1 \quad (1 \ll L \ll |\omega x|),$$

the $G_{L+\nu}$ being dominant solutions of recurrence relation (104). If $\eta\omega \ll L^2$, we immediately obtain

$$\lim_{L\to\infty} \frac{a_L G_{L+\nu}(\omega x)}{a_{L-1} G_{L+\nu-1}(\omega x)} = \frac{1}{x}\left(x_0 + \frac{C_1}{L}\right) \to \begin{cases} x_0/x, & \text{if } x_0 \neq 0, \\ C_1/xL, & \text{if } x_0 = 0, \end{cases} \quad (116)$$

so that the series (103) is absolutely convergent for all $x > x_0$ and diverges at $x = x_0$. The $L < 0$ part of the series can be truncated when ν is an integer, and this will happen when $\eta = 0$ and A_{lm} is an eigenvalue of the ordinary spheroidal wave equation. If $\nu = l + m$ then the series $\Sigma_{L=0}^{\infty} a_L F_{L+l+m}(0,\omega x)$ will represent the regular ordinary spheroidal wave function.

Expressing, as it does, the limiting form of the ratio of successive terms in the series, expression (116) tells us not only that the series converges, but also says much about how rapid the convergence is. If for some $\epsilon > 0$ we wish to find an N such that

$$|a_N G_{N+\nu}(\eta,\omega x)/a_0 G_\nu(\eta,\omega x)| < \epsilon,$$

then we can use (116) to estimate

$$|a_N G_{N+\nu}(\omega x)/a_0 G_\nu(\omega x)| = \epsilon \approx (x_0/x)^N \quad (117)$$

(assuming $|C_1| \ll N x_0$), from which

$$N \approx \log \epsilon / \log(x_0/x), \quad (118)$$

which again expresses the divergence of the series as $x \to x_0$. As an example, suppose one wished to compute the irregular ordinary spheroidal wave function (with $l = m = 0$) for $\omega = 1$ at $\lambda = 3$ to an accuracy of approximately seven decimal places. In this case $x_0 = 2$, $x = 4$, and $\epsilon = 10^{-7}$. Equation (118) then estimates $N \approx 23$. When actually computed, it turns out that $a_{23} = 1.51 \times 10^{-23}$ and $G_{23}(0,4) = 4.31 \times 10^{14}$. The product of the two, $a_{23} G_{23}(0,4) = 6.5 \times 10^{-9}$, is not unreasonably distant from the desired value 10^{-7}. Thus expression (118), though not exact, is a useful guide for computational purposes.

C. Solutions as $\omega \to 0$

The solutions ν to Eq. (113) are periodic with period 1. As $\omega \to 0$ the differential equation (101) takes the limiting form

$$\lim_{\omega \to 0} \frac{d^2 h}{dz^2} + \left(1 - \frac{2\eta}{z} + \frac{C_2}{z^2}\right) h = 0, \quad (119)$$

and has solutions

$$\lim_{\omega \to 0} h_1(z) = F_{\nu_0}(\eta,z)$$

and

$$\lim_{\omega \to 0} h_2(z) = G_{\nu_0}(\eta,z), \quad (120)$$

where

$$\nu_0 = -\tfrac{1}{2}[1 \pm \sqrt{1 - 4C_2}]. \quad (121)$$

If we normalize the a_L such that $a_{\max} = a_0$, then the minimal solution sequence a_L to recurrence relation (112) has the property that $a_L \to 0$ as $\omega \to 0$ for all $L \neq 0$. The small ω form of expansion (103) will then be dominated by the single term $a_0 u_\nu(\eta,\omega x)$. For larger ω the only solutions ν to Eq. (113) that can be used to generate the a_L and give a convergent series that actually solves the differential equation are those ν that are contiguous with the ν_0 as $\omega \to 0$. The values of ν_0 are the same as the values of the order ν_a of the asymptotic solutions (22) and (23) given in Sec. II.

D. Asymptotic behavior

Two independent solutions to the generalized spheroidal wave equation

$$x(x - x_0)\frac{d^2 y}{dx^2} + (B_1 + B_2 x)\frac{dy}{dx} + [\omega^2 x(x - x_0) - 2\eta\omega(x - x_0) + B_3] y = 0$$

can now be written

$$y_+(x) = x^{-B_2/2} \sum_{L=-\infty}^{\infty} a_L [G_{L+\nu}(\eta,z) + iF_{L+\nu}(\eta,z)],$$

$$y_-(x) = x^{-B_2/2} \sum_{L=-\infty}^{\infty} a_L [G_{L+\nu}(\eta,z) - iF_{L+\nu}(\eta,z)]. \quad (122)$$

The asymptotic form of y_+ and y_- can be expressed as

$$\lim_{x \to \infty} y_\pm(x) = x^{-B_2/2} \exp[\pm i(\omega x - \eta \ln(2\omega x) - \phi_\pm)], \quad (123)$$

where ϕ_+ and ϕ_- are obtained from Eqs. (109) and (122):

$$\phi_\pm = \pm i \ln\left[\sum_{L=-\infty}^{\infty} a_L \exp \mp i\left[(L+\nu)\frac{\pi}{2} - \sigma_L\right]\right]. \quad (124)$$

The σ_L are defined by Eq. (110). The asymptotic behavior of any combination of these solutions $Y(x) = A_{\text{out}} y_+(x) + A_{\text{in}} y_-(x)$ is therefore obtainable from the values of $Y(x)$ and $Y_{,x}(x)$ at any convenient x at which the matching coefficients A_{out} and A_{in} may be determined. Expressions (122), (123), and (124) are all analytic in ω and x, so there is no reason the phase of the x at which the matching is done need be the same as the phase of the asymptotic limit desired. This analyticity with the asymptotic form is the crucial property that makes a representation "truly useful," and the Coulomb wave-function expansions (122) probably express it as well as is possible by anything short of an actual integral representation for the generalized spheroidal wave functions.

E. Values on the ω branch cut

The irregular generalized spheroidal wave functions have branch cuts in x that emanate from $x = 0$ and $x = x_0$. These may be treated in the usual manner using the known values of the indices k_1 and k_2 of Eq. (21). There is, however, a branch cut in the frequency ω that is an important consideration in some physical problems.[4] This branch cut starts at $\omega = 0$ and extends downward along the negative imaginary ω axis. The Coulomb wave-function expansions (122) allow the values of the irregular generalized spheroidal wave functions y_+ and y_- to be determined on each side of this cut. In terms of the regular and irregular confluent hypergeometric functions $M(a,b,2iz)$ and $U(a,b,2iz)$ as defined by Slater,[28] the Coulomb wave functions can be expressed as

$$G_{L+\nu}(\eta,z) \pm iF_{L+\nu}(\eta,z) = (-)^L e^{\pi\eta/2} e^{\mp i\pi(\nu+1/2)} \left[\frac{\Gamma(L+\nu+1 \pm i\eta)}{\Gamma(L+\nu+1 \mp i\eta)}\right]^{1/2}$$
$$\times (2z)^{L+\nu+1} e^{\pm iz} U(L+\nu+1 \pm i\eta, 2L+2\nu+2, \mp 2iz) \quad (125)$$

and

$$F_{L+\nu}(\eta,z) = \frac{[\Gamma(L+\nu+1+i\eta)\Gamma(L+\nu+1-i\eta)]^{1/2}}{2e^{\pi\eta/2}\Gamma(2L+2\nu+2)} (2z)^{L+\nu+1} e^{\pm iz} M(L+\nu+1 \pm i\eta, 2L+2\nu+2, \mp 2iz). \quad (126)$$

Slater's Eq. 13.1.10 then gives the ω branch cut information

$$U(L+\nu+1 \pm i\eta, 2L+2\nu+2, \mp 2ize^{2n\pi i})$$
$$= e^{-4n\pi i\nu} U(L+\nu+1 \pm i\eta, 2L+2\nu+2, \mp 2iz)$$
$$+ (1 - e^{-4n\pi i\nu})[\Gamma(-2L-2\nu-1)/\Gamma(-L-\nu \pm i\eta)] M(L+\nu+1 \pm i\eta, 2L+2\nu+2, \mp 2iz). \quad (127)$$

Recall that $z=\omega x$. Equations (125) and (127) can be inserted into expansions (122), and the reflection property of the gamma function eventually allows the desired result:

$$y_\pm(\omega e^{2n\pi i}) = e^{-2n\pi i\nu}y_\pm(\omega) + \sin 2n\pi\nu \csc 2\pi\nu \times (e^{2\pi\eta} - e^{\mp 2\pi i\nu})[y_+(\omega) - y_-(\omega)]. \tag{128}$$

Here $y_+(\omega) - y_-(\omega) = 2ix^{-B_2/2}\Sigma a_L F_{L+\nu}(\eta,\omega x)$. This expression is also valid in the limit when ν is an integer. It should be kept in mind that the expansion coefficients a_L and the phase parameter ν are implicit functions of ω. They appear, however, to be entire, and their values do not change across the ω branch cut.

F. An alternate normalization for the Coulomb wave functions

There exist other normalizations for the Coulomb wave functions that should have definite computational advantages over the usual Coulomb wave functions discussed above. These are exemplified by a normalization first proposed by Gautschi,[6] who defined functions $f_{L+\nu}$ and $g_{L+\nu}$ by

$$f_{L+\nu}(\eta,z) = (2L+2\nu+1)e^{\pi\eta/2}\frac{\Gamma(L+\nu+1)}{[\Gamma(L+\nu+1+i\eta)\Gamma(L+\nu+1-i\eta)]^{1/2}}F_{L+\nu}(\eta,z), \tag{129}$$

$$g_{L+\nu}(\eta,z) = (2L+2\nu+1)e^{\pi\eta/2}\frac{\Gamma(L+\nu+1)}{[\Gamma(L+\nu+1+i\eta)\Gamma(L+\nu+1-i\eta)]^{1/2}}G_{L+\nu}(\eta,z). \tag{130}$$

The factor $(2L+2\nu+1)e^{\pi\eta/2}$ is not absolutely necessary, but it does no harm to retain it. The differential and recurrence relations obeyed by both $f_{L+\nu}$ and $g_{L+\nu}$ are

$$f_{L+\nu,z} = \frac{L+\nu+1}{2L+2\nu-1}f_{L+\nu-1} - Q_L f_{L+\nu} - \frac{L+\nu}{2L+2\nu+3}\left[1+\frac{\eta^2}{(L+\nu+1)^2}\right]f_{L+\nu+1}, \tag{131}$$

$$\frac{1}{z}f_{L+\nu} = \frac{1}{2L+2\nu-1}f_{L+\nu-1} - Q_L f_{L+\nu} + \frac{1}{2L+2\nu+3}\left[1+\frac{\eta^2}{(L+\nu+1)^2}\right]f_{L+\nu+1}, \tag{132}$$

where $Q_L = \eta/[(L+\nu)(L+\nu+1)]$ as in Eq. (106). Asymptotic forms for f and g may be expressed as

$$g_{L+\nu}(\eta,z) \pm i f_{L+\nu}(\eta,z) \xrightarrow{z\to\infty} \exp\left[\pm i\left(z - \eta\ln 2z - (L+\nu)\frac{\pi}{2} + \sigma_L^{(\pm)}\right)\right], \tag{133}$$

where

$$\sigma_L^{(\pm)} = \mp i\ln\left[(2L+2\nu+1)e^{\pi\eta/2} \times \frac{\Gamma(L+\nu+1)}{\Gamma(L+\nu+1\mp i\eta)}\right].$$

If we write our solutions (122) to the generalized spheroidal wave equation as

$$y_\pm(x) = x^{-B_2/2}\sum_{L=-\infty}^{\infty} a_L[g_{L+\nu}(\eta,z) \pm i f_{L+\nu}(\eta,z)], \tag{134}$$

then the asymptotic forms of y_+ and y_- are given by

$$\lim_{x\to\infty} y_\pm(x) = x^{-B_2/2}\exp[\pm i(\omega x - \eta\ln(2\omega x) - \tilde{\phi}_\pm)], \tag{135}$$

where

$$\tilde{\phi}_\pm = \pm i\ln\left[\sum_{L=-\infty}^{\infty} a_L \exp \mp i\left[(L+\nu)\frac{\pi}{2} - \sigma_L^{(\pm)}\right]\right]. \tag{136}$$

The expansion coefficients a_L will satisfy the three-term recurrence relation

$$\alpha_L a_{L+1} + \beta_L a_L + \gamma_L a_{L-1} = 0, \tag{137}$$

where now the recurrence coefficients are defined by

$$\alpha_L = -\omega/(2L+2\nu+1) \times[(L+\nu+1)(L+\nu+2)x_0 - (L+\nu+2)C_1 - C_3],$$

$$\beta_L = (L+\nu)(L+\nu+1) + C_2 + \omega Q_L[(L+\nu)(L+\nu+1)x_0 - C_1 - C_3],$$

$$\gamma_L = -\omega/(2L+2\nu+1) \times[(L+\nu)(L+\nu-1)x_0 + (L+\nu-1)C_1 - C_3] \times[1+\eta^2/(L+\nu)^2].$$

The C_1, C_2, and C_3 are given in terms of the B_1, B_2, and B_3 in Eqs. (102). The recurrence relations (131) and (132) for Gautschi's Coulomb wave functions should be compared with the corresponding relations (104) and (105) for the usual Coulomb wave functions, and the definitions of the α_L, β_L, and γ_L for Eq. (137) compared with the corresponding definitions for the α_L, β_L, and γ_L for Eq. (112). No square roots appear in any of the relations using Gautschi's normalization, a property that will greatly enhance the speed with which expansions (134) can be evaluated and, by eliminating spurious branch cuts associated with the unnecessary square roots, will probably enlarge the parameter regions for which the Coulomb wave-function expansion is valid.

VII. SOLUTIONS BY EXPANSION IN SERIES OF CONFLUENT HYPERGEOMETRIC FUNCTIONS

One last set of representations for the generalized spheroidal wave functions may be obtained by expanding the solutions $y(x)$ to Eq. (1) in series of the confluent hypergeometric functions $\tilde{M}(a,b,z)$ and $U(a,b,z)$. Four new representations are obtained. The expansions for the solution that is regular as $x \to x_0$ are shown to be uniformly convergent both at $x = x_0$ and as $x \to \infty$. However, convergence of these series does not appear to be rapid, and the representations have not yet been implemented on a computer.

A. The confluent hypergeometric function expansion

Again, start with Eq. (1):

$$x(x-x_0)\frac{d^2y}{dx^2} + (B_1 + B_2x)\frac{dy}{dx} + [\omega^2 x(x-x_0) - 2\eta\omega(x-x_0) + B_3]y = 0.$$

Solutions can be expanded in the form

$$y_1(x) = e^{+i\omega x}\sum_{L=-\infty}^{\infty} a_L \tilde{M}\left(\frac{1}{2}B_2 + i\eta, L + \nu_1, -2i\omega x\right), \quad (138)$$

$$y_2(x) = e^{-i\omega x}\sum_{L=-\infty}^{\infty} b_L \tilde{M}\left(\frac{1}{2}B_2 - i\eta, L + \nu_2, +2i\omega x\right), \quad (139)$$

$$y_+(x) = e^{+i\omega x}\sum_{L=-\infty}^{\infty} a_L U\left(\frac{1}{2}B_2 + i\eta, L + \nu_1, -2i\omega x\right), \quad (140)$$

$$y_-(x) = e^{-i\omega x}\sum_{L=-\infty}^{\infty} b_L U\left(\frac{1}{2}B_2 - i\eta, L + \nu_2, +2i\omega x\right). \quad (141)$$

To demonstrate this, substitute

$$y(x) = e^{+i\omega x}f(x), \quad z = -2i\omega x, \quad \text{and} \quad z_0 = -2i\omega x_0, \quad (142)$$

then the differential equation for f in terms of z is

$$z(z-z_0)\frac{d^2f}{dz^2} + (D_1 + D_2 z - z^2)\frac{df}{dz} + (D_3 + D_4 z)f = 0, \quad (143)$$

where

$$D_1 = -2i\omega B_1, \quad D_2 = B_2 - 2i\omega x_0, \quad D_3 = B_3 + 2\eta\omega x_0 + i\omega B_1, \quad D_4 = -\tfrac{1}{2}B_2 - i\eta. \quad (144)$$

The solutions $f(z)$ to Eq. (143) can be expanded in a series of the confluent hypergeometric functions $M_L(z)$ and $U_L(z)$, where M_L and U_L denote, respectively, the regular and irregular confluent hypergeometric functions

$$M_L(z) \equiv \tilde{M}(-D_4, L+\nu, -2i\omega x)$$

and

$$U_L(z) \equiv U(-D_4, L+\nu, -2i\omega x)$$

where $z = -2i\omega x$. Here $\tilde{M}(a,b,z)$ and $U(a,b,z)$ are defined by the integral representations

$$\tilde{M}(a,b,z) = \frac{1}{\Gamma(a)}\int_0^1 e^{zt}t^{a-1}(1-t)^{b-a-1}\,dt, \quad \mathrm{Re}(b) > \mathrm{Re}(a) > 0, \quad (145)$$

and

$$U(a,b,z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zt}t^{a-1}(1+t)^{b-a-1}\,dt, \quad \mathrm{Re}(a) > 0, \quad \mathrm{Re}(z) > 0. \quad (146)$$

Properties of these functions may be found in Slater.[28] The definition (145), used here for the regular confluent hypergeometric function $\tilde{M}(a,b,z)$, differs from the usual normalization of the Kummer series by a factor $\Gamma(b-a)/\Gamma(b)$:

$$\tilde{M}(a,b,z) = \frac{\Gamma(b-a)}{\Gamma(b)}\sum_{n=0}^{\infty}\frac{(a)_n z^n}{(b)_n n!},$$

where $(a)_n$ denotes Pochammer's symbol $(a)_n \equiv a(a+1)(a+2)\cdots(a+n-1)$ and $(a)_0 \equiv 1$. Thus normalized $\tilde{M}(a,b,z)$ obeys the same differential and recurrence relations as does $U(a,b,z)$. Let $\mathscr{F}_L(z)$ denote any linear combination of the $M_L(z)$ and $U_L(z)$ defined above. Then $\mathscr{F}_L$ solves the confluent hypergeometric equation

$$z\mathscr{F}_{L,zz} + (L+\nu-z)\mathscr{F}_{L,z} + D_4\mathscr{F}_L = 0, \quad (147)$$

satisfies the differential relations

$$\mathscr{F}_{L,z} = \mathscr{F}_L - \mathscr{F}_{L+1} \quad (148)$$

$$z\mathscr{F}_{L,z} = (1-L-\nu)\mathscr{F}_L + (L+\nu-1+D_4)\mathscr{F}_{L-1}, \quad (149)$$

and obeys the recurrence relation

$$z\mathscr{F}_{L+1} - (L+\nu-1+z)\mathscr{F}_L + (L+\nu-1+D_4)\mathscr{F}_{L-1} = 0. \quad (150)$$

The $M_L(z)$ form the solution sequence to the recurrence relation (150) that is minimal as $L \to +\infty$, and the $U_L(z)$ form a solution that is dominant. The Wronskian of $M_L(z)$ and $U_L(z)$ is

$$M_L(z)U_{L,z}(z) - U_L(z)M_{L,z}(z) = [\Gamma(L+\nu+D_4)/\Gamma(-D_4)]z^{-L-\nu}e^z. \quad (151)$$

The solutions $f(z)$ to Eq. (143) may now be expressed as

$$f(z) = \sum_{L=-\infty}^{\infty} a_L\mathscr{F}_L(z), \quad (152)$$

where ν must be chosen such that the coefficients a_L form a solution sequence minimal as $L \to \pm\infty$ of the recurrence relation

$$\alpha_L a_{L+1} + \beta_L a_L + \gamma_L a_{L-1} = 0, \quad (153)$$

where

$$\alpha_L = -(L+\nu+D_4)(L+\nu+1-D_2+z_0),$$
$$\beta_L = (L+\nu)(L+\nu-D_2+2z_0-1) + (D_4-1)z_0 + (D_1+D_2+D_3), \quad (154)$$
$$\gamma_L = -(L+\nu-1)z_0 - D_1,$$

or, in the current case where the D_i are given by Eqs. (144),

$$\alpha_L = -(L+\nu+1-B_2)(L+\nu-\tfrac{1}{2}B_2-i\eta),$$
$$\beta_L = (L+\nu)(L+\nu-1-B_2-2i\omega x_0) + i\omega x_0(B_2 - B_1/x_0) + B_2 + B_3,$$
$$\gamma_L = +2i\omega x_0(L+\nu-1+B_1/x_0).$$

The parameter ν must solve the implicit equation

$$\beta_0 = \frac{\alpha_{-1}\gamma_0}{\beta_{-1}-}\,\frac{\alpha_{-2}\gamma_{-1}}{\beta_{-2}-}\,\frac{\alpha_{-3}\gamma_{-2}}{\beta_{-3}-}\cdots + \frac{\alpha_0\gamma_1}{\beta_1-}\,\frac{\alpha_1\gamma_2}{\beta_2-}\,\frac{\alpha_2\gamma_3}{\beta_3-}\cdots. \quad (155)$$

The solutions ν to Eq. (155) are probably periodic. Experience with the Coulomb wave-function expansion suggests that the correct values of ν will be those that are contiguous as $\omega\to 0$ with the $\omega=0$ roots of (155), that is, the values ν_0 that make $\beta_0=0$ when $\omega=0$:

$$\nu_0=\tfrac{1}{2}[B_2+1\pm\sqrt{B_2(B_2-2)-4B_3+1}]. \quad (156)$$

B. Convergence properties

If the parameter ν is chosen to satisfy Eq. (155), then the a_L of Eq. (153) will be minimal as $L\to\infty$ and successive a_L will have the limiting ratios

$$\lim_{L\to+\infty}\frac{a_L}{a_{L-1}} = \frac{-2i\omega x_0}{L}\left[1+\frac{1}{L}\left(B_2+\frac{B_1}{x_0}-\nu\right)+O(L^{-2})\right], \quad (157)$$

and

$$\lim_{L\to-\infty}\frac{a_L}{a_{L+1}}=1+\frac{1}{L}\left(2-\frac{1}{2}B_2-i\eta\right)+O(L^{-2}). \quad (158)$$

Assume that both the M_L and the U_L are dominant solutions to recurrence relation (150) as $L\to-\infty$, and denote them again by $\mathscr{F}_L$. Successive $\mathscr{F}_L$ will have the limiting ratios

$$\lim_{L\to-\infty}\frac{\mathscr{F}_L}{\mathscr{F}_{L+1}}=1+\frac{1}{L}\left(\frac{1}{2}B_2+i\eta\right)+O(L^{-2}). \quad (159)$$

As $L\to+\infty$ the M_L are minimal,

$$\lim_{L\to+\infty}\frac{M_L}{M_{L-1}}=1-\frac{1}{L}\left(\frac{1}{2}B_2+i\eta\right)+O(L^{-2}), \quad (160)$$

while the U_L are dominant:

$$\lim_{L\to+\infty}\frac{U_L}{U_{L-1}}=\frac{L}{-2i\omega x}\left[1+\frac{1}{L}(2\nu-2)+O(L^{-2})\right]. \quad (161)$$

From Eqs. (158) and (159) we see that

$$\lim_{L\to-\infty}\frac{a_L\mathscr{F}_L}{a_{L+1}\mathscr{F}_{L+1}}=1+\frac{2}{L}+O(L^{-2}), \quad (162)$$

so that the negative L part of series (152) is absolutely (albeit slowly) convergent for $\mathscr{F}_L$ either M_L or U_L. From (157) and (160),

$$\lim_{L\to+\infty}\frac{a_LM_L}{a_{L-1}M_{L-1}} = \frac{-2i\omega x_0}{L}\left[1+\frac{1}{L}\left(\frac{1}{2}B_2+\frac{B_1}{x_0}-i\eta\right)+O(L^{-2})\right], \quad (163)$$

and from (157) and (161)

$$\lim_{L\to+\infty}\frac{a_LU_L}{a_{L-1}U_{L-1}} = \frac{x_0}{x}\left[1+\frac{1}{L}\left(B_2+\frac{B_1}{x_0}-\nu-2i\omega x(\nu-2)\right) + O(L^{-2})\right]. \quad (164)$$

Therefore Σa_LU_L converges for all x such that $|x|>|x_0|$, and Σa_LM_L converges, amusingly enough, for all x.

The preceding arguments may be repeated using in Eq. (142) the alternate substitutions

$$y(x)=e^{-i\omega x}f(x),\quad z=+2i\omega x,\quad\text{and}\quad z_0=+2i\omega x_0, \quad (165)$$

which yield two final representations for the generalized spheroidal wave functions:

$$y_2(x)=e^{-i\omega x}\sum_{L=-\infty}^{\infty}b_L\widetilde{M}\left(\frac{1}{2}B_2-i\eta,L+\nu,+2i\omega x\right), \quad (166)$$

and

$$y_-(x)=e^{-i\omega x}\sum_{L=-\infty}^{\infty}b_LU\left(\frac{1}{2}B_2-i\eta,L+\nu,+2i\omega x\right), \quad (167)$$

where the expansion coefficients b_L are again a solution sequence minimal as $L\to\pm\infty$ of a three-term recurrence relation

$$\alpha_Lb_{L+1}+\beta_Lb_L+\gamma_Lb_{L-1}=0, \quad (168)$$

with recurrence coefficients α_L, β_L, and γ_L given by

$$\alpha_L=-(L+\nu+1-B_2)\left(L+\nu-\frac{1}{2}B_2+i\eta\right),$$
$$\beta_L=(L+\nu)(L+\nu-1-B_2+2i\omega x_0) - i\omega x_0(B_2-B_1/x_0)+B_2+B_3,$$
$$\gamma_L=-2i\omega x_0(L+\nu-1+B_1/x_0),$$

which together with our first two confluent hypergeometric function solutions

$$y_1(x)=e^{+i\omega x}\sum_{L=-\infty}^{\infty}a_L\widetilde{M}\left(\frac{1}{2}B_2+i\eta,L+\nu,-2i\omega x\right), \quad (169)$$

and

$$y_+(x)=e^{+i\omega x}\sum_{L=-\infty}^{\infty}a_LU\left(\frac{1}{2}B_2+i\eta,L+\nu,-2i\omega x\right), \quad (170)$$

constitute the last of our new representations for the generalized spheroidal wave functions. Note that the ν that solve equation (155) for the expansion coefficients b_L using the α, β, and γ of Eq. (168) will probably not be the same ν that solve the equation for the a_L using the α, β, and γ of Eq. (153). Computer programs have not been written to generate any of these confluent hypergeometric function series, but Eq. (162) suggests that

$$\lim_{L\to-\infty}a_L\mathscr{F}_L\approx O(L^{-2}),$$

so that roughly 10^N terms will be needed if the series (166), (167), (169), or (170) are to be summed to N figures of accuracy. While one could hope that this is only a worst-case estimate and, at least for the series of the regular functions $\widetilde{M}(\frac{1}{2}B_2\pm i\eta,L+\nu,\mp 2i\omega x)$ and $|x|\approx 1$, that the series can in practice be made to converge much faster (perhaps with the help of a sequence accelerating algorithm), such speculation must be regarded as "wishful thinking" pending more detailed analysis.

C. Asymptotic behavior

The limiting forms of the confluent hypergeometric functions for large values of the argument are

$$\lim_{|z|\to\infty} M_L(z) = e^{\mp i\pi D_4} z^{D_4} + [\Gamma(L+\nu+D_4)/\Gamma(-D_4)] \times e^z z^{-L-\nu-D_4}[1+O(z^{-1})], \quad (171)$$

and

$$\lim_{|z|\to\infty} U_L(z) = z^{D_4}. \quad (172)$$

The upper sign is taken in (171) if $-\pi/2 < \arg(z) < 3\pi/2$, and the lower sign is taken if $-3\pi/2 < \arg(z) \leqslant -\pi/2$. The factor $\Gamma(L+\nu+D_4)z^{-L-\nu-D_4}$ makes the large $|z|$ limit of the negative L part of series (166) and (169) difficult to evaluate. However, the series (167) and (170) involving the irregular functions $U_L(z)$ are relatively simple, and we may express the limiting forms of these solutions as

$$\lim_{|x|\to\infty} y_+(x) = \lim_{|x|\to\infty} e^{+i\omega x} \sum_{L=-\infty}^{\infty} a_L \times U\left(\frac{1}{2}B_2 + i\eta, L+\nu, -2i\omega x\right) = (-2i\omega x)^{-B_2/2 - i\eta} e^{+i\omega x} \sum_{L=-\infty}^{\infty} a_L, \quad (173)$$

and

$$\lim_{|x|\to\infty} y_-(x) = \lim_{|x|\to\infty} e^{-i\omega x} \sum_{L=-\infty}^{\infty} b_L \times U\left(\frac{1}{2}B_2 - i\eta, L+\nu, +2i\omega x\right) = (+2i\omega x)^{-B_2/2 + i\eta} e^{-i\omega x} \sum_{L=-\infty}^{\infty} b_L. \quad (174)$$

As noted before, these sums are only slowly convergent, and expansions (167) and (170) probably have no computational advantage over the Jaffé-type solutions discussed in Sec. IV. However, the convergence, no matter how slow, of the regular solution expansions (166) and (167) over the entire interval $[x_0 \leqslant x < \infty)$ may give them some unique analytic utility.

VIII. A CONFLUENT GENERALIZED SPHEROIDAL WAVE EQUATION

A. The confluent equation

When $x_0 = 0$, Eq. (1) becomes

$$x^2 y_{,xx} + (B_1 + B_2 x) y_{,x} + (\omega^2 x^2 - 2\eta\omega x + B_3) y = 0. \quad (175)$$

If $y(x) = x^{-B_2/2} h$ and $z = \omega x$, then the differential equation for $h(z)$ is

$$z^2 h_{,zz} + C_1 \omega h_{,z} + [z^2 - 2\eta z + C_2 + C_3\omega/z] h = 0, \quad (176)$$

where

$$C_1 = B_1, \quad C_2 = B_3 - \tfrac{1}{2}B_2(\tfrac{1}{2}B_2 - 1),$$

and

$$C_3 = -\tfrac{1}{2}B_1 B_2.$$

In Sec. VI we showed that h could be expanded as

$$h(x) = \sum_{L=-\infty}^{\infty} a_L u_{L+\nu}(\eta, \omega x) \quad (177)$$

(where $u_{L+\nu}$ is a Coulomb wave function), and that the expansion converges even in the present case when $x_0 = 0$. This property is intriguing, because as $x_0 \to 0$ the point $x = 0$ becomes a confluent singular point, and convergent expansions of solutions to differential equations near such points are generally difficult to obtain. Note that the point $x = 0$ is an irregular singular point only when $B_1 \neq 0$: when $B_1 = 0$, Eq. (175) is a simple confluent hypergeometric equation, Eq. (176) is the Coulomb wave equation, and the solutions $y(x)$ can be expressed as

$$y(x) = x^{-(1/2)B_2} u_\nu(\eta, \omega x),$$

where

$$\nu(\nu+1) = -C_2 = \tfrac{1}{2}B_2(\tfrac{1}{2}B_2 - 1) - B_3. \quad (178)$$

The solution regular at $x = 0$ is $u_\nu = F_\nu(\eta, \omega x)$, and an irregular solution is given by $u_\nu = G_\nu(\eta, \omega x)$.

However, when $B_1 \neq 0$ the point $x = 0$ is an irregular singular point and expansion (177) must be used for the two solutions, neither of which will converge at $x = 0$. When $x_0 \neq 0$ the solutions near $x = 0$ could be generated via the Jaffé expansion $\Sigma a_n [(x - x_0)/x]^n$. When $x_0 = 0$, the Jaffé expansion does not exist and another approach must be taken towards generating solutions good near that point. This can be done by exploiting the symmetry that exists between the point $x = 0$ and the point at ∞: both are confluent singular points and with the substitutions

$$y(x) = e^{i\omega x + B_1/2x} x^{1 - B_2/2} f(\xi), \quad \xi = iB_1/2x,$$

Eq. (175) becomes

$$\xi^2 f_{,\xi\xi} + \tilde{C}_1 \omega F_{,\xi} + [\xi^2 - 2\tilde{\eta}\xi + \tilde{C}_2 + \tilde{C}_3\omega/\xi] f = 0, \quad (179)$$

where

$$\tilde{C}_1 = C_1 = B_1, \quad \tilde{\eta} = -i(\tfrac{1}{2}B_2 - 1),$$

$$\tilde{C}_2 = C_2 = B_3 - \tfrac{1}{2}B_2(\tfrac{1}{2}B_2 - 1), \quad \xi = iB_1/2x,$$

$$\tilde{C}_3 = -(1 + i\eta)B_1.$$

Hence solutions to Eq. (175) can also be written

$$y(x) = x^{1 - B_2/2} e^{i\omega x + B_1/2x} \sum_{L=-\infty}^{\infty} b_L u_{L+\nu}(\tilde{\eta}, \xi). \quad (180)$$

Expansion (180) is uniformly convergent as $x \to 0$. The expansion coefficients a_L in expression (177) and the coefficients b_L in expression (180) are both defined by Eqs. (112) and (113) using, respectively, the C_i of (176) and the $\tilde{C}_i$ of (179).

B. The Kerr limit of black hole rotation

An example of the confluent equation occurs at the Kerr limit of black hole rotation, where $b = 0$ and $a = \tfrac{1}{2}$ in Eqs. (15) and (19). If it were physically possible, the confluence of the event horizon at $x = x_0$ with the singularity at $x = 0$ would result in a naked singularity. The current theory of gravitation does not allow naked singularities to form,[42] but an understanding of the behavior of the solutions to the wave equation (15) at the Kerr limit might allow some insight regarding the behavior of solutions near that limit. At the Kerr limit, Eq. (19) becomes

$$x^2 y_{,xx} + [2(1-s-i\omega)x - i(\omega - m)]y_{,x} + [\omega^2 x^2 + 2(\omega + is)\omega x + \tfrac{7}{4}\omega^2 + (2s-1)i\omega - 2s - A_{lm}]y = 0, \quad (181)$$

which is of the form (175). The substitutions $y = x^{s+i\omega-1} \times h(z)$ and $z = \omega x$ yield

$$z^2 h_{,zz} - i\omega(\omega - m)h_{,z} + [z^2 - 2\eta z + C_2 + i\omega(\omega - m)(1 - s - i\omega)/z]h = 0, \quad (182)$$

where $C_2 = \frac{7}{4}\omega^2 - s(s+1) - A_{lm}$ and $\eta = -\omega - is$, which corresponds to Eq. (176). Similarly, the substitution $y = x^{s+i\omega} \exp i[\omega x - (\omega - m)/2x]f(\xi)$ with $\xi = (\omega - m)/2x$ gives

$$\xi^2 f_{,\xi\xi} - i\omega(\omega - m)f_{,\xi} + [\xi^2 - 2\tilde{\eta}\xi + C_2 + i\omega(\omega - m)(1 + s - i\omega)/\xi]f = 0, \quad (183)$$

where $\tilde{\eta} = -\omega + is$. Two interesting limiting cases of Eqs. (182) and (183) occur when $\omega \approx 0$ and when $\omega \approx m$. The occurrence of the product $i\omega(\omega - m)$ allows us to treat both cases in the same manner.

Let $C_1 = -i(\omega - m)$, $C_3 = i(\omega - m)(1 - s - i\omega)$, $\widetilde{C}_3 = i(\omega - m)(1 + s - i\omega)$, and $\tilde{\eta} = -\omega + is$. Then expansions corresponding to (177) and (180) that are, respectively, convergent for x and $1/x$ bounded away from 0 are

$$y^{(\infty)}(x) = x^{s+i\omega-1} \sum_{L=-\infty}^{\infty} a_L u_{L+\nu}(-\omega - is, \omega x), \quad (184)$$

$$y^{(0)}(x) = x^{s+i\omega} e^{i\omega x - i(\omega - m)/2x} \times \sum_{L=-\infty}^{\infty} b_L u_{L+\nu}\left(-\omega + is, \frac{(\omega - m)}{2x}\right). \quad (185)$$

Again, $u_{L+\nu}(\eta, z)$ denotes a Coulomb wave function. The ν in Eq. (184) is chosen to satisfy (113) and makes the a_L the minimal solution of

$$\alpha_L a_{L+1} + \beta_L a_L + \gamma_L a_{L-1} = 0, \quad (186)$$

where

$$\alpha_L = -i\omega(\omega - m)R_{L+1}(\eta)(L + \nu + 1 + s + i\omega)/(2L + 2\nu + 3),$$

$$\beta_L = (L+\nu)(L+\nu+1) + C_2 + i\omega(\omega - m) \times (s + i\omega)Q_L(\eta), \quad (187)$$

$$\gamma_L = +i\omega(\omega - m)R_L(\eta)(L + \nu - s - i\omega)/(2L + 2\nu - 1).$$

The ν and b_L of Eq. (185) are generated from

$$\tilde{\alpha}_L b_{L+1} + \tilde{\beta}_L b_L + \tilde{\gamma}_L b_{L-1} = 0, \quad (188)$$

where the $\tilde{\alpha}_L, \tilde{\beta}_L$, and $\tilde{\gamma}_L$ are the same as the α, β, and γ of Eq. (187) but with η replaced by $\tilde{\eta}$ and C_3 replaced by $\widetilde{C}_3$ (i.e., s is replaced by $-s$). As $\omega \to 0$ or as $\omega \to m$, only the a_0 and b_0 terms contribute to the sums in (184) and (185), so that limiting forms for $y^{(\infty)}$ and $y^{(0)}$ are (with $a_0 = b_0 = 1$)

$$\lim_{\omega \to 0,m} y^{(\infty)}(x) \sim x^{s+i\omega-1} u_\nu(-\omega - is, \omega x), \quad (189)$$

$$\lim_{\omega \to 0,m} y^{(0)}(x) \sim x^{s+i\omega} e^{i\omega x - i(\omega - m)/2x} u_\nu \times (-\omega + is, (\omega - m)/2x), \quad (190)$$

where $\nu(\nu + 1) = -C_2(\omega = 0, m)$. Expressions (189) and (190) are valid for values of the rotation parameter such that $0 < b \ll 1$, not just $b = 0$. This generalization may be demonstrated by retaining b and interchanging ω and $\omega - m$ when deriving the definitions (187) for the recurrence coefficients α_L, β_L, and γ_L.

The physically relevant field function is Teukolsky's $R_{lm}(r)$, which is related to $y(x)$ by equation (17):

$$R(r) = (r - r_-)^{k_-}(r - r_+)^{k_+} y(x),$$

where $x = r - r_-$. Again,

$$r_\pm = (1 \pm b)/2, \quad k_- = -s + i(\omega r_- - am)/b,$$
$$k_+ = -s - i(\omega r_+ - am)/b, \quad \text{and} \quad a = \tfrac{1}{2}(1 - b^2)^{1/2}.$$

Taking the limit

$$\lim_{b \to 0} (r - r_-)^{k_-}(r - r+)^{k_+} = x^{-2s - i\omega} e^{i(\omega - m)/2x},$$

we find the expansions for $R(r)$ that are, respectively, convergent away from $r = \frac{1}{2}$ and away from $r = \infty$ are, with $x = r - \frac{1}{2}$,

$$R_\pm^{(\infty)}(x) = x^{-s-1} e^{i(\omega - m)/2x} \times \sum_{L=-\infty}^{\infty} a_L [G_{L+\nu}(-\omega - is, \omega x) \pm iF_{L+\nu}(-\omega - is, \omega x)], \quad (191)$$

$$R_\pm^{(0)}(x) = x^{-s} e^{i\omega x} \times \sum_{L=-\infty}^{\infty} b_L \left[G_{L+\nu}\left(-\omega + is, \frac{(\omega - m)}{2x}\right) \pm iF_{L+\nu}\left(-\omega + is, \frac{(\omega - m)}{2x}\right)\right]. \quad (192)$$

Equations (109) give the behavior of the Coulomb wave functions for large magnitudes of the argument, and from them we obtain the desired behavior of the two solutions near $x = 0$:

$$\lim_{x \to 0} R_+^{(0)}(x) \sim [(\omega - m)/2x]^{i\omega} e^{i(\omega - m)/2x}, \quad (193)$$

$$\lim_{x \to 0} R_-^{(0)}(x) \sim x^{-2s}[(\omega - m)/2x]^{-i\omega} e^{-i(\omega - m)/2x}. \quad (194)$$

With the sign convention $(e^{-i\omega t})$ of Eq. (13), it is $R_+^{(0)}(x)$ that describes the case of radiation going into the singularity.

A black hole rotates at the Kerr limit with angular velocity $d\phi/dt = 1$ in the normalized units used here, and a wave train with frequency $\omega = m$ at this limit corotates with the singularity. If $|\omega| \ll 1$ and $|\omega x| \ll 1$, expansion (192) is dominated by the b_0 term and may be approximated (with $b_0 = 1$) by

$$R_{\pm}^{(0)}(x) \approx x^{-s} e^{i\omega x}[G_\nu(-\omega + is, (\omega - m)/2x) \pm iF_\nu(-\omega + is, (\omega - m)/2x)], \quad (195)$$

where $\nu(\nu+1) = -C_2(\omega)$. This result also holds for $|\omega - m| \ll 1$ and $|\omega x| \ll \frac{1}{2}$. Similarly, if $|x/(\omega - m)| \gg 1$, then by Eq. (116) the a_0 term will dominate expansion (191) and $R_{\pm}^{(\infty)}$ may be approximated (with $a_0 = 1$) by

$$R_{\pm}^{(\infty)}(x) \approx x^{-s-1} e^{i(\omega - m)/2x} \times [G_\nu(-\omega - is, \omega x) \pm iF_\nu(-\omega - is, \omega x)]. \quad (196)$$

These approximations may be used whenever the rotation parameter $b \ll 1$. A different approach to approximating the Teukolsky function R_{lm} near the Kerr limit may be found in Teukolsky and Press,[43] and has been used by Detweiler.[44]

IX. SUMMARY

A. Review of the representations

In this study we have demonstrated ten analytic series representations for the solutions to the generalized spheroidal wave equation

$$x(x - x_0)\frac{d^2y}{dx^2} + (B_1 + B_2 x)\frac{dy}{dx} + [\omega^2 x(x - x_0) - 2\eta\omega(x - x_0) + B_3]y = 0$$

on the interval $[0 < x < \infty)$. They are, together with the asymptotic form, the following.

(i) The regular power series solutions of the Jaffé type (Sec. IV A):

$$y_1(x) = e^{+i\omega x} x^{-(1/2)B_2 - i\eta} \sum_{n=0}^{\infty} a_n^r \left(\frac{x - x_0}{x}\right)^n, \quad (39)$$

and

$$y_2(x) = e^{-i\omega x} x^{-(1/2)B_2 + i\eta} \sum_{n=0}^{\infty} b_n^r \left(\frac{x - x_0}{x}\right)^n. \quad (49)$$

These two expansions are proportional by a factor $e^{2i\omega x_0} a_0^r / b_0^r$ and represent the generalized spheroidal wave function that is regular at $x = x_0$. They converge for all x such that $|(x - x_0)/x| < 1$. The convergence is uniform only when ω is an eigenfrequency and the expansion coefficients form minimal solutions to their respective recurrence relations (40) and (50). When ω is not an eigenfrequency the convergence of these series is not uniform and the analytic forms of $y_1(x)$ and $y_2(x)$ as $x \to \infty$ cannot be deduced.

(ii) The irregular confluent hypergeometric function solutions of Hylleraas type (Sec. IV C):

$$y_+(x) = e^{+i\omega x} \sum_{n=0}^{\infty} a_n^r (B_2 + B_1/x_0)_n \times U(\tfrac{1}{2}B_2 + i\eta + n, -B_1/x_0, -2i\omega x), \quad (73)$$

$$y_-(x) = e^{-i\omega x} \sum_{n=0}^{\infty} b_n^r (B_2 + B_1/x_0)_n \times U(\tfrac{1}{2}B_2 - i\eta + n, -B_1/x_0, +2i\omega x). \quad (74)$$

These solutions are always independent and correspond, respectively, to the asymptotic forms

$$a_0^r x^{-(1/2)B_2} \exp[+i(\omega x - \eta \ln x)]$$

and

$$b_0^r x^{-(1/2)B_2} \exp[-i(\omega x - \eta \ln x)].$$

In general the expansions for y_3 and y_4 do not converge as $x \to x_0$, and these solutions are usually irregular at that point. The exception occurs when ω is an eigenfrequency and either the a_n^r or the b_n^r (but never both) are minimal as $n \to \infty$. In this case these irregular Hylleraas solutions become regular eigensolutions proportional to the regular Jaffé solutions y_1 and y_2 [Eqs. (39) and (49)].

(iii) The irregular Coulomb wave-function solutions of the generalized Stratton type (Sec. VI):

$$y_+(x) = x^{-B_2/2} \sum_{L=-\infty}^{\infty} a_L [G_{L+\nu}(\eta, z) + iF_{L+\nu}(\eta, z)],$$

$$y_-(x) = x^{-B_2/2} \sum_{L=-\infty}^{\infty} a_L [G_{L+\nu}(\eta, z) - iF_{L+\nu}(\eta, z)]. \quad (22)$$

The asymptotic forms of these expansions are given by Eqs. (123):

$$\lim_{x \to \infty} y_{\pm}(x) = x^{-B_2/2} \exp[\pm i(\omega x - \eta \ln(2\omega x) - \phi_{\pm})], \quad (123)$$

where the $\phi_{\pm}$ and the necessary σ_L are given by Eqs. (124) and (110). The expansion coefficients a_L and the phase factor ν are defined by Eqs. (112), (113), and (121).

(iv) The asymptotic solutions in terms of Coulomb wave functions (Sec. II):

$$\lim_{x \to \infty} y_{\pm}(x) = x^{B_1/2x_0}(x - x_0)^{-(1/2)(B_2 + B_1/x_0)} \times [G_{\nu_a}(\eta, \omega x) \pm iF_{\nu_a}(\eta, \omega x)] \times [1 + O(x^{-3})], \quad (197)$$

cf. Eqs. (22) and (23). The asymptotic phase parameter ν_a usually differs markedly from the Coulomb wave function phase parameter ν, and the asymptotic form can provide a check on the full expansion in the regions of large x for which both are valid.

(v) The confluent hypergeometric function expansions of Sec. VII:

$$y_1(x) = e^{+i\omega x} \sum_{L=-\infty}^{\infty} a_L \widetilde{M}\left(\frac{1}{2}B_2 + i\eta, L + \nu, -2i\omega x\right),$$

$$y_2(x) = e^{-i\omega x} \sum_{L=-\infty}^{\infty} b_L \widetilde{M}\left(\frac{1}{2}B_2 - i\eta, L + \nu, +2i\omega x\right),$$

$$y_+(x) = e^{+i\omega x} \sum_{L=-\infty}^{\infty} a_L U\left(\frac{1}{2}B_2 + i\eta, L + \nu, -2i\omega x\right),$$

$$y_-(x) = e^{-i\omega x} \sum_{L=-\infty}^{\infty} b_L U\left(\frac{1}{2}B_2 - i\eta, L + \nu, +2i\omega x\right),$$

cf. equations (166), (167), (169), and (170). The description of these solutions given in Sec. VII is brief enough (as befits the preliminary nature of their derivation) to make further discussion unnecessary.

B. Notes on the computer implementation

The expressions for which FORTRAN subroutines have been written to evaluate are those for the regular Jaffé solution

$$y_1(x) = e^{+i\omega x} x^{-(1/2)B_2 - i\eta} \sum_{n=0}^{\infty} a_n^r \left(\frac{x-x_0}{x}\right)^n, \qquad (198)$$

the Coulomb wave function expansions (122)

$$y_\pm(x) = x^{-B_2/2} \sum_{L=-\infty}^{\infty} a_L [G_{L+\nu}(\eta,z) \pm iF_{L+\nu}(\eta,z)], \qquad (199)$$

and the associated asymptotic forms (22) and (23)

$$\lim_{x\to\infty} y_\pm(x) = x^{B_1/2x_0}(x-x_0)^{-(1/2)(B_2+B_1/x_0)} \times [G_{\nu_a}(\eta,\omega x) \pm iF_{\nu_a}(\eta,\omega x)] \times [1 + O(x^{-3})]. \qquad (200)$$

The Jaffé solutions are regular and analytic as $x\to x_0$, but for general ω are divergent as $x\to\infty$. The Coulomb wave function expansions are analytic as $x\to\infty$, but diverge as $x\to x_0$. The combination of the two representations provides a powerful computational tool for analysis of physical systems described by generalized spheroidal wave equations. The parameter regions in which the Coulomb wave-function expansion is valid often overlap the regions of validity of the Jaffé expansions, and frequently those of the asymptotic Coulomb wave-function solutions as well, so the three different methods of solution can be used as checks against each other.

The regular Jaffé solution is a simple power series and coding it was straightforward. However, the Coulomb wave-function expansions are irregular as $x\to x_0$, and usually have a branch cut associated with that point. Additional branch cuts arise in the continued fractions that define the expansion coefficients a_L and the phase parameter ν. Some of these cuts no doubt are inherent to the fractions themselves, but others are spurious and are due to the square roots that occur in the recurrence relations both for the a_L and for the Coulomb wave functions. We showed in Sec. VI F how the square roots could be avoided by use of Gautschi's normalizations, but the current (July 1985) version of the program implements Eq. (122) with the usually defined Coulomb wave functions [Eq. (111)]. The branch cuts are a genuine problem since the generalized spheroidal wave functions are functions of seven complex parameters: $x,x_0,\eta,\omega,B_1,B_2$ and B_3. The Coulomb wave functions are computed using an analytic extension of Steed's algorithm,[45] and branch cuts in this subroutine alone restrict the product ωx to lie in the fourth quadrant of the complex ωx plane.

It probably is not possible with algorithms of this complexity to fully predict the ranges of the parameters for which they are valid, but fortunately there are enough self-consistency checks (computation of Wronskians, independent sums of series for the derivatives, etc.) that the accuracy of the program that implements the Coulomb wave-function expansion can be determined internally as it is run. Although one usually cannot predict *a priori* whether the program will run with a given set of arguments, the relative error of the calculation is accurately supplied at execution. External checks on the program's results, while comforting when they are obtainable, are not strictly necessary for reliable use.

The single precision version of the program typically returns five or six decimal places of accuracy on a 36-bit DEC20. (A double precision version using the COMPLEX*16 variable type available on VAX computers gives between twelve and sixteen places.) While these algorithms have been found to be quite powerful in the analysis of the perturbation response of Schwarzschild black holes,[4] the programs are not a complete panacea to the problem of generating generalized spheroidal wave functions, as there are values of the parameters for which it is not possible to find a value of the phase parameter ν that satisfies Eq. (113). However, it is our belief that the elements of analyticity inherent to Jaffé's solutions and the Coulomb wave-function expansion provide, in the regions where they are valid, a refreshing alternative to the usual recourse of brute force numerical integration of the generalized spheroidal wave equation.

The programs are fully portable and will be described in detail in a forthcoming article.[46] Future study should be in the direction of implementing Gautschi's normalization for the Coulomb wave functions (Sec. VI F), and the irregular Hylleraas solutions (Sec. IV C). The high frequency approximation of Jen and Hu,[40] for the ordinary spheroidal wave functions should be extended to the generalized functions. Further investigation should be made of the confluent hypergeometric function expansions discussed in Sec. VII.

ACKNOWLEDGMENTS

This investigation of generalized spheroidal wave functions was undertaken at the suggestion of J. P. Simons. The author is grateful for useful and encouraging discussions with J. Ball, N. H. F. Beebe, R. Borscherdt, W. C. Lineburger, D. Norcross, R. Pack, R. H. Price, A. Ramsay, W. Reinhardt, and P. K. P. Spudich. This investigation is at its heart a computational study, and the analytic results described herein were obtained only after extensive numeric experimentation. This work would not have been possible without the competent assistance of the operations staff of the University of Utah College of Science Computer, and it is with the deepest satisfaction that I acknowledge the contributions, both professional and personal, of Pieter Bowman, Arlan Duncan, Alan Lichty, Lon Willet, Douglas Hendry, and Nelson H. F. Beebe.

A preliminary study for this work was partially supported by National Science Foundation Grant No. CHE-75-19476.

APPENDIX A: LAGUERRE POLYNOMIALS

The Laguerre polynomials used in this paper are the Laguerre polynomials as defined by Slater,[27] and in Gradshteyn and Ryzhik.[29] They are generated by

$$\sum_{n=0}^{\infty} z^n L_n^\alpha(x) = (1-z)^{-\alpha-1} e^{xz/(z-1)} \quad (|z|<1),$$

obey the recurrence relation

$$(n+1)L^{\alpha}_{n+1}(z) - (2n+\alpha+1-z)L^{\alpha}_{n}(z) + (n+\alpha) \times L^{\alpha}_{n-1}(z) = 0,$$

$$xL^{\alpha}_{0}(x) = -L^{\alpha}_{1}(x) + (\alpha+1)L^{\alpha}_{0}(x),$$

and satisfy the differential property

$$x\frac{d}{dx}L^{\alpha}_{n}(x) = -(n+1)L^{\alpha}_{n+1}(x) + (2n+\alpha+1) \times L^{\alpha}_{n}(x) - (n+\alpha)L^{\alpha}_{n-1}(x)$$

$$\frac{d}{dx}L^{\alpha}_{0}(x) = 0.$$

Laguerre polynomials are solutions to the confluent hypergeometric equation

$$x\frac{d^2}{dx^2}L^{\alpha}_{n}(x) + (\alpha+1-x)\frac{d}{dx}L^{\alpha}_{n}(x) + nL^{\alpha}_{n}(x) = 0,$$

and $L^{\alpha}_{n}(x)$ is related to Kummer's function ${}_1F_1(a,b,x)$ by

$$L^{\alpha}_{n}(x) = \frac{\Gamma(\alpha+n+1)}{n!\Gamma(\alpha+1)}\,{}_1F_1(-n,\alpha+1,x).$$

There are several different normalizations of Laguerre polynomials currently in use. Three of them are listed below together with the names of the authors that have used them, and their relation to the Laguerre polynomials as normalized by Slater.

(1) Messiah[47] and Morse and Feshbach[10]:

L^{m}_{n} (Messiah, Morse, and Feshbach) $= (n+m)!L^{m}_{n}$ (Slater).

(2) Titchmarsh[48]:

L^{m}_{n} (Titchmarsh) $= n!L^{m}_{n}$ (Slater).

(3) Hylleraas[7]:

L^{m}_{n} (Hylleraas) $= L^{m}_{n-m}$ (Slater).

APPENDIX B: AN INTEGRAL TRANSFORM

An outline of a proof of the validity of Eq. (62) is given using the standard theory of integral transforms.[49]

Start with Eq. (1):

$$x(x-x_0)\frac{d^2y}{dx^2} + (B_1+B_2x)\frac{dy}{dx} + [\omega^2x(x-x_0) - 2\eta\omega(x-x_0) + B_3]y = 0.$$

With the substitution $y(x) = e^{i\omega x}f(x)$ and the restriction $i\eta = B_2/2 - 1$, the differential equation for f is

$$\mathscr{L}_x\{f(x)\} = 0, \qquad \text{(B1)}$$

where the differential operator $\mathscr{L}_x$ is defined by

$$\mathscr{L}_x \equiv x(x-x_0)\frac{d^2}{dx^2} + [B_1 + B_2x + 2i\omega x(x-x_0)]\frac{d}{dx} + [2i\omega x(B_2-1)(x-x_0) + i\omega x_0(B_2 + B_1/x_0) + B_3].$$

With the further substitution $f(x) = x^{1+B_1/x_0}g(x)$, the differential equation for $g(x)$ is

$$\mathscr{M}_x\{g(x)\} = 0,$$

where the differential operator $\mathscr{M}_x$ is in turn defined by

$$\mathscr{M}_x \equiv x(x-x_0)\frac{d^2}{dx^2} + [-B_1 - 2x_0 + (2 + B_2 + 2B_1/x_0)x + 2i\omega x(x-x_0)]\frac{d}{dx} + [(B_2 + B_1/x_0)(2i\omega x - i\omega x_0 + 1 + B_1/x_0) + B_3],$$

so that $\mathscr{L}_x\{f\} = x^{1+B_1/x_0}\mathscr{M}_x\{g\}$. The adjoint operator to $\mathscr{L}_x$ is

$$\bar{\mathscr{L}}_x = x(x-x_0)\frac{d^2}{dx^2} + [-B_1 - 2x_0 + (4-B_2)x - 2i\omega x(x-x_0)]\frac{d}{dx} + [2i\omega(B_2-3)x - i\omega x_0(B_2 - B_1/x_0 - 4) + 2 - B_2 + B_3].$$

The kernel

$$K(x,t) \equiv e^{2i\omega x(t-x_0)/x_0}(t-x_0)^{B_2 + B_1/x_0 - 1}$$

has the property that $\mathscr{M}_x\{K(x,t)\} = \bar{\mathscr{L}}_t\{K(x,t)\}$. Hence if we write

$$\bar{f}(x) = x^{1+B_1/x_0}\int_c K(x,t)f(t)dt, \qquad \text{(B2)}$$

we can operate with $\mathscr{L}_x$ on $\bar{f}$ and find successively

$$\mathscr{L}_x\{\bar{f}(x)\} = \mathscr{L}_x\left\{x^{1+B_1/x_0}\int_c K(x,t)f(t)dt\right\} = x^{1+B_1/x_0}\int_c(\mathscr{M}_x\{K(x,t)\})f(t)dt$$
$$= x^{1+B_1/x_0}\int_c(\bar{\mathscr{L}}_t\{K(x,t)\})f(t)dt = x^{1+B_1/x_0}\left[\int_c K(x,t)\mathscr{L}_t\{f(t)\}dt + \int_c\frac{d}{dt}P(x,t)dt\right],$$

where the bilinear concomitant $P(x,t)$ is given by

$$P(x,t) = t(t-x_0)\left[f(t)\frac{d}{dt}K(x,t) - K(x,t)\frac{d}{dt}f(t)\right] + \left[2i\omega t^2 + \left(B_2 + \frac{2B_1}{x_0} - 2i\omega x_0\right)t - B_1 - x_0\right]K(x,t)f(t).$$

Therefore two functions $\bar{f}(x)$ and $f(x)$ related by Eq. (B2) will both satisfy differential equation (B1) provided the contour c is chosen such that the integral converges and the value of $P(x,t)$ is the same at each end of the contour.

APPENDIX C: A SECOND SOLUTION BY EXPANSION IN IRREGULAR CONFLUENT HYPERGEOMETRIC FUNCTIONS

The validity of Eq. (70) is proven for arbitrary η, and convergence properties of the expansion are discussed. Start with Eq. (1):

$$x(x-x_0)\frac{d^2y}{dx^2} + (B_1 + B_2x)\frac{dy}{dx} + [\omega^2 x(x-x_0) - 2\eta\omega(x-x_0) + B_3]y = 0.$$

The substitution $y(x) = e^{+i\omega x}f(x)$ yields Eq. (56):

$$x(x-x_0)f_{,xx} + [B_1 + B_2x + 2i\omega x(x-x_0)]f_{,x} + [(B_2 + 2i\eta)i\omega x + 2\eta\omega x_0 + i\omega B_1 + B_3]f = 0,$$

which, with the substitutions $z = -2i\omega x$ and $z_0 = -2i\omega x_0$, can be more suggestively written as

$$z(z-z_0)\frac{d^2f}{dz^2} + (D_1 + D_2z - z^2)\frac{df}{dz} + (D_3 + D_4z)f = 0, \tag{C1}$$

where

$$D_1 = -2i\omega B_1, \quad D_2 = B_2 - 2i\omega x_0, \quad D_3 = B_3 + 2\eta\omega x_0 + i\omega B_1, \quad D_4 = -\tfrac{1}{2}B_2 - i\eta. \tag{C2}$$

We expand $f(z)$ as

$$f(z) = \sum_{n=0}^{\infty} b_n U(a+n, -B_1/x_0, z), \tag{C3}$$

where the $U(a+n, -B_1/x_0, z)$ are irregular solutions to the confluent hypergeometric equation

$$zU_{n,zz} - (B_1/x_0 + z)U_{n,z} - (a+n)U_n = 0. \tag{C4}$$

We have denoted $U(a+n, -B_1/x_0, z)$ by U_n for notational convenience, and the parameter a is to be determined. The confluent hypergeometric functions used here are those defined by Slater.[28] They satisfy the differential property

$$zU_{n,z} = -(n+a)U_n + (n+a)\times(n+a+1+B_1/x_0)U_{n+1} \tag{C5}$$

and are a solution that is minimal as $n\to\infty$ of the recurrence relation

$$U_{n-1} - [2(n+a) + B_1/x_0 + z]U_n + (n+a)(n+a+1+B_1/x_0)U_{n+1} = 0. \tag{C6}$$

Substituting (C3) into (C1) and using (C4)–(C6), we obtain

$$\sum_{n=0}^{\infty} b_n\{(n+a-B_2/2-i\eta)U_{n-1} + (n+a)(n+a+1+B_1/x_0)(n+a+B_2/2-i\eta+B_1/x_0)U_{n+1} - [2(n+a)^2 + (2B_1/x_0 - 2i\eta - 2i\omega x_0)(n+a) - (2\eta\omega x_0 + (B_2/2+i\eta)B_1/x_0 + i\omega B_1 + B_3)]U_n\} = 0. \tag{C7}$$

The coefficient of U_{-1} must vanish if the series is to start at $n=0$, so we must fix a to be $a = B_2/2 + i\eta$. Equation (C7) can then be re-indexed to yield

$$(\bar{\alpha}_0 b_1 + \bar{\beta}_0 b_0)U_0 + \sum_{n=1}^{\infty}(\bar{\alpha}_n b_{n+1} + \bar{\beta}_n b_n + \bar{\gamma}_n b_{n-1})U_n = 0, \tag{C8}$$

where

$$\bar{\alpha}_n = n+1,$$
$$\bar{\beta}_n = -[2n^2 + 2(B_2 + B_1/x_0 + i\eta - i\omega x_0)n + (B_2 + B_1/x_0)(B_2/2 + i\eta - i\omega x_0) - B_3],$$
$$\bar{\gamma}_n = (n + B_2/2 + i\eta - 1)(n + B_2/2 + i\eta + B_1/x_0)\times(n - 1 + B_2 + B_1/x_0).$$

Equation (C8) can hold only if the coefficient of each U_n vanishes. Letting $b_n = \Gamma(B_2 + B_1/x_0 + n)a_n$ in expansion (C3) and recurrence relation (C8), we can obtain as the recurrence relation for the a_n

$$\begin{aligned} \alpha_0 a_1 + \beta_0 a_0 &= 0, \\ \alpha_n a_{n+1} + \beta_n a_n + \gamma_n a_{n-1} &= 0, \quad n = 1,2,..., \end{aligned} \tag{C9}$$

where

$$\begin{aligned} \alpha_n &= (n+1)(n + B_2 + B_1/x_0), \\ \beta_n &= -2n^2 - 2[B_2 + i(\eta - \omega x_0) + B_1/x_0]n - [(B_2/2 + i\eta)(B_2 + B_1/x_0) - i\omega(B_1 + B_2x_0) - B_3], \\ \gamma_n &= (n - 1 + B_2/2 + i\eta)(n + B_2/2 + i\eta + B_1/x_0), \end{aligned} \tag{C10}$$

which is the same recurrence relation as for the Jaffé coefficients, Eq. (41).

Convergence of the series (C3) may be analyzed by considering the sequence of functions $\{\bar{U}_n,\ n = 0,1,2,...\}$ defined by $\bar{U}_n = \Gamma(c+n)U(a,b,z)$, where $a = B_2/2 + i\eta$, $b = -B_1/x_0$, and $c = B_2 + B_1/x_0$. The $\bar{U}_n$ are a minimal solution to the recurrence relation

$$(n+a)(n+a+1-b)/(n+c)\bar{U}_{n+1} - (2n + 2a - b + z)\bar{U}_n + (n-1+c)\bar{U}_{n-1} = 0, \tag{C11}$$

which, after dividing by n and retaining terms to $O(1/n)$, takes the limiting form

$$[1+(2a+1-b-c)/n]\overline{U}_{n+1} - [2+(2a-b+z)/n]\overline{U}_n + [1+(c-1)/n]\overline{U}_{n-1} + O(n^{-2})\approx 0$$

Hence

$$\lim_{n\to\infty}(\overline{U}_{n+1}/\overline{U}_n) = 1-\sqrt{-z/n}.$$

See the convergence discussion in Sec. IV A. Here $z=-2i\omega x$ and the branch of the square root is taken such that $\mathrm{Re}(\sqrt{-z})\geqslant 0$ [the $\overline{U}_n$ being a minimal solution to (C6)]. We do not consider the case when z is positive real. Since

$$\lim_{n\to\infty}(a_{n+1}/a_n) = 1\pm\sqrt{-2i\omega x_0/n}$$

[see Eqs. (42) and (44)], our final result is that the limiting ratio of successive terms of series (C3) is given by

$$\lim_{n\to\infty}\frac{b_{n+1}U(a+n+1,b,-2i\omega x)}{b_n U(a+n,b,-2i\omega x)} = 1-\frac{\sqrt{-2i\omega x}\pm\sqrt{-2i\omega x_0}}{\sqrt{n}}. \qquad \text{(C12)}$$

The $(+)$ sign is obtained when ω is an eigenfrequency and the a_n are themselves a minimal solution to recurrence relation (C9). Then the series converges at $x=x_0$. When ω is not an eigenfrequency the a_n are dominant and the $(-)$ sign prevails in Eq. (C12). In this case the series converges for all $x>x_0$, but diverges when $x=x_0$. This is precisely the behavior expected of a second solution.

[1]W. H. Press and S. A. Teukolsky, Astrophys. J. **185**, 649 (1973).

[2]C. W. Clark, Phys. Rev. A **20**, 1875 (1979).

[3]S. Detweiler, in *Sources of Gravitational Radiation*, edited by L. Smarr (Cambridge U. P., London, 1979), p. 217.

[4]E. W. Leaver, "A spectral decomposition of the perturbation response of the Schwarzschild geometry," submitted to Phys. Rev. D.

[5]W. G. Baber and H. R. Hassé, Proc. Cambridge Philos. Soc. **25**, 564 (1935).

[6]W. Gautschi, SIAM Rev. **9**, 24 (1967).

[7]E. Hylleraas, Z. Phys. **71**, 739 (1931).

[8]G. Jaffé, Z. Phys. **87**, 535 (1934).

[9]J. Meixner, F. S. Schäfke, and G. Wolf, "Mathieu Functions and Spheroidal Functions and their Mathematical Foundations (further studies)," in *Lecture Notes in Mathematics*, Vol. 837 (Springer, Berlin, 1980).

[10]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[11]E. W. Leaver, Proc. R. Soc. Lond. Ser. A **402**, 285 (1985).

[12]H. Eyring, J. Walter, and G. E. Kimball, *Quantum Chemistry* (Wiley, New York, 1948).

[13]J. A. Wheeler, Phys. Rev. **97**, 511 (1955).

[14]T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).

[15]F. J. Zerilli, Phys. Rev. D **2**, 2141 (1970).

[16]S. Chandrasekhar, Proc. R. Soc. Lond. Ser. A **343**, 289 (1975).

[17]S. Chandrasekhar and S. Detweiler, Proc. R. Soc. Lond. Ser. A **344**, 441 (1975).

[18]S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).

[19]R. H. Price, Phys. Rev. D **5**, 2419 (1972).

[20]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), exercise 32.10.

[21]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).

[22]A. N. Lowan, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (National Bureau of Standards, Washington, 1964), Chap. 21. There are other definitions for the oblate coordinates λ and μ.

[23]E. D. Fackerell and R. G. Crossman, J. Math. Phys. **18**, 1849 (1977).

[24]R. A. Breuer, M. P. Ryan, and S. Waller, Proc. R. Soc. Lond. Ser. A **358**, 71 (1977).

[25]C. Hunter and B. Guerrieri, Stud. Appl. Math. **66**, 217 (1982).

[26]V. Ferrari and B. Mashhoon, Phys. Rev. D **30**, 295 (1984).

[27]A. H. Wilson, Proc. R. Soc. Lond. Ser. A **118**, 617 (1928).

[28]L. J. Slater, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (National Bureau of Standards, Washington, 1964), Chap. 13.

[29]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).

[30]S. Chandrasekhar, Proc. R. Soc. Lond. Ser. A **392**, 1 (1984).

[31]A. L. Rabenstein, *Introduction to Ordinary Differential Equations* (Academic, New York, 1966).

[32]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1926), p. 201.

[33]I. S. Gradshteyn and I. M. Ryzhik, Ref. 29, Eq. 7.414.8.

[34]J. A. Stratton, Proc. Natl. Acad. Sci. U.S.A. **21**, 51, 316 (1935). We use the term "Stratton solution" somewhat loosely, as a very important generalization was made in a subsequent paper coauthored by L. J. Chu (see Ref. 38). Contributions were also made by Morse [P. M. Morse, Proc. Natl. Acad. Sci. U.S.A. **21**, 56 (1935): also see Ref. 10].

[35]P. M. Morse and H. Feshbach, Ref. 10, Eq. 11.3.87.

[36]P. M. Morse and H. Feshbach, Ref. 10, Eq. 11.3.39.

[37]P. M. Morse and H. Feshbach, Ref. 10, p. 1503. There are brackets missing from the expression for β_n in some editions of the book.

[38]The generalized expansions (86) and (87) were first given by Chu and Stratton [L. J. Chu and J. A. Stratton, J. Math. Phys. (Cambridge, Mass.) **20**, 259 (1941)], and reprinted in *Spheroidal Wave Functions* [J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and F. J. Corbató (Wiley, New York, 1956)]. Our present equations (89) and (90) are completely equivalent to Eqs. (95) and (96) of Chu and Stratton, the important difference being that in the present study we regard the separation constant A_{lm} as being fixed, and use these equations to define the characteristic value of the phase parameter ν.

[39]P. M. Morse and H. Feshbach, Ref. 10, p. 1507.

[40]B. P. Sinha and R. H. Macphie, J. Math. Phys. **16**, 2378 (1975).

[41]L. Jen and C. Hu, IEEE Trans. Antennas Propag. **31**, 382 (1983).

[42]R. Penrose, Nuovo Cimento **1** (special number), 252 (1969).

[43]S. A. Teukolsky and W. H. Press, Astrophys. J. **193**, 443 (1974).

[44]S. Detweiler, Astrophys. J. **239**, 292 (1980).

[45]A. R. Barnett, D. H. Feng, J. W. Steed, and L. J. B. Goldfarb, Comput. Phys. Commun. **8**, 377 (1974).

[46]E. W. Leaver, "GSWFN—Generalized spheroidal wave-functions for complex arguments," to be submitted to Trans. Math. Software.

[47]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1958).

[48]E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second Order Differential Equations* (Oxford U. P., New York, 1962).

[49]E. L. Ince, Ref. 32, Chap. 5.

Prolongation structures and Lie algebra real forms

Chris Radford[a)]
Department of Applied Mathematics, University of Sydney, New South Wales 2006, Australia

(Received 26 June 1985; accepted for publication 26 November 1985)

A technique useful as an aid for finding finite-dimensional representations of a Wahlquist–Estabrook prolongation algebra is introduced. The technique is illustrated with an examination of equations of the form $\partial^2 z^a/\partial x^1\,\partial x^2 = f^a(z^b)$.

I. INTRODUCTION

Considerable success has been achieved over the last two decades in solving certain classes of nonlinear partial differential equations, using the related methods of Bäcklund transformations and inverse spectral transforms (see, for example, Refs. 1 and 2). Given a differential equation (or system) the problem of finding a Bäcklund transformation or linear scattering problem (if they exist) falls into two basic initial steps. The first, the application of the Wahlquist–Estabrook prolongation method,[3] is reasonably systematic. The second step is to find a finite-dimensional representation for the infinite-dimensional, constrained, free Lie algebra which results from an application of the prolongation technique. With the increasing complexity of the equations being attacked by these methods, it is this second step that has proved most difficult to complete. Dodd and Fordy[4] collected and applied several theorems from Lie algebra theory in a work that may be regarded as an initial step in a systematic approach to the prolongation algebra problem. The present work may be regarded, in part, as an adjunct to their paper.

In the present paper we will exploit the fact that many of the equations encountered in applications possess a prolongation algebra that has a certain involutive symmetry. The technique introduced is particularly useful for hyperbolic (or elliptic) equations—these equations are not quasipolynomial flows and hence cannot be dealt with (systematically) by the Dodd and Fordy process. For example, equations of the form $\partial^2 z/\partial x^1\,\partial x^2 = f(z)$ possess the simple symmetry of interchange of the coordinates x^1 and x^2. This symmetry carries through to the prolongation algebra (see Sec. III), and may be used to considerably simplify the task of finding finite-dimensional representations.

Before discussing particular differential equations, we briefly review some of the theory on real forms of the complex semisimple Lie algebras. This we do in Sec. II, while Secs. III and IV are devoted to applications.

II. REAL FORMS OF COMPLEX SEMISIMPLE LIE ALGEBRAS

Let $\mathscr{g}$ be a complex Lie algebra (see Refs. 5 and 6 for general theory), a real subalgebra $\mathscr{g}_0$ of $\mathscr{g}$ is *a real form* of $\mathscr{g}$ if $\mathscr{g}$ is the complexification of $\mathscr{g}_0$ (see Refs. 5–7) and, in particular, the article by Macdonald,[8] i.e., $\mathscr{g} = \mathscr{g}_0 \oplus i\mathscr{g}_0$ (direct sum). Any real form $\mathscr{g}_0$ determines (and, conversely, is determined by) a mapping c: $\mathscr{g}\to\mathscr{g}$, viz., $X + iY \to X - iY$; $X, Y\in\mathscr{g}_0$. The map c has the following properties: (i) c is semilinear, $c(\alpha X + \beta Y) = \bar{\alpha}c(X) + \bar{\beta}c(Y)$, for $X, Y\in\mathscr{g}$ and $\alpha,\beta\in C$; (ii) c is an involution, $c^2(X) = X$, for all $X\in\mathscr{g}$; and (iii) $c[X,Y] = [c(X),c(Y)]$, for $X,Y\in\mathscr{g}$.

A mapping c:$\mathscr{g}\to\mathscr{g}$ with these properties is called a conjugation of $\mathscr{g}$ (see Ref. 8). There is a one-to-one correspondence between conjugations of $\mathscr{g}$ and real forms of $\mathscr{g}$: the real form determined by the conjugation c is just $\mathscr{g}_0 = \{X\in\mathscr{g}: c(X) = X\}$.

We now specialize our considerations somewhat and take $\mathscr{g}$ to be a complex simple Lie algebra. Now[5–7] $\mathscr{g}$ may be decomposed as a vector space direct sum as follows:

$$\mathscr{g} = \mathscr{n}^- \oplus \mathscr{h} \oplus \mathscr{n}^+ ,$$

where $\mathscr{h}$ is a Cartan subalgebra, $\mathscr{n}^+$ the maximal nilpotent algebra of positive root vectors, and $\mathscr{n}^-$ the subalgebra of negative root vectors. If we choose, as the basis of $\mathscr{g}$, the Cartan subalgebra and the root vectors we obtain the Cartan–Weyl commutation relations. These are

$$[H, E_\alpha] = \alpha(H)E_\alpha\,; \quad H\in\mathscr{h}, \quad E_\alpha\in\mathscr{n}^+\oplus\mathscr{n}^-, \quad \text{with } \alpha\in\Sigma \quad (\text{the root space}^{5-7}),$$

$$[E_\alpha, E_\beta] = \begin{cases} 0, & \text{if } \alpha+\beta\neq 0 \text{ and } \alpha+\beta\notin\Sigma, \\ H_\alpha\in\mathscr{h}, & \text{if } \alpha+\beta = 0, \\ N_{\alpha,\beta}E_{\alpha+\beta}, \quad E_{\alpha+\beta}\in\mathscr{n}^-\oplus\mathscr{n}^+, & \text{if } \alpha+\beta\in\Sigma, \end{cases} \tag{2.1}$$

where the constants $N_{\alpha,\beta}$ satisfy $N_{\alpha,\beta} = -N_{-\alpha,-\beta}$ and certain other relations.[5–7]

Let c be a conjugation of $\mathscr{g}$. Then c acts on the root space Σ as follows: For each root $\alpha\in\Sigma$ define α^* by

$$\alpha^*(H) = \overline{\alpha[c(H)]}$$

(overbar denotes complex conjugation), which defines an involutory isometry * on the Euclidean root space Σ. Making use of the commutation relation (2.1) we determine the effect of c on the Cartan–Weyl basis as follows[7]:

$$c\colon H\to c(H)\,, \quad \text{with } [c(H), E_{\alpha^*}] = \alpha^*(H)E_{\alpha^*}\,,$$
$$c\colon E_\alpha \to c(E_\alpha) = \rho_\alpha E_{\alpha^*}\,,$$

with the complex constants ρ_α satisfying (2.2)

$$\bar{\rho}_\alpha\rho_{\alpha^*} = 1\,, \quad \rho_\alpha\rho_{-\alpha} = 1\,, \quad \text{and}$$
$$\rho_{\alpha+\beta}N_{\alpha,\beta} = \rho_\alpha\rho_\beta N_{\alpha^*,\beta^*}\,.$$

The above theory (and the theorems presented in Ref.

[a)] Present address: Department of Mathematics, Statistics and Computing Science, The University of New England, Armidale, New South Wales 2351, Australia.

 0022-2488/86/051266-05$02.50

4) may be collected together to give a prescription for finding a finite-dimensional representation ("closure") of a given prolongation algebra. The techniques in Ref. 4 rely on the identification of nilpotent elements. This, however, is not absolutely necessary, for one can use the Jordan–Chevalley decomposition (Humphreys[5]) to (in turn) identify an element of the prolongation algebra with the various canonical elements (under the Jordan–Chevally decomposition) of the finite simple Lie algebra into which one seeks to embed the prolongation algebra (see Sec. III). So the procedure for finding a closure of a prolongation algebra may be roughly described as follows.

(i) Determine (if it exists!) the action of the involutory automorphism on the prolongation algebra.

(ii) Select an element of the prolongation algebra to be identified with the various classes of canonically decomposed elements of the finite simple Lie algebra—using the theorems on nilpotent elements (*à la* Dodd and Fordy[4]) if possible.

(iii) Beginning with the algebra A_1 (and proceeding to $A_1, A_2, \ldots$, until closure is obtained), examine each conjugation of A_l and each conjugacy class of canonically decomposed elements [for (i) and (ii), respectively] until a nontrivial closure is obtained (i.e., until a non-Abelian finite-dimensional Lie algebra, embedded in A_l, is obtained).

The beauty in applying the theory of real forms is that, at each step in the calculation of (iii), the number of generators in the prolongation algebra is (in most cases) effectively halved.

Before we apply these methods to particular equations, we present some results (to be used in our applications) on the real forms of A_1 and A_2.

Conjugations of A_1: The three-dimensional, complex, simple Lie algebra A_1 has (in standard notation) a Cartan–Weyl basis $\{H, E_\alpha, E_{-\alpha}\}$ and canonical commutation relations

$$[H,E_\alpha] = \tfrac{1}{2}E_\alpha, \quad [H,E_{-\alpha}] = -\tfrac{1}{2}E_{-\alpha},$$
$$[E_\alpha,E_{-\alpha}] = H.$$

There are only two conjugations with respect to real form. They are

(1) $\alpha^* = \alpha$, giving the split real form $SL(2,R)$,

(2) $\alpha^* = -\alpha$, giving the compact real form $SU(3)$.

For (1) we have $c(E_\alpha) = e^{i\theta}E_\alpha$, $c(E_{-\alpha}) = e^{-i\theta}E_{-\alpha}$, and $c(H) = H$, where θ is a real constant. For (2) we have $c(E_\alpha) = \rho E_{-\alpha}$, $c(E_{-\alpha}) = (1/\rho)E_\alpha$, and $c(H) = -H$, where ρ is a real constant.

Conjugations of A_2: The root system for A_2, in terms of a pair of simple roots α_1 and α_2, may be presented as

$$\Sigma \begin{cases} \left.\begin{matrix} \alpha_1 \qquad\qquad \alpha_2 \\ \alpha_1+\alpha_2 \end{matrix}\right\} \Sigma^+ \quad \text{(positive roots)}, \\ \left.\begin{matrix} -\alpha_1 \qquad\qquad -\alpha_2 \\ -\alpha_1-\alpha_2 \end{matrix}\right\} \Sigma^- \quad \text{(negative roots)}. \end{cases}$$

The conjugations may be calculated in a straightforward manner (using either combinations of the Weyl group and the outer automorphisms or by searching for all involutive isometries of the Euclidean root space). There are eight of them, as shown in Table I.

III. THE EQUATION $\partial^2 z/\partial x^1 \partial x^2 = f(z)$

In Ref. 9, Shadwick determined all equations of the above form that possess a non-Abelian prolongation algebra in one prolongation variable. From Ref. 9, the prolongation structure equations are (strictly, the prolongation equations restricted to the solution submanifold)

$$y^k_{,x^1} = p_1 B^k + A^k, \quad y^k_{,x^2} = -p_2 B^k + C^k,$$

with (3.1)

$$A^k = A^k(z,y^j), \quad C^k = C^k(z,y^j),$$

and

$$B^k = B^k(y^j)$$

(here $y^k_{,x^1} \equiv \partial y^k/\partial x^1$, etc.). The y^k are the prolongation variables, the space of these variables carries the representation of the prolongation algebra. The integrability conditions for (3.1) are simply the original differential equation together with the contact conditions $p_1 = z_{,x^1}$, $p_2 = z_{,x^2}$, and the Lie algebraic constraints given by

$$[B,A] = \frac{\partial A}{\partial z}, \quad [C,B] = \frac{\partial C}{\partial z}, \quad [A,C] = 2f(z)B, \tag{3.2}$$

where

$$A = \sum_k A^k \frac{\partial}{\partial y^k}, \quad [B,A] = \left[B^j \frac{\partial A^k}{\partial y^j} - A^j \frac{\partial B^k}{\partial y^j}\right] \frac{\partial}{\partial y^k}$$

(summation convention on repeated indices assumed). The first two of these equations have the formal solutions

$$A = e^{z\,\mathrm{ad}\,B}A_0, \quad C = e^{-z\,\mathrm{ad}\,B}C_0, \tag{3.3}$$

where A_0 and C_0 are algebra elements independent of z and ad refers to the adjoint representation $(\mathrm{ad}\,X)Y = [X, Y]$. Using (3.3), the last equation of (3.2) may be written as

$$[e^{2z\,\mathrm{ad}\,B}A_0, C_0] = 2f(z)B. \tag{3.4}$$

We now observe that Eqs. (3.2) possess an involutive symmetry (which we denote by *) given by

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} \mapsto \begin{pmatrix} A^* \\ B^* \\ C^* \end{pmatrix} = \begin{pmatrix} C \\ -B \\ A \end{pmatrix}. \tag{3.5}$$

TABLE I. The conjugations of A_2.

α_1^*	α_2^*	
α_1	α_2	giving the split form
α_1	$-\alpha_1-\alpha_2$	
$-\alpha_1-\alpha_2$	α_2	
α_2	α_1	diagram automorphism
$-\alpha_2$	$-\alpha_1$	
$-\alpha_1$	$\alpha_1+\alpha_2$	
$\alpha_1+\alpha_2$	$-\alpha_2$	
$-\alpha_1$	$-\alpha_2$	giving the compact form

Properties (ii) and (iii) of Sec. II are easily verified for *. Property (i), semilinearity, is required to hold so as to guarantee the reality of the function $f(z)$ defined by the last equation of (3.2). Thus * defines a conjugation in the sense of Sec. II. Equation (3.3) and (3.4) may now be written as

$$A = e^{z\,\mathrm{ad}\,B}A_0, \quad C = e^{-z\,\mathrm{ad}\,B}A_0^*, \qquad [e^{2z\,\mathrm{ad}\,B}A_0, A_0^*] = 2f(z)B. \tag{3.6}$$

We now wish to find those functions $f(z)$, for which $z_{,x_1x_2} = f(z)$ has a nontrivial prolongation structure (at least for closures within the algebras A_1 and A_2). We proceed as follows.

(A) Start with one of the algebras A_l and reduce B to canonical form using the Jordan–Chevally decomposition (Humphries[5]).

(B) Elucidate the various forms of the algebra elements A and C for all canonical forms B and all possible conjugations.

(C) Using the last equation of (3.6) and the various forms for A and C deduced from (B), determine the distinct functions $f(z)$.

The algebra A_1: Using the notation of Sec. II there are only two possible forms for B: B semisimple, in which case we may take (using the automorphisms of A_1) B proportional to H; and B nilpotent, in which case we may take B proportional to E_α. Combining these two cases with the two conjugations listed in Sec. II and the fact that $B^* = -B$ we obtain three possible cases:

(1) $B = iH, \quad \alpha^* = \alpha, \quad (E_\alpha)^* = e^{i\theta}E_\alpha, \quad \theta$ real;

(2) $B = H, \quad \alpha^* = -\alpha, \quad (E_\alpha)^* = \rho E_{-\alpha}, \quad \rho$ real;

and

(3) $B = E_\alpha, \quad \alpha^* = \alpha, \quad (E_\alpha)^* = -E_\alpha$.

Case (1): Write

$$A_0 = a_1 + H + a_2E_\alpha + a_3E_{-\alpha},$$

a_1, a_2, and a_3 constants. Then,

$$C_0 = A_0^* = \bar a_1H + e^{i\theta}\bar a_2E_\alpha + e^{-i\theta}\bar a_3E_{-\alpha}.$$

We also have

$$e^{2z\,\mathrm{ad}\,B}A_0 = a_1H + a_2e^{iz}E_\alpha + a_3e^{-iz}E_{-\alpha}.$$

So from (3.6) and A_1, commutation relations of Sec. II

$$2if(z)H = (a_2\bar a_3e^{i(z-\theta)} - a_3\bar a_2e^{-i(z-\theta)})H + \tfrac12(a_1\bar a_2e^{i\theta} - \bar a_1a_2e^{iz})E_\alpha + \tfrac12(a_3\bar a_1 - a_1\bar a_3e^{-\theta}e^{-iz})E_{-\alpha},$$

whence $a_1 = 0$. By rescaling the coordinates (x^1, x^2) and performing a translation on z we may take $a_2\bar a_3 = 1$ and $\theta = 0$ and we have arrived at the sine–Gordon equation with

$$A = e^{(i/2)z}E_\alpha + e^{-(i/2)z}E_{-\alpha},$$
$$C = e^{-(i/2)z}E_\alpha + e^{(i/2)z}E_{-\alpha},$$
$$f(z) = \sin z.$$

Case (2): The calculation proceeds in a similar manner to case (1). We find two possible cases: the Liouville equation

$$A = e^{(1/2)z}E_\alpha, \quad C = e^{(1/2)z}E_{-\alpha}, \quad f(z) = \tfrac12 e^z;$$

and the sinh–Gordon equation

$$A = e^{(1/2)z}E_\alpha + e^{-(1/2)z}E_{-\alpha},$$
$$C = e^{(1/2)z}E_{-\alpha} + e^{-(1/2)z}E_\alpha,$$
$$f(z) = \sinh + z.$$

Case (3): Write

$$A_0 = a_1H + a_2E_\alpha + a_3E_{-\alpha}.$$

So

$$A_0^* = \bar a_1H - \bar a_2E_\alpha - \bar a_3E_{-\alpha},$$

and

$$e^{2z\,\mathrm{ad}\,B}A_0 = (a_1 + 2a_3z)H + (a_2 - a_1z - a_3z)E_\alpha + a_3E_{-\alpha}.$$

From (3.6) we find $a_3 = 0$, and hence

$$f(z) = -\tfrac14(a_1\bar a_2 + \bar a_1a_2) + (a_1\bar a_1/4)z.$$

As one would expect, these are precisely the equations obtained by Shadwick[9] for the one pseudopotential case—A_1 and its various real forms are the only nontrivial algebras with a (real or complex) one-dimensional representation.

The algebra A_2: Using the Jordan–Chevalley decomposition we obtain the following canonical forms for B:

$B = a_1H_1 + a_2H_2$

(at least one of a_1, a_2 nonzero),

$\{H_1, H_2\}$ the basis of a Cartan subalgebra,

$B = \delta_1E_{\alpha_1} + \delta_2E_{\alpha_2}, \quad \delta_j = 0, 1$, not both zero,

$B = a(H_1 + 2H_2) + E_{\alpha_2}, \quad a \neq 0$.

To obtain the necessary results, we must test each conjugation of Sec. II with each of the canonical forms for B in Eqs. (3.6). It turns out (after some tedious calculations!) that only one new equation arises (aside from those associated with A_1). This is the Bullough–Dodd equation.[10] It is associated with the first canonical form of B (B purely semisimple), and several of the conjugations, for example,

$$\alpha_1^* = -\alpha_2, \quad \alpha_2^* = -\alpha_1,$$
$$B = 6(H_1 + H_2),$$
$$A = ae^zE_{\alpha_1} + 12be^zE_{\alpha_2} + \sqrt{12}ce^{-2z}E_{-\alpha_1-\alpha_2},$$
$$C = ae^zE_{-\alpha_2} + 12be^zE_{-\alpha_1} + \sqrt{12}ce^{-2z}E_{\alpha_1+\alpha_2},$$
$$f(z) = abe^{2z} - ce^{-4z},$$

a, b, and c real constants.

New equations for A_l ($l > 3$)? Although the calculations rapidly become unmanageable (by hand at least), results thus far obtained indicate that no new equations arise in the algebra A_3. The obvious question now is whether there are any equations of the given form, aside from those listed above, that possess nontrivial prolongation structures. We hope to return to this question in a future publication.

IV. THE SYSTEM OF EQUATIONS $\partial^2z^k/\partial x^1\partial x^2 = f^k(z^l)$

Equations in this class include (in addition to those of Sec. II) the generalized two-dimensional Toda-lattice equations.[11–13]

The prolongation algebra equations are

$$y^k_{,x^1} = p^a_1 B^k_a + A^k \quad \text{(summation on the index } a\text{)},$$
$$y^k_{,x^2} = -p^a_2 B^k_a + C^k, \tag{4.1}$$
$$\text{with } B^k_a = B^k_a(y^j), \quad A^k = A^k(z^a, y^j),$$
$$\text{and } C^k = C^k(z^a, y^j), \quad a = 1,2,...,m.$$

The integrability conditions for (4.1) give the original differential equation together with the contact conditions $p^a_1 = z^a_{,x^1}$, $p^a_2 = z^a_{,x^2}$ and the Lie algebra constraints given by

$$[B_a, B_b] = 0,$$
$$[B_a, A] = \frac{\partial A}{\partial z^a}, \quad [C, B_a] = \frac{\partial C}{\partial z^a}, \tag{4.2}$$
$$[A, C] = 2f^a(z^b)B_a \quad \text{(summation on } a\text{)}.$$

As in Sec. II we can obtain the formal solutions for A and C,

$$A = e^{z^a \operatorname{ad} B_a} A_0, \quad C = e^{-z^a \operatorname{ad} B_a} C_0, \tag{4.3}$$

which allow us to rewrite the last equation of (4.2) (using $[B_a, B_b] = 0$) as

$$[e^{2z^a \operatorname{ad} B_a} A_0, C_0] = 2f^a B_a. \tag{4.4}$$

As in Sec. II we have a conjugation *, which acts as follows:

$$*: \begin{pmatrix} A \\ B_a \\ C \end{pmatrix} \mapsto \begin{pmatrix} A^* \\ B^*_a \\ C^* \end{pmatrix} = \begin{pmatrix} C \\ -B_a \\ A \end{pmatrix}, \tag{4.5}$$

and Eqs. (4.3) and (4.5) are rewritten as

$$A = e^{z^a \operatorname{ad} B_a} A_0, \quad C = e^{-z^a \operatorname{ad} B_a} A^*_0,$$
$$[e^{2z^a \operatorname{ad} B_a} A_0, A^*_0] = 2f^a B_a \tag{4.6}$$

(summation on the repeated index a is always assumed).

Clearly the above equations contain many infinite classes of systems of differential equations possessing nontrivial prolongation structures (the generalized Toda lattices of Refs. 11–13 being but one example). The main problem in finding new examples of such systems is to satisfy the last equation of (4.6), and this usually requires some form of *Ansatz*. In the present paper we will briefly consider two such *Ansätze* (one being a generalization of that of Ref. 12).

Example (1): We consider here the case in which all the B_a are nilpotent [the generalization of case (3) of Sec. III]. The B_a then form a commuting subalgebra of nilpotent elements. The most obvious *Ansatz* is to take one of the B_a to be a principal nilpotent element (Kostant[14]) of the closed Lie algebra. We may then take the B_a ($a = 1,2,...,l$; l the rank of the algebra) to be a basis of the l-dimensional commuting subalgebra $\operatorname{Ker}(\operatorname{ad} B_1)$, where B_1 is principal nilpotent—see Ref. 14 for results on principal nilpotent elements. Equations (4.6) can then be satisfied by taking A_0 to be in the Cartan subalgebra. For example, consider the algebra A_2 and the conjugation associated with the diagram automorphism, $\alpha^*_j = \alpha_{3-j}$, $j = 1,2$; notation is that of Sec. II. We may take

$$B_1 = E_{\alpha_1} + E_{\alpha_2} \quad \text{(principal nilpotent)},$$
$$B_2 = E_{\alpha_1 + \alpha_2},$$
$$(E_{\alpha_j})^* = -E_{\alpha_{3-j}}, \quad \text{so } B^*_a = -B_a,$$
$$A_0 = 3(a_1 H_1 + a_2 H_2),$$

where $\{H_1, H_2\}$ is the Cartan subalgebra and a_1 and a_2 are real constants. Then

$$\begin{aligned} e^{2z^a \operatorname{ad} B_a} A_0 = {} & 3(a_1 H_1 + a_2 H_2) \\ & + z_1[(a_2 - 2a_1)E_{\alpha_1} + (a_1 - 2a_2)E_{\alpha_2}] \\ & + [\sqrt{\tfrac{3}{2}}\,(a_1 - a_2)(z^1)^2 \\ & - (a_1 + a_2)z^2]E_{\alpha_1 + \alpha_2}, \end{aligned}$$

and

$$A^*_0 = 3(a_1 H_1 + a_1 H_2),$$

so that

$$\begin{aligned} & [e^{2z^a \operatorname{ad} B_a} A_0, A^*_0] \\ & = \tfrac{1}{2}(2a_2 - a_1)(a_2 - 2a_1)z^1(E_{\alpha_1} + E_{\alpha_2}) \\ & \quad + \tfrac{1}{2}(a_1 + a_2)[\sqrt{\tfrac{3}{2}}\,(a_1 - a_2)(z^1)^2 \\ & \quad - (a_1 + a_2)z^2]E_{\alpha_1 + \alpha_2}, \end{aligned}$$

yielding the following coupled system of differential equations:

$$z^1_{,x^1 x^2} = \tfrac{1}{4}(2a_2 - a_1)(a_2 - 2a_1)z^1,$$
$$z^2_{,x^1 x^2} = \tfrac{1}{4}[\sqrt{\tfrac{3}{2}}\,(a_1^2 - a_2^2)(z^1)^2 - (a_1 + a_2)^2 z^2].$$

The system can be easily generalized to a system with l degrees of freedom associated with the algebra A_l, although we will not pursue this here.

Example (2): We consider here the case where the B_a are semisimple. Hence, the B_a may be taken to be elements of a Cartan subalgebra of the closed Lie algebra $\mathscr{g}$. This still leaves us with a very wide class of problems, so again we adopt an *Ansatz* to satisfy Eqs. (4.6). To this end we introduce two new definitions.

(i) We will say that a subset R of the root system Σ of $\mathscr{g}$ is **-admissible* if, for each $\alpha \in R$, $\alpha^* + \beta \notin \Sigma$, for all $\beta \in R$.

(ii) A *-admissible root system R will be called **-reduced* if, for each $\alpha \in R$, there exists $\beta \in R$ such that $\alpha^* + \beta = 0$.

If we choose the A_0 to be linear combinations of the root vectors corresponding to a *-reduced root system, then we can ensure (for appropriate choices of the B_a) that the last equation of (4.6) is nontrivially satisfied. In fact, the generalized Toda lattice systems of Ref. 12 may (with the correct choice for the B_a) be constructed in this way, with * being the conjugation that gives the compact real form of $\mathscr{g}$; i.e., $\alpha^* = -\alpha$ (and, in this case, the *-admissible root system is an admissible root system in the usual sense).

We now consider a particular example, the algebra A_l with $l = 2k$, and * the conjugation given by $\alpha^*_j = -\alpha_{l+1-j}$, $j = 1,2,...,l = 2k$; α_j being a set of simple roots for A_l. The following root system is *-reduced:

$$R = \{\alpha_1, \alpha_2, ..., \alpha_l, -\alpha_1 - \alpha_2 \cdots - \alpha_l\}.$$

Here, R is also an admissible root system in the usual sense. We then take

$$A_0 = \sqrt{(l+1)}\left[\sum_{j=1}^{l=2k} a_j E_{\alpha_j} + a_{l+1} E_{-\alpha_1 \cdots \alpha_l}\right],$$

and

$$B_a = \begin{cases} (l+1)(H_j + H_{l+1-j}), & a = j \quad 1 \leqslant j \leqslant k, \\ i(l+1)(H_j - H_{l+1-j}), & a = j+k, \quad 1 \leqslant j \leqslant k, \end{cases}$$

where $\{H_j\}_{j=1}^{l}$ is the Cartan subalgebra and the a_j and a_{l+1} are real constants. Now write $2\alpha_j(z^a B_a) = u_j + iv_j$, $j = 1,2,\ldots, l = 2k$, so that

$$u_j = u_{l+1-j} = \begin{cases} 2z^1 - z^2, & j = 1, \\ -z^{j-1} + 2z^j - z^{j+1}, & 2 \leqslant j \leqslant k-1, \\ -z^{k-1} + z^k, & j = k; \end{cases}$$

and

$$v_j = -v_{l+1-j} = \begin{cases} 2z^{k+1} - z^{k+2}, & j = 1, \\ -z^{k+j-1} + 2z^{k+j} - z^{k+j+1}, & 2 \leqslant j \leqslant k-1, \\ -z^{2k-1} + 3z^{2k}, & j = k. \end{cases}$$

Also,

$$2(\alpha_1 + \alpha_2 + \cdots + \alpha_l)(z^a B_a) = 2z^1,$$

whence

$$e^{2z^a \operatorname{ad} B_a} A_0 = \sqrt{(l+1)}\left[\sum_{j=1}^{l=2k} a_j e^{u_j + iv_j} E_{\alpha_j} + a_{l+1} e^{-2z^1} E_{-\alpha_1 \cdots \alpha_l}\right],$$

and

$$A_0^* = \sqrt{(l+1)}\left[\sum_{j=1}^{l} a_{l+1-j} E_{-\alpha_j} + a_{l+1} E_{\alpha_1 + \alpha_2 + \cdots + \alpha_l}\right],$$

so that

$$\begin{aligned}
[e^{2z^a \operatorname{ad} B_a} A_0, A_0^*] &= (l+1)\left[\sum_{j=1}^{l} a_j a_{l+1-j} e^{u_j + iv_j} H_j - (a_{l+1})^2 e^{-2z^1} \sum_{j=1}^{l} H_j\right] \\
&= \sum_{j=1}^{k} a_j a_{l+1-j} e^{u_j}[(l+1)(H_j + H_{l+1-j})\cos v_j + i(l+1)(H_j - H_{l+1-j})\sin v_j] \\
&\quad - \left[\sum_{j=1}^{k}(l+1)(H_j + H_{l+1-j})\right](a_{l+1})^2 e^{-2z^1}.
\end{aligned}$$

And thus, by Eqs. (4.6), our nonlinear system of differential equations with nontrivial prolongation structure is

$$z^a_{,x^1x^2} = f^a, \quad a = 1,2,\ldots,l;$$

with

$$f^a = \begin{cases} \frac{1}{2} a_j a_{l+1-j} e^{u_j} \cos v_j - \frac{1}{2}(a_{l+1})^2 e^{-2z^1}, \\ \quad a = j, \quad 1 \leqslant j \leqslant k, \\ \frac{1}{2} a_j a_{l+1-j} e^{u_j} \sin v_j, \\ \quad a = k+j, \quad 1 \leqslant j \leqslant k \end{cases}$$

(with the u_j and v_j as above).

ACKNOWLEDGMENTS

The author wishes to acknowledge the help of Phil Harmsworth who provided many of the references and participated in many enlightening discussions.

[1] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[2] C. Rogers and W. F. Shadwick, "Bäcklund Transformations and Their Applications," in *Mathematics in Science and Engineering*, Vol. 161 (Academic, New York, 1982).
[3] H. D. Wahlquist and F. D. Estabrook, J. Math. Phys. **16**, 1 (1975).
[4] R. Dodd and A. Fordy, Proc. R. Soc. London Ser. A **385**, 389 (1983).
[5] N. Jacobson, *Lie Algebras* (Wiley–Interscience, New York, 1962).
[6] J. E. Humphreys, "Introduction to Lie Algebras and Representation Theory," in *Graduate Texts in Mathematics*, Vol. 9 (Springer, Berlin, 1972).
[7] Zh-X. Wan, "Lie Algebras," in *International Series in Pure and Applied Mathematics*, Vol. 104 (Pergamon, New York, 1975).
[8] I. G. Macdonald, in *Representation Theory of Lie Groups, London Mathematical Society Lecture Note Series*, Vol. 34 (Cambridge U. P., Cambridge, England, 1979).
[9] W. F. Shadwick, J. Math. Phys. **19**, 2312 (1978).
[10] R. Bullough and R. Dodd, Proc. R. Soc. London Ser. A **352**, 481 (1977).
[11] B. Kostant, in *Representation Theory of Lie Groups, London Mathematical Society Lecture Note Series*, Vol. 34 (Cambridge U. P., Cambridge, England, 1979).
[12] A. V. Mikhailov, M. A. Olshanetsky, and A. M. Perelomov, Commun. Math. Phys. **79**, 473 (1981).
[13] G. Wilson, Ergodic Theor. Dyn. Syst. **1**, 361 (1982).
[14] B. Kostant, Am. J. Math. **81**, 973 (1959).

Painless nonorthogonal expansions[a)]

Ingrid Daubechies[b)]
Theoretische Natuurkunde, Vrije Universiteit Brussel, Pleinlaan 2, B 1050 Brussels, Belgium
A. Grossmann
Centre de Physique Theorique II, CNRS-Luminy, Case 907, F 13288 Marseille Cedex 9, France
Y. Meyer
Centre de Mathématiques, Ecole Polytechnique, 91128 Palaiseau Cedex, France

(Received 18 September 1985; accepted for publication 11 December 1985)

In a Hilbert space $\mathscr{H}$, discrete families of vectors $\{h_j\}$ with the property that $f = \Sigma_j \langle h_j | f \rangle h_j$ for every f in $\mathscr{H}$ are considered. This expansion formula is obviously true if the family is an orthonormal basis of $\mathscr{H}$, but also can hold in situations where the h_j are not mutually orthogonal and are "overcomplete." The two classes of examples studied here are (i) appropriate sets of Weyl–Heisenberg coherent states, based on certain (non-Gaussian) fiducial vectors, and (ii) analogous families of affine coherent states. It is believed, that such "quasiorthogonal expansions" will be a useful tool in many areas of theoretical physics and applied mathematics.

I. INTRODUCTION

A classical procedure of applied mathematics is to store some incoming information, given by a function $f(x)$ (where x is a continuous variable, which may be, e.g., the time) as a discrete table of numbers $\langle g_j | f \rangle = \int dx\, g_j(x) f(x)$ rather than in its original (sampled) form. In order to have a mathematical framework for all this, we shall assume that the possible functions f are elements of a Hilbert space $\mathscr{H}$ [we take here $\mathscr{H} = L^2(\mathbf{R})$]; the functions g_j are also assumed to be elements of this Hilbert space.

One can, of course, choose the functions g_j so that the family $\{g_j\}$ ($j \in J$, J a denumerable set) is an orthonormal basis of $\mathscr{H}$. The decomposition of f into the g_j is then quite straightforward: one has

$$f = \sum_j \langle g_j | f \rangle g_j ,$$

where the series converges strongly. The requirement that the g_j be orthonormal leads, however, to some less desirable features. Let us illustrate these by means of two examples.

Take first $g_j(x) = p_j(x)\, w(x)^{1/2}$, where the p_j are orthonormal polynomials with respect to the weight function w. In this case local changes of the function f will affect the whole table of numbers $\langle g_j | f \rangle$ ($j \in J$), which is a feature we would like to avoid.

An orthonormal basis $\{g_j\}$ ($j \in J$), which would enable us to keep nonlocality under control, is given by our second example. We cut $\mathbf{R}$ (the set of real numbers, which is the range of the continuous variable x) into disjoint intervals of equal length, and we construct the g_j starting from an orthonormal basis for one interval. Schematically, consider h_n, an orthonormal basis of $L^2([0,a))$,

$$J = \{(n,m);\ n,m \in \mathbf{Z}, \text{ the set of integers}\} ,$$

$$g_{n,m}(x) = \begin{cases} h_n(x - ma), & \text{for } ma < x < (m+1)a, \\ 0, & \text{otherwise.} \end{cases}$$

If now the function f undergoes a local change, confined to the interval $[ka, la]$, only the numbers $\langle g_{n,m} | f \rangle$ with $k < m < l - 1$ will be affected, reflecting the locality of the change. This choice for the g_j also has, however, its drawbacks: some of the functions g_j are likely to be discontinuous at the edges of the intervals, thereby introducing discontinuities in the analysis of f, which need not have been present in f itself. This is particularly noticeable if one takes the following natural choice for the h_n:

$$h_n(x) = a^{-1/2} e^{i2\pi x/a} .$$

In this case even very smooth functions f will give values $g_{n,m}(f)$ significantly different from zero for rather high values of n, reflecting high-frequency components artificially introduced by the cutting of $\mathbf{R}$ into intervals.

We shall now see how these undesirable features can be avoided by taking radically different options for the choice of the g_j. In particular, we shall not restrict ourselves to orthonormal bases. Let us start by asking which properties we want to require for the g_j.

The storage of the function f in the form of a discrete table of numbers $\langle g_j | f \rangle$ ($j \in J$) only makes sense if one is certain that f is completely characterized by the numbers $\langle g_j | f \rangle$ ($j \in J$). In other words, we want

$$\langle g_j | f \rangle = \langle g_j | h \rangle, \quad \text{for all } j \text{ in } J ,$$

to imply $f = h$, which is equivalent to saying that the vectors $\{g_j\}$ ($j \in J$) span a dense set, i.e., that the orthonormal complement $\{g_j; j \in J\} = \{0\}$. This will be our first requirement.

In all the cases we shall discuss, the set $\{g_j\}$ ($j \in J$) is such that the map

$$T:\ f \to (\langle g_j | f \rangle)_{j \in J}$$

defines a bounded operator from $\mathscr{H}$ to $l^2(J)$, the Hilbert space of all square integrable sequences labeled by J. In other

[a)] This work was carried out within the framework of the R. C. P. "Ondelettes."
[b)] "Bevoegdverklaard Navorser" at the National Foundation for Scientific Research, Belgium.

 0022-2488/86/051271-13$02.50

words, J is countable, and there exists a positive number B such that for all f in $\mathscr{H}$ one has

$$\sum_{j\in J}|\langle g_j|f\rangle|^2\leqslant B\,\|f\|^2 .$$

This can also be stated in the following, equivalent, form: If $|g_j\rangle\,\langle g_j|$ is defined as the operator associating to every vector h in $\mathscr{H}$ the vector $\langle g_j|h\rangle g_j$, then

$$\sum_{j\in J}|g_j\rangle\,\langle g_j|\in\mathscr{B}(\mathscr{H})$$

(the set of all bounded operators in $\mathscr{H}$), with

$$\Big\|\sum_{j\in J}|g_j\rangle\,\langle g_j|\Big\|\leqslant B .$$

In order to reconstruct f from the discrete table $(\langle g_j|f\rangle)_{j\in J}$, one needs to invert the map T

$$T:\ f\rightarrow(\langle g_j|f\rangle)_{j\in J}$$

from $\mathscr{H}$ to $l^2=l^2(J)$.

In general the image $T\mathscr{H}$ is not all of l^2, but only a subspace of l^2; one can see this, for instance, if the g_j constitute what is often called, in the physics literature, an "overcomplete" set, i.e., if each g_j is in the closed linear span of the remaining ones: $\{g_k;\ k\in J,\ k\neq j\}$. Strictly speaking, there is then no inverse map T^{-1}. This is, of course, no real difficulty: One can define a map $\widetilde{T}$ from $l^2(J)$ to $\mathscr{H}$, which is zero on the orthogonal complement of $T\mathscr{H}$ and which inverts T when restricted to $T\mathscr{H}$.

If the spectrum of the positive operator $\Sigma_{j\in J}|g_j\rangle\,\langle g_j|$ reaches down to zero, this inverse is an unbounded operator, and the recovery of f from $(\langle g_j|f\rangle)_{j\in J}$ becomes an ill-posed problem. This is avoided if we require the spectrum of $\Sigma_{j\in J}|g_j\rangle\,\langle g_j|$ to be bounded away from zero, i.e., if we impose that there exist positive constants A,B such that

$$A\mathbf{1}\leqslant\sum_{j\in J}|g_j\rangle\,\langle g_j|\leqslant B\mathbf{1} .$$

(Here $\mathbf{1}$ is the identity operator in $\mathscr{H}$. The inequality sign $\leqslant$ between two operators L and T means that their difference $T-L$ is positive definite.) Equivalently, for all f in $\mathscr{H}$, we require

$$A\,\|f\|^2\leqslant\sum_{j\in J}|\langle g_j|f\rangle|^2\leqslant B\,\|f\|^2 . \tag{1.1}$$

This is the second condition we impose on the set $\{g_j\}$ ($j\in J$). The only new condition is the lower bound.

A set of vectors $\{g_j\}$ ($j\in J$) in a Hilbert space $\mathscr{H}$, satisfying condition (1.1) with $A,B>0$, is called a *frame*.[1] Note that in general the vectors $\{g_j\}_{j\in J}$ will not be a basis in the technical sense, even though their closed linear span is all of $\mathscr{H}$. This is so because the vectors g_j need not be "ω independent," even though they will usually be linearly independent. That is, a vector g_j usually cannot be written as a finite linear combination of vectors $g_{j'}$ (with $j'\neq j$) but it may well belong to the closed linear span of the infinitely many remaining members of the family. Frames were introduced in the context of nonharmonic Fourier series, where the functions g_j are exponentials.[1,2] As far as we know, this is the only context in which frames have been put to use. One of the aims of the present paper is to provide examples of frames in other contexts. Notice that the results on frames in connection with nonharmonic Fourier series can be rewritten as estimates for entire functions in the Paley–Wiener space[1,2]; one of the results we shall derive here can be rewritten as an analogous estimate for entire functions of growth less than $(2,\frac{1}{2})$ (see Ref. 3).

Notice that, even for functions g_j satisfying the condition (1.1), the effective inversion of the map T: $f\rightarrow(\langle g_j|f\rangle)_j$ may be a complicated matter. The condition (1.1) on the g_j ensures that the operator is $\widetilde{T}$ is bounded ($\|\widetilde{T}\|\leqslant A^{-1/2}$) but does not provide a way of calculating it. We are still left with a problem where we have to invert large matrices, although some convergence questions are under control. Assuming for a moment that $\widetilde{T}$ is given, we may define the family $e_k=\widetilde{T}d_k$, where the d_k ($k\in J$) form the natural orthonormal basis of $l^2(J)$. For $c=(c_j)_{j\in J}\in l^2$, the image $\widetilde{T}c$ is then given by

$$\widetilde{T}c=\sum_j c_j e_j ,$$

where the series converges strongly, by the boundedness of $\widetilde{T}$. This then implies, for all f in $\mathscr{H}$,

$$f=\sum_j\langle g_j|f\rangle e_j , \tag{1.2}$$

again with strong convergence of the series. While (1.2) looks identical to the familiar expansion of f into biorthogonal bases, it really is very different because the $(g_j)_{j\in J}$ need not be a basis at all, technically speaking.

There exists, however, a particular class of frames for which these computational problems do not arise. These are the frames for which the ratio B/A reaches its "optimal" value, $B/A=1$. One has then, for all f in $\mathscr{H}$,

$$\sum_{j\in J}|\langle g_j|f\rangle|^2=A\,\|f\|^2 \tag{1.3}$$

or, equivalently,

$$\sum_{j\in J}|g_j\rangle\,\langle g_j|=A\mathbf{1} .$$

So the map T is now a multiple of an isometry from $\mathscr{H}$ into l^2; as such, it is inverted, on its range, by a multiple of its adjoint T^*. Moreover TT^* is a multiple of the orthogonal projection operator on the range T, which can be thus easily characterized.

It is evident that (1.3) is satisfied whenever the g_j constitute an g_j constitute an orthonormal basis (with $A=1$ then). We shall see that there are other, more interesting examples of frames satisfying (1.3), in which the vectors g_j are not mutually orthogonal, and where the set $\{g_j\}$ ($j\in J$) is "overcomplete" in the sense defined above. We shall say that a frame is *tight* if it satisfies condition (1.3) or, equivalently, if the inequalities in (1.1) can be tightened into equalities. The inversion formula allowing one to recover the vector f from $(\langle g_j|f\rangle)_{j\in J}$ is particularly simple for tight frames. For any f in $\mathscr{H}$ one has

$$f=A^{-1}\sum_{j\in J}\langle g_j|f\rangle g_j , \tag{1.4}$$

where the series converges strongly (as in the case of a general frame). The expansion (1.4) is thus entirely analogous to an expansion with respect to an orthonormal basis, even though the g_j need not be orthogonal. We believe that tight frames and the associated simple (painless!) quasiorthogonal expansions will turn out to be very useful in various questions of signal analysis, and in other domains of applied mathematics. Closely related expansions have already been used in the analysis of seismic signals.[4]

The vectors g_j constituting a tight frame need not be normalized. On the other hand, an orthogonal basis consisting of vectors of different norm, does not constitute a tight frame.

In real life, of course, one will have to deal with finite sets of vectors g_j, i.e., one will have to truncate the infinite set J to a finite subset. The reconstruction problem then becomes ill-posed, and extra conditions, using *a priori* information on f, will be needed to stabilize the reconstruction procedure.[5] We shall not address this question here.

In this paper, we shall discuss two classes of examples of sets $\{g_j\}$ ($j \in J$). In both cases, this discrete set of vectors is obtained as a discrete subset of a continuous family which forms an orbit of a unitary representation of a particular group. Schematically, such families can be described as follows. Consider the following.[6]

(i) $U(\cdot)$ is an irreducible unitary representation, on $\mathscr{H}$, of a locally compact group $\mathscr{G}$.

(ii) $d\mu(\cdot)$ is the left-invariant measure on $\mathscr{G}$.

(iii) Let g be an admissible vector in $\mathscr{H}$ for U (see Ref. 6), i.e., a nonzero vector such that

$$c_g = \|g\|^{-2} \int d\mu(y)\,|\langle g, U(y)g\rangle|^2 < \infty\,, \tag{1.5}$$

the integral being taken over $\mathscr{G}$.

(Notice that there are many irreducible unitary representations for which no admissible vectors exist. However, if there is one admissible vector, there is a dense set of them, and we call the representation square integrable.[6])

(iv) Then

$$\int d\mu(y)\, U(y)\,|g\rangle\,\langle g|\,U(y)^* = c_g \mathbf{1}\,, \tag{1.6}$$

where the integral is to be understood in the weak sense. If the group $\mathscr{G}$ is unimodular [i.e., if $d\mu(\cdot)$ is both left and right invariant], the existence of one admissible vector in $\mathscr{H}$ implies that all vectors in $\mathscr{H}$ are admissible; moreover, one has in this case that $c_g = c\|g\|^2$ for some c independent of g [see Ref. 6(b)].

(v) In order to obtain possible sets $\{g_j\}$ ($j \in J$) we choose (1) an admissible vector g in $\mathscr{H}$ and (2) a "lattice" of discrete values for the group element y: $\{y_j;\ j \in J\}$.

The vectors g_j are then defined as

$$g_j = U(y_j)g\,.$$

By imposing appropriate restrictions on g and on J, we shall obtain families $\{g_j\}$ that are frames—or tight frames—in $\mathscr{H}$.

With this procedure it is possible to adjust the "spacing" of the "lattice" $\{y_j;\ j \in J\}$ according to the desired degree of "oversampling." In the two cases that we shall consider, this flexibility can be exploited at little computational cost, since the action of $U(y)$ on g is very simple and the new g_j—obtained after an adjustment of the "lattice"—can be easily and quickly calculated.

In this paper, we shall discuss sets $\{g_j\}$ ($j \in J$) constructed along the lines described above for two different groups; the Weyl–Heisenberg group, and the affine or $ax+b$ group.

In Sec. II we treat the Weyl–Heisenberg case. We start, in Sec. II A, by giving a short review of the definition and main properties of this group and of the associated "overcomplete" set, generally called the set of coherent states. A particular discrete set of coherent states is associated to the so-called von Neumann lattice and to a particular choice of g; it has been discussed and used many times (see, e.g., Refs. 7 and 8). It is well known that the set of coherent states associated to the von Neumann lattice is complete, i.e., that its linear span is dense in $\mathscr{H}$ (see Refs. 8–10). It thus meets the first of the two requirements listed above. We show in Sec. II B that the second requirement is not met: the coherent states associated to the von Neumann lattice do *not* constitute a frame. In Sec. II C we shall see that a similar lattice, with density twice as high, does lead to a frame. In II D we concentrate on analogous families of states based on function g with compact support, as opposed to the most commonly discussed canonical coherent states, where g is a Gaussian. We derive sufficient conditions ensuring that the $g_j = U(y_j)g$ constitute a frame. In Sec. II E we show how g can be chosen in such a way that the frame generated is tight. In Sec. II F we analyze this situation and describe in more detail the necessary and sufficient conditions that g has to satisfy in order to generate a tight frame.

In Sec. III we discuss the $ax+b$ group. Again we start, in Sec. III A, with a short review of definitions and properties, including the so-called affine coherent states. The affine coherent states were first defined in Ref. 11; detailed studies of them can be found, e.g., in Refs. 4 and 12; for applications of these states to signal analysis, see Ref. 4. In Sec. III B we discuss discrete "lattices" of affine coherent states based on "band-limited" functions g, i.e., on functions such that the Fourier transform of g has compact support. We derive sufficient conditions for these discrete sets to be frames. In Sec. III C we show how certain specific choices of g lead to tight frames; in Sec. III D we again analyze the construction, and derive necessary and sufficient conditions on g, ensuring that certain frames will be tight.

As can be readily seen from Sec. II E and Sec. III C, the construction of tight frames associated with the Weyl–Heisenberg group is essentially the same as that of tight frames associated with the $ax+b$ group. Tight frames associated with the $ax+b$ group were first introduced[13] in a different context closer to pure mathematics. In Ref. 13(b) one can find a definition of "quasiorthogonal families" very close to our tight frames, and a short discussion of the similarities between a "quasiorthogonal family" and an orthonormal basis. For the many miraculous properties of this orthonormal basis, see Ref. 14.

Finally, let us note that while we have restricted our discussion to $\mathscr{H} = L^2(\mathbf{R})$, it is possible to extend the discussion to $L^2(\mathbf{R}^n)$, as well for the Weyl–Heisenberg group as for the $ax+b$ group. In the latter case the unitary representation $U(\cdot)$ underlying the construction of frames, is no longer irreducible. A more detailed analysis shows, however, that the essential feature is cyclicity of the representation rather than its irreducibility.[15]

II. THE WEYL–HEISENBERG CASE

A. Review of definitions and basic properties

The Weyl–Heisenberg group is the set $\mathbb{T}\times\mathbb{R}\times\mathbb{R}$ (where $\mathbb{T}$ is the set of complex numbers of modulus 1), with the group multiplication law

$$(z,q,p)(z',q',p') = (e^{i(pq'-p'q)/2}zz',q+q',p+p') .$$

We shall here be concerned with the irreducible unitary representation of this group acting in the Hilbert space $\mathscr{H} = L^2(\mathbb{R},dx)$, and given by

$$(W(z,q,p)f)(x) = ze^{-ipq/2}e^{ipx}f(x-q) .$$

The *Weyl operators* $W(q,p)$ are defined as

$$W(q,p) = W(1,q,p);$$

they satisfy the relations

$$W(q,p)W(q',p') = \exp[i(pq'-p'q)/2]W(q+q',p+p'),$$

an exponentiated form of the Heisenberg commutation relations. By a theorem of von Neumann, the above relations determine the irreducible family W up to unitary equivalence. A well-known property[16] of Weyl operators is the following; for all f_1,f_2,g_1,g_2, in $\mathscr{H}$ one has

$$\int\int dpdq\langle f_1,W(q,p)g_1\rangle\langle W(q,p)g_2,f_2\rangle = 2\pi\langle f_1|f_2\rangle\,\langle g_2|g_1\rangle. \qquad (2.1)$$

Comparing this with (1.5) and (1.6), one sees that all elements of $\mathscr{H}$ are admissible and that $c_g = 2\pi\|g\|^2$. These two features are a consequence of the unimodularity of the Weyl–Heisenberg group.

The family of canonical coherent states is defined as a particular orbit under this set of unitary operators. The canonical coherent states can be defined as the family of vectors $W(q,p)\Omega$, where Ω is the ground state of the harmonic oscillator:

$$\Omega(x) = \pi^{-1/4}\exp(-x^2/2) .$$

One readily sees that this is equivalent to the customary definition of a canonical coherent state as the function

$$\pi^{-1/4}e^{-ipq/2}e^{ipx}\exp(-(x-q)^2/2) .$$

We shall often work with orbits of Weyl operators other than the canonical coherent states. We therefore introduce the notation

$$|q,p;g\rangle = W(q,p)g , \qquad (2.2)$$

where g is any nonzero element of $\mathscr{H}$. The canonical coherent states are thus $|p,q;\Omega\rangle$. As a consequence of (2.1), one has

$$\int\int dqdp|q,p;g\rangle\,\langle q,p;g| = 2\pi\|g\|^2\mathbf{1} .$$

B. The von Neumann lattice and Zak transform

Take $a,b>0$. For any integer m,n, consider

$$|ma,nb;\Omega\rangle = W(ma,nb)\Omega . \qquad (2.3)$$

It is known[8–10] that the linear span of the set $\{|ma,nb;\Omega\rangle;\ m,n \text{ in } \mathbb{Z}\}$ is dense in $\mathscr{H}$ if and only if $ab\leqslant 2\pi$. At the critical density $ab=2\pi$, this set of points $\{(ma,nb)\}$ in phase space is called a von Neumann lattice.[7] In quantum mechanics, the associated set of canonical coherent states has a nice physical interpretation. It corresponds to choosing exactly one state per "semiclassical Gibbs cell," i.e., per cell of area h (Planck's constant).

Notice that the discrete set of Weyl operators $\{W(ma,n2\pi/a);\ m,n\in\mathbb{Z}\}$ is Abelian. This feature is exploited in the construction of the kq transform, or Zak transform,[17] which will turn out to be useful in what follows.

Denote by $\square$ the semiopen rectangle $\square=[-\pi/a,\pi/a)\times[-a/2,a/2)$.

The Zak transform is a unitary map from $L^2(\mathbb{R})$ onto $L^2([-\pi/a,\pi/a)\times[-a/2,a/2)) = L^2(\square)$ and is defined as follows. For a function f in $C_0^\infty(\mathbb{R})$ (infinitely differentiable functions with compact support), one defines its Zak transform Uf by

$$(Uf)(k,q) = \left(\frac{a}{2\pi}\right)^{1/2}\sum_l e^{ikal}f(q-la) , \qquad (2.4)$$

where, for any q, only a finite number of terms in the sum contribute, due to the compactness of the support of f. The map U, defined by (2.4), is isometric from $C_0^\infty(\mathbb{R})\subset L^2(\mathbb{R})$ to $L^2(\square)$; there exists therefore an extension, which we shall also denote by U, to all of $L^2(\mathbb{R})$. It turns out that this extension maps $L^2(\mathbb{R})$ onto all of $L^2(\square)$; this is the Zak transform.

We ask now whether the family (2.3) constitutes a frame, i.e., whether the spectrum of the positive operator

$$P=\sum_m\sum_n|ma,n2\pi/a;\Omega\rangle\,\langle ma,n2\pi/a;\Omega|$$

is bounded away from zero. We shall see that the answer to this question is straightforward for the unitarily equivalent operator UPU^{-1}. The same technique was used in Ref. 10(a) to investigate the question whether the linear span of (2.3) and of similar families is dense.

An easy calculation leads to

$$[UW(ma,n2\pi/a)f](k,q) = e^{-ikma}e^{iqn2\pi/a}(Uf)(k,q) . \qquad (2.5)$$

Hence, for $f\,L^2(\square)$, we have

$$\langle f,UPU^{-1}f\rangle = \sum_m\sum_n\left|\int\int_\square dk\,dq\,e^{ikma}e^{iqn2\pi/a}(U\Omega)(k,q)f(k,q)\right|^2 = \int dk\int dq|U\Omega(k,q)f(k,q)|^2 ,$$

where we have used the basic unitary property of Fourier series expansions. This shows that the operator P is unitarily equivalent to multiplication by $|U\Omega(k,q)|^2$ in $L^2(\square)$. The spectrum of P is therefore exactly the numerical range of the function $|U\Omega(k,q)|^2$. The function $U\Omega$ is given by

$$U\Omega(k,q) = [a\pi^{-3/2}/2]^{1/2}\sum_l \exp[ikla - (q-la)^2/2]$$
$$= [a\pi^{-3/2}/2]^{1/2}\exp(-q^2/2)$$
$$\times\theta_3(a(k-iq)/2, \exp(-a^2/2)),$$

where θ_3 is one of Jacobi's theta functions[18]

$$\theta_3(z,u) = 1 + 2\sum_l u^{l^2}\cos(2lz).$$

The zeros of $U\Omega$ are therefore completely determined by the zeros of θ_3; one finds that the function $U\Omega$ has zeros at the corner of the semiopen rectangle $[-\pi/a,\pi/a) \times [-a/2,a/2)$, and nowhere else. (The fact that $U\Omega$ is zero at the corner can also be seen easily from its series expansion.) This is enough, however, to ensure that the spectrum of the multiplication operator by $|U\Omega(k,q)|^2$, and therefore also of P, contains zero. Therefore the family (2.3) associated to the von Neumann lattice is *not* a frame.

C. A frame of canonical coherent states

Since the family (2.3) with $ab = 2\pi$ is not a frame, it is clear that we have to look at lattices with higher density, i.e., with $ab < 2\pi$. [If $ab > 2\pi$, the linear span of the vectors (2.3) is not even dense.] The construction above, which uses the Zak transform, will not work for arbitrary a and b; if $b \neq 2\pi/a$, then Eq. (2.5) will no longer be true in general. This is due to the fact that in general the operators $W(ma,nb)$ do not mutually commute. It is, however, possible to use again the same construction in the case where the density is an integral multiple of the density for the von Neumann lattice. For the sake of definiteness, we shall consider the case where $ab = \pi$.

We now have to study the operator

$$P = \sum_m\sum_n \left| ma,\frac{n\pi}{a};\Omega\right\rangle \left\langle ma,\frac{n\pi}{a};\Omega\right| .$$

For $n = 2l$, it is clear from (2.3) that

$$|m,2l;\ a,\pi/a\rangle = W(ma,12\pi/a)\Omega = |ma,2\pi l/a;\Omega\rangle .$$

On the other hand,

$$|m,2l+1;\ a,\pi/a\rangle = e^{imab/4}W(ma,2\pi l/a)W(0,\pi/a)\Omega$$
$$= e^{imab/4}|ma,2\pi l/a;\ W(0,\pi/a)\Omega\rangle .$$

Hence

$$P = \sum_m\sum_n \left| ma,\frac{2\pi n}{a};\Omega\right\rangle\left\langle ma,\frac{2\pi n}{a};\Omega\right|$$
$$+ \sum_m\sum_n \left| ma,\frac{2\pi n}{a};\ W\left(0,\frac{\pi}{a}\right)\Omega\right\rangle$$
$$\times\left\langle ma,\frac{2\pi n}{a};W\left(0,\frac{\pi}{a}\right)\Omega\right| .$$

Using (2.5) again, we then see that

$$\langle f,UPU^{-1}f\rangle = \int dk\int dq|f(k,q)|^2(|U\Omega(k,q)|^2$$
$$+ |[UW(0,\pi/a)\Omega](k,q)|^2),$$

where U is again the Zak transform as defined above, in Sec. II B. A calculation of $UW(0,\pi/a)\Omega$ gives

$$[UW(0,\pi/a)\Omega](k,q)$$
$$= 2^{-1/2}\pi^{-3/4}a^{1/2}e^{i\pi q/a}\exp(-q^2/2)$$
$$\times\theta_3[(ak-aiq-\pi)/2, \exp(-a^2/2)],$$

hence

$$|U\Omega(k,q|^2 + |[UW(0,\pi/a)\Omega](k,q)|^2$$
$$= 2^{-1}\pi^{-3/2}\exp(-q^2)a$$
$$\times\{|\theta_3[a(k-iq)/2, \exp(-a^2/1)]|^2$$
$$+ \theta_3[(ak-iqa-\pi)/2, \exp(-a^2/2)]|^2\} .$$

This function is continuous and has no zeros, since the zeros of $\theta_3[u,\exp(-a^2/2)]$ occur only at $u = \pi(m+\frac{1}{2}+ia^2(n+\frac{1}{2}))$. There exist therefore $A,B > 0$ such that

$$A \leqslant |U\Omega(k,q)|^2 + |[UW(0,\pi/a)\Omega](k,q)|^2 \leqslant B;$$

this implies that the set of canonical coherent states $\{|ma,n\pi/a;\ \Omega\rangle\}$ $(m,n\in\mathbb{Z})$ is a frame, with

$$A \leqslant \sum_m\sum_n \left| ma,\frac{n\pi}{a};\Omega\right\rangle\left\langle ma,\frac{n\pi}{a};\Omega\right| \leqslant B .$$

A numerical estimate of A and B gives, in the case $a = 2$,

$A \geqslant 1.60$,

$B \leqslant 2.43$.

Remark: The above analysis also works if the density of the chosen lattice is another, higher multiple of the critical von Neumann density, i.e., for $ab = 2\pi/n$, where $n = 3,4,\ldots$. The ratio B/A of the upper and lower bound of the frame is clearly a decreasing function of n.

D. Lattices with analyzing wavelets of compact support

We shall now consider families of the type $|na,mb;h\rangle$, where $h(x)$ is a function of compact support.

As an example, consider first the case where $h(x)$ is the characteristic function of an interval $[-L/2,L/2]$, i.e., $h(x) = 1$ if x belongs to this interval, and is zero otherwise. It is then easy to see that, with the choice $a = L$ and $b = 2\pi L$ (hence again $ab = 2\pi$), the family $\{|ma,nb;h\rangle\}$ $(m,n\in\mathbb{Z})$ consisting of the functions $\exp[2\pi inx/L]h(x)$ is an orthonormal basis of $L^2(\mathbb{R})$ and therefore certainly a frame.

For reasons explained in the Introduction, however, we prefer to work with smoother functions h. We shall see that under fairly general conditions, a lattice based on continuous functions of compact support also gives rise to a frame. The price to be paid is a higher density of the lattice; furthermore, the frame will not be tight in general.

Theorem 1: Let $h(x)$ be a continuous function on $\mathbb{R}$, with support in the interval $[-L/2,L/2]$. Assume that $h(x)$ is bounded away from zero in a subinterval $[-\mu L/2,\mu L/2]$ $(0<\mu<1)$:

$$|h(x)| \geqslant k, \quad \text{if}|x| \leqslant \mu L/2 \quad (\mu < 1) .$$

Define now a lattice in phase space by taking $a = \mu L$ and $b = 2\pi/L$ (hence $ab = 2\pi\mu$, but the "oversampling parameter" μ^{-1} need not be an integer, contrary to Sec. II C). Consider the set of states

$$\{|ma,nb;h\rangle\} \quad (m,n\in\mathbb{Z});$$

then this set is a frame, with

$$A \geq L \inf_{|x| < \mu L/2} |h(x)|^2 \geq k$$

and

$$B \leq L \sup_{x \in \mathbf{R}} \left[\sum_m |h(x + m\mu)|^2 \right]$$
$$\leq L(1 + 2[\mu^{-1}])(\|h\|_\infty)^2 ,$$

where $[\mu^{-1}]$ is the largest integer not exceeding μ^{-1}.

Proof: For typographical convenience, write $\Delta = [-L/2, L/2]$. Let f be any element of $L^2(\mathbf{R})$. Then

$$\langle ma, nb; h \mid f \rangle = e^{-imn\mu\pi} \int_\Delta dx\, h(x) \times e^{-2i\pi n x/L} f(x + m\mu L) .$$

Hence, by considering the above integral as $L^{1/2}$ times the nth Fourier coefficient of the function $h(x)f(x + m\mu L)$ defined on the interval Δ,

$$\sum_n |\langle ma, nb; h \mid f \rangle|^2 = L \int_\Delta dx |h(x)|^2 |f(x + \mu m L)|^2$$
$$\geq k^2 L \int_{\mu\Delta} dx |f(x + \mu m L)|^2 .$$

This implies

$$\sum_m \sum_n |\langle ma, nb; h \mid f \rangle|^2 \geq k |L \int dx |f(x)|^2 .$$

On the other hand, we clearly have

$$\sum_m \sum_n |\langle ma, nb; h \mid f \rangle|^2$$
$$= L \int_\Delta dx \left(\sum_m |h(x + m\mu L)|^2\right) |f(x)|^2$$
$$\leq bL \int dx |f(x)|^2 ,$$

with

$$b = \sup_{x \in \mathbf{R}} \left[\sum_m |h(x + m\mu L)|^2 \right]$$
$$\leq (2[\mu^{-1}] + 1)(\|h\|_\infty)^2 ,$$

and so our assertions are proved.

E. Tight frames with analyzing wavelets of compact support

We keep the assumptions and notations of Sec. II D. The arguments of that subsection show that

$$\sum_n \left| \left\langle m\mu L, \frac{2\pi n}{L}; h \mid f \right\rangle \right|^2$$
$$= L \int_\Delta dx |h(x)|^2 |f(x + \mu m L)|^2$$
$$= L \int dx |h(x - \mu m L)|^2 |f(x)|^2 .$$

Consequently we have

$$\sum_n \sum_m \left| \left\langle m\mu L, \frac{2\pi n}{L}; h \middle| f \right\rangle \right|^2$$
$$= L \int dx |f(x)|^2 \left[\sum_m |h(x + \mu m L)|^2 \right] ,$$

and we obtain the following result.

Theorem 2: Let $h(x)$ be continuous on $\mathbf{R}$, with support in $[-L/2, L/2]$, and bounded away from zero on $[-\mu L/2, \mu L/2]$, where $0 < \mu < 1$. Assume furthermore that the function $\Sigma_m |h(x + \mu m L)|^2$ is a constant, i.e., independent of x. Then the family $\{|m\mu L, 2\pi n/L; h\rangle\}$ $(m, n \in \mathbf{Z})$ is a tight frame.

Remark: By the assumptions on h, the sum $\Sigma_m |h(x + \mu m L)|^2$ has only finitely many nonzero terms, and defines a continuous function of x.

We shall now give a procedure for constructing functions h that are k times continuously differentiable and satisfy the condition in Theorem 2:

$$\sum_m |h(x + m\mu L)|^2 = \text{const.} \tag{2.6}$$

Here k may be any positive integer or even ∞. We start by choosing a function g that is $2k$ times continuously differentiable and such that $g(x) = 0$ for $x \leq 0$, and $g(x) = 1$ for $x \geq 1$. Assume in addition that g is everywhere increasing.

For the sake of simplicity, we shall now assume that $\mu \geq \frac{1}{2}$. We then define h as follows:

$$h(x) = \begin{cases} 0, & \text{for } x \leq -L/2, \\ \{g[(x/L + 1/2)/(1-\mu)]\}^{1/2}, & \text{for } -L/2 \leq x \leq -L(2\mu - 1)/2, \\ 1, & \text{for } -L(2\mu - 1)/2 \leq x \leq L(2\mu - 1)/2, \\ \{1 - g([x/L - (2\mu - 1)/2]/(1-\mu))\}^{1/2}, & \text{for } L(2\mu - 1)/2 \leq x \leq L/2, \\ 0, & \text{for } x \geq L/2. \end{cases}$$

The function $h(x)$ defined in this way is non-negative, with support $[-L/2, L/2]$, and equal to 1 on $[-(2\mu - 1)L/2, (2\mu - 1)L/2]$. Since g is a C^{2k} function, one sees that h is indeed a C^k function. The points $x = \pm(2\mu - 1)L/2$, where h becomes constant, have been chosen so that their distance to the furthest edge of supp(f) is exactly μL. It is now easy to check that h fulfills the condition (2.6): for $|x| \leq (2\mu - 1)L/2$, one has

$$\sum_m |h(x + m\mu L)|^2 = |h(x)|^2 = 1 ;$$

and for x in $[(2\mu - 1)L/2, L/2]$, one has

$$\sum_m |h(x + m\mu L)|^2$$
$$= |h(x)|^2 + |h(x - \mu L)|^2$$
$$= 1 - g((x/L - (2\mu - 1)/2)/(1-\mu))$$
$$+ g(((x - \mu L)/L + 1/2)/(1-\mu)) = 1 .$$

For x outside $[-(2\mu - 1)L/2, L/2]$ the result follows by simple translation. Hence

$$\sum_m |h(x + m\mu L)|^2 = 1 ,$$

which implies

$$\sum_m \sum_n \left| m\mu L, \frac{n2\pi}{L};\ h \right\rangle \left\langle m\mu L, \frac{n2\pi}{L};\ h \right| = L\mathbf{1},$$

and we have constructed a tight frame!

The above construction may be clarified by the following easy example.

Example: We define a function h satisfying the condition (2.6) as follows:

$$h(x) = \begin{cases} 0, & \text{if} |x| > \pi/2, \\ \cos x, & \text{if} |x| < \pi/2. \end{cases}$$

Hence $L = \pi$. We take $\mu = \frac{1}{2}$. Then (see also Fig. 1), with χ the characteristic function of the interval $[-\pi/2, \pi/2]$, one has

$$\sum_m |h(x + \mu m L)|^2 = \sum_m \cos^2\left(x + \frac{m\pi}{2}\right) \chi\left(x + \frac{m\pi}{2}\right)$$
$$= \cos^2 x + \sin^2 x = 1 .$$

In this example, the corresponding function g is the function

$$g(x) = \begin{cases} 0, & \text{if } x > 0, \\ \sin^2(\pi x/2), & \text{if } 0 < x < 1, \\ 1, & \text{if } x > 1. \end{cases}$$

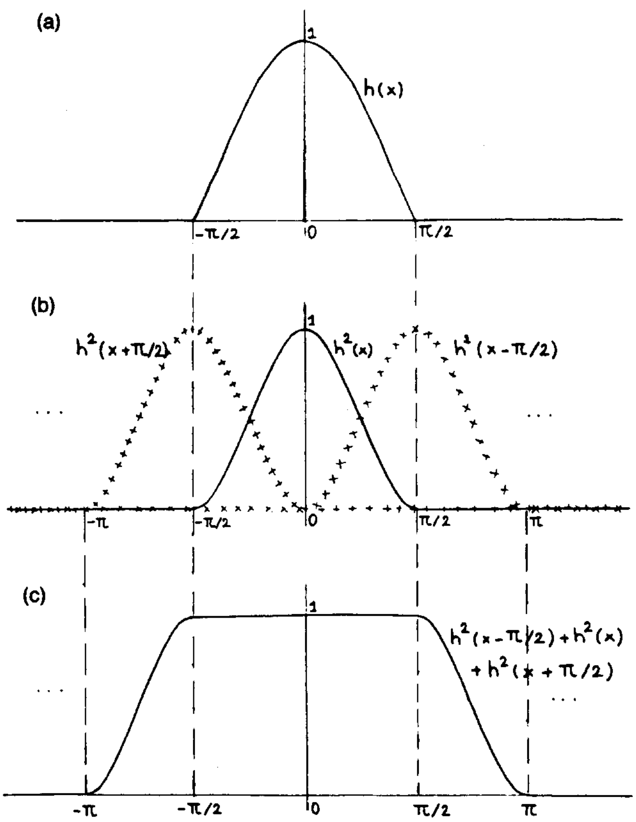

FIG. 1. (a) The function $h(x) = \cos x\, \chi_{[-\pi/2, \pi/2]}(x)$. (b) $h^2(x)$ and two translated copies, $h^2(x + \pi/2)$ and $h^2(x - \pi/2)$. (c) The sum $h^2(x - \pi/2) + h^2(x) + h^2(x + \pi/2)$ is equal to 1, for $-\pi/2 < x < \pi/2$. Analogously $\Sigma_{l=-N}^{N} h^2(x + l\pi/2) = 1$, for $-N\pi/2 < x < N\pi/2$, and $\Sigma_{l\in\mathbb{Z}} h^2(x + l\pi/2) = 1$, for all x.

Remark: In our construction, we have assumed that $\mu > \frac{1}{2}$. For smaller values of μ, a similar but more complicated construction can be made (see Appendix).

F. A closer look at condition (2.5)

Let h be a function continuous on $\mathbb{R}$, vanishing on the set $\mathbb{R} \setminus [-L/2, L/2]$ and nowhere else. The discussion of the preceding sections shows that the family of functions

$$h_{mn}(x) = e^{2i\pi n x/L} h(x + \mu m L) \quad (0 < \mu < 1; n, m \in \mathbb{Z})$$

is a tight frame if and only if (2.6) holds, i.e., one has

$$\sum_n f(x + na) = \text{const}, \tag{2.7}$$

with $a = \mu L$ and $f(x) = |h(x)|^2$.

In this subsection we shall study the class of functions that satisfy (2.7); at first, we shall not require f to be positive or to have compact support (as opposed to the assumptions on f in the preceding subsections). However, we need to impose some assumptions on f in order to ensure that the left-hand side of (2.7) is well defined. It will be convenient to work with the space $\mathscr{C}$ defined as follows.

Definition:

$\mathscr{C} = \{f\colon \mathbb{R} \to \mathbb{C};\ f$ is measurable,

and there exists a $C > 0$ and a $K > 1$,

such that $|f(x)| < C(1 + |x|)^{-K}\}$.

It is clear that, for f in $\mathscr{C}$, the series $\Sigma_n f(x + na)$ is absolutely convergent, uniformly on the interval $[-a/2, a/2]$. We shall now derive a necessary and sufficient condition for elements of $\mathscr{C}$ to satisfy (2.7). Take f in $\mathscr{C}$. Denote by Δ the interval $[-a/2, a/2]$. Define, for q in Δ,

$$F(q) = \sum_n f(q + na) .$$

F is bounded, and hence belongs to $L^2(\Delta)$. We can therefore write its Fourier series as

$$F(q) = \sum_n c_n e^{2\pi i n q/a};$$

this series converges in the L^2 sense and also pointwise almost everywhere. The coefficients c_n are given by

$$c_n = \frac{1}{a} \int_\Delta dq\, e^{-2\pi i n q/a} \sum_n f(q + na)$$
$$= \frac{1}{a} \int dq\, e^{-2\pi i n q/a} f(q) = (2\pi)^{1/2} \frac{1}{a} \hat{f}\left(\frac{2\pi n}{a}\right),$$

where the interchange of integration and summation is justified, since the series converges absolutely. We thus have

$$\sum_n f(q + na) = (2\pi)^{1/2} \frac{1}{a} \sum_n \hat{f}\left(\frac{2\pi n}{a}\right) e^{2\pi i n q/a}. \tag{2.8}$$

This is Poisson's summation formula; see, e.g., Ref. 19, for a derivation of this formula for other classes of functions. It is now clear that the condition $\hat{f}(2\pi n/a) = 0$ for $n \neq 0$ is necessary and sufficient for $F = \text{const}$. Recapitulating, take f in $\mathscr{C}$. Then $\Sigma_n f(q + na)$ is independent of q if and only if, for every nonzero integer n, one has $\hat{f}(2\pi n/a) = 0$.

This motivates the following definition.

Definition:

$$\mathscr{I}_a = \{f\colon \mathbf{R}\to\mathbf{C};\ f\in\mathscr{C} \text{ and } \hat f(2\pi n/a) = 0,\ \text{for } n\in\mathbf{Z},\ n\neq 0\}.$$

The set $\mathscr{I}_a$ has many interesting properties. We enumerate a few of them.

(1) $\mathscr{I}_a$ is an ideal under convolution in $\mathbf{C}$, i.e., if $f\in\mathscr{I}_a$, $g\in\mathscr{C}$, then $f*g\in\mathscr{I}_a$.

(2) $\mathscr{I}_a$ is invariant under translations: if $f\in\mathscr{I}_a$ then, for every $y\in\mathbf{R}$, the function $x\to f(y-x)$ also belongs to $\mathscr{I}_a$.

(3) If $f\in\mathscr{I}_a$, then, for every $y>0$, the function $x\to f(yx)$ belongs to $\mathscr{I}_{ya}$.

(4) If $f\in\mathscr{I}_a$ then the integral of f can be replaced by a discrete sum:

$$\int dx\, f(x) = a\sum_n f(na) = a\sum_n f(q+na) \quad (\text{for all } q).$$

Proof:

(1) It is easy to check that for $f,g\in\mathscr{C}$, one has $f*g\in\mathscr{C}$. Since $\mathscr{C}\subset L^1(\mathbf{R})$, we have $(f*\hat g)(k)\hat g(k)$ for all real k; hence $(f*\hat g)(2\pi n/a) = 0$ if $\hat f(2n/a) = 0$.

The assertions (2) and (3) are trivial.

$$(4)\ \int dx\, f(x) = (2\pi)^{1/2}\hat f(0) = a(2\pi)^{1/2}\frac{1}{a}\sum_n \hat f\left(\frac{2\pi n}{a}\right)e^{2\pi i n q/a},$$

since $\hat f(2\pi n/a) = 0$ for $n\neq 0$. Then (2.8) gives

$$\int dx\, f(x) = a\sum_n f(q+na).$$

Remark: Notice that (2.7) can be given an interpretation in terms of Zak transform, defined in Sec. II B. To impose condition (2.7) on a function f amounts to requiring that its Zak transform $(Uf)(k,q)$, defined on $[-\pi/a, \pi/a)\times[-a/2,a/2)$, should be constant along the line $k=0$.

III. THE AFFINE CASE

A. Review of definitions and basic properties

The group of shifts and dilations, or the "$ax+b$ group," is the set $\mathbf{R}^*\times\mathbf{R}$ (where $\mathbf{R}^*$ is the set of nonzero real numbers) with the group law

$$(a,b)(a',b') = (aa', ab'+b).$$

We shall here be concerned with the following representation of this group on $L^2(\mathbf{R})$:

$$[U(a,b)f](x) = |a|^{-1/2}f((x-b)/a). \tag{3.1}$$

This representation is irreducible and square integrable, so there exists a dense set of admissible vectors. The admissibility condition (1.5) can in this case be rewritten as

$$c_g = 2\pi\int dp|p|^{-1}|\hat g(p)|^2 < \infty, \tag{3.2}$$

where $\hat g$ is the Fourier transform of g:

$$\hat g(p) = (2\pi)^{-1/2}\int dx\, e^{-ipx}g(x)dx.$$

The fact that not every element of $L^2(\mathbf{R})$ is admissible with respect to the representation (3.1) stems from the nonunimodularity of the $ax+b$ group. The left-invariant measure on the $ax+b$ group is $a^{-2}\,da\,db$; the right-invariant measure is $|a|^{-1}\,da\,db$.

If g is an admissible vector, we define

$$|a,b;g\rangle = U(a,b)g;$$

such families of vectors can be called "affine coherent states."[11,12] The notation just used does not differ from the notation (2.2), used for the Weyl–Heisenberg group. However, it should be clear from the context which family is used at any one time. The general expression (1.4) in the Introduction can then be written for the $ax+b$ group in the following form:

$$\int a^{-2}\,da\,db\,|a,b;g\rangle\langle a,b;g| = c_g\mathbf{1},$$

where c_g is defined by (3.2).

B. Frames of affine coherent states, based on band-limited analyzing wavelets

The families that we shall consider are defined as

$$\{|a_n^+,b_{mn};g\rangle, |a_n^-,b_{mn};g\rangle\} \quad (m,n\in\mathbf{Z}),$$

where

$$a_n = \exp(\alpha n), \quad b_{mn} = \beta m a_n,$$

for some positive numbers α,β. We shall now derive restrictions on these numbers under which this discrete family is a frame.

The function g is supposed to be band limited, i.e., it is square integrable and its Fourier transform has compact support. We shall also assume that the support of $\hat g$ contains only strictly positive frequencies, i.e., is contained in an interval $[l,L]$, with $0<l<L<\infty$. This will enable us to decouple positive and negative frequencies in our calculations, which will turn out to be very convenient. Note that the requirement $l>0$ automatically guarantees that g is admissible [since the condition (3.2) is trivially satisfied].

Let f be any element of $L^2(\mathbf{R})$. We want to show that, under certain conditions on α,β,g to be derived here, we have

$$A\,\|f\|^2 \leqslant \sum_m\sum_n\{|\langle a_n^+,b_{nm};g|f\rangle|^2 + |\langle a_n^-,b_{mn};g|f\rangle|^2\} \leqslant B\,\|f\|^2,$$

with $A>0$, $B<\infty$.

An easy calculation leads to

$$\sum_m|\langle a_n^+,b_{mn};g|f\rangle|^2 = \frac{1}{a_n^+}\sum_m\left|\left(\int_l^L dw\, e^{-iw\beta m}\hat g(w)\hat f\left(\frac{w}{a_n^+}\right)\right)\right|^2.$$

If we impose on β the condition

$$\beta = 2\pi/(L-l),$$

this simplifies to

$$\sum_m |\langle a_n^+, b_{nm}; g | f\rangle|^2 = \frac{1}{a_n^+}\int dw \left|\hat g(w)|^2|\hat f\left(\frac{w}{a_n}\right)\right|^2$$
$$= \int dw |\hat g(a_n^+ w)|^2 . \qquad (3.3)$$

Define now

$F_+(s) = f(e^s)$,

$G(s) = \hat g(e^s)$.

Since $a_n^+ = \exp(\alpha n) > 0$ for all $n\in\mathbf{Z}$, and since supp $\hat g \subset \mathbf{R}_+$ we can make the substitution $t = e^s$ in the integral (3.3) and write

$$\sum_n \sum_m |\langle a_n^+, b_{mn}; g | f\rangle|^2$$
$$= \int ds \Big[\sum_n |G(s+\alpha n)|^2\Big] e^s |F+(s)|^2 . \qquad (3.4)$$

Since supp $G = [\log l, \log L]$ is compact, only a finite number of terms contribute in the sum $\Sigma_n |G(s+\alpha n)|^2$ for any s. If we define now

$$A = \inf_{s\in\mathbf{R}} \Big[\sum_n |G(s+\alpha n)|^2\Big],$$
$$B = \sup_{s\in\mathbf{R}} \Big[\sum_n |G(s+\alpha n)|^2\Big],$$

then clearly

$$A \int_0^\infty dw |\hat f(w)|^2$$
$$<\sum_n \sum_m |\langle a_n^+, b_{mn}; g | f\rangle|^2 < B \int_0^\infty dw |\hat f(w)|^2 , \qquad (3.5)$$

where we have used

$$\int ds\, e^s |F_+(s)|^2 = \int_0^\infty dw |\hat f(w)|^2 .$$

A similar calculation can be made for vectors involving a_n^-. Introducing $F_-(s) = \hat f(-e^s)$, one finds

$$\sum_n \sum_m |\langle a_n^-, b_{mn}; g | f\rangle|^2$$
$$= \int ds \Big[\sum_n |G(s+\alpha n)|^2\Big] e^s |F_-(s)|^2 ,$$

hence

$$A \int_{-\infty}^0 dw |\hat f(w)|^2$$
$$<\sum_n \sum_m |\langle a_n^-, b_{mn}; g | f\rangle|^2 < B \int_{-\infty}^0 dw |\hat f(w)|^2 .$$

Combining this with (3.5) we find thus

$$A \|f\|^2 < \sum_n \sum_m \{ |\langle a_n^+, b_{mn}; g | f\rangle|^2$$
$$+ |\langle a_n^-, b_{mn}; g | f\rangle|^2\} < B \|f\|^2 . \qquad (3.6)$$

If we can derive conditions on α,β,g, ensuring that $A > 0, B < \infty$, then (3.6) implies that under these conditions the set $\{|a^\pm, b_{mn}; g\rangle;\ m,n\in\mathbf{Z}\rangle\}$ is a frame. Since supp $\hat g = [\log l,\ \log L]$, it is clear that A is zero unless $\alpha < \log(L/l)$. If we assume that $\hat g(w)$ is a continuous function without zeros in the interior of its support, then this condition is also sufficient to ensure that $A > 0$. Indeed, we then have

$$A > \inf\{|G(s)|^2;\ \log l + (\log(L/l) - \alpha)/2$$
$$< s < \log L (\log - (L/l) - \alpha)/2\} .$$

As for B, it is not hard to show that

$$B < \{2[\alpha^{-1}\log(L/l)] + 1\}\|\hat g\|_\infty^2 < \infty ,$$

where again we have used the notation $[\mu]$ for the largest integer not exceeding μ.

We have thus derived a set of sufficient conditions ensuring that our construction leads to a frame. The theorem below brings all these conditions together, rewritten in a slightly different form, and states our main conclusion.

Theorem: Let g: $\mathbb{R}\to\mathbb{C}$ satisfy the following conditions: (i) $\hat g$ has compact support $[l,kl]$, with $l > 0, k > 1$; and (ii) $|\hat g|$ is a continuous function, without zeros in the open interval (l,kl). Take $a \in (0,k)$. Define, for $m,n\in\mathbf{Z}$,

$$a_n^+ = \pm a^n ,$$
$$b_{mn} = 2\pi/[(k-1)l]m\, a^n .$$

Then the set $\{|a_n^\pm, b_{mn}; g\rangle;\ m,n\in\mathbf{Z}\}$ is a frame, i.e.,

$$A\mathbf{1} < \sum_n \sum_m \{|a_n^+, b_{mn}; g\rangle\langle a_n^+, b_{mn}; g|$$
$$+ |a_n^-, b_{mn}; g\rangle\langle a_n^-, b_{mn}; g|\} < B\mathbf{1} .$$

The lower and upper bounds A and B are given by

$$A = \inf_{w\in\mathbf{R}} \sum_n |\hat g(a^n w)|^2$$
$$> \inf\{|\hat g(w)|^2;\ w\in[l(k/a)^{1/2}, l(ka)^{1/2}]\} ,$$
$$B = \sup_{w\in\mathbf{R}} \sum_n |\hat g(a^n w)|^2$$
$$< \{2[\log(k/a)] + 1\}\|\hat g\|_\infty^2 .$$

Remarks: (1) The same conclusions can be drawn under slightly less restrictive conditions on g. Strictly speaking we only need $\|\hat g\|_\infty < \infty$ and $\inf_{w\in\Delta} |\hat g(w)| > 0$ for any closed interval Δ contained in (l,kl); both these conditions are of course satisfied if $\hat g$ is continuous and has no zeros in (l,kl).

(2) As the calculations preceding the above theorem show, the positive and negative frequencies decouple neatly. It is therefore possible to choose a different function g_- (and accordingly, also a different lattice a_n, b_{mn}) for the negative frequency domain than for the positive frequency domain.

We have thus constructed a frame, based on a band-limited function g, under fairly general conditions on g. In general, the ratio B/A, comparing the upper with the lower bound, will be larger than 1. Again, however, as in the Weyl–Heisenberg case, it is possible to choose g in such a way that the frame becomes tight, i.e., $B/A = 1$; such tight frames have been used previously by one of us (Y. M.) in Ref. 13(a); they were also used in Ref. 13(b). The construction of such a frame follows more or less the same lines as in the Weyl–Heisenberg case (see Sec. II E); we shall show in the next subsection how the construction works in the present case.

C. Tight frames based on band-limited functions

We shall stick to the same construction as in the preceding subsection, and try to find a function g such that the frame based on g is quasiorthogonal.

Going back to (3.4), it is clear that the frame will be quasiorthogonal if and only if

$$\sum_n |G(s+\alpha n)|^2 = \text{const}, \tag{3.7}$$

where $G(s) = \hat{g}(e^s)$, and $\alpha = \log(a)$ with $a<k$, $\operatorname{supp}\hat{g} = [l,kl]$ $(l>0,\ k>1)$.

This condition (3.7) is exactly the same as the condition (2.6) in the Weyl–Heisenberg case; the analog of L is here $\log(kl) - \log(l) = \log(k)$, while the role of μ is played by $\alpha/\log k = \log(a)/\log(k) < 1$. The only difference is that the function G need not be centered around zero, as was supposed in Sec. II E.

We can therefore copy the construction made in Sec. II E to define a suitable G, hence a suitable g. Explicitly, and directly in terms of $\hat{g}$ rather than in terms of G, this gives

$$\hat{g}(w) = \begin{cases} 0, & \text{for } w<l, \\ [q(\log(w/l)/\log(k/a))]^{1/2}, & \text{for } l<w<lk/a, \\ 1, & \text{for } lk/a<w<al \\ [1-q(\log(w/l)/\log(k/a))]^{1/2} & \text{for } al<w<kl, \\ 0, & \text{for } w>kl, \end{cases} \tag{3.8}$$

where q is a C^{2k} function such that

$$q(x) = \begin{cases} 0, & \text{for } x<0, \\ 1, & \text{for } x>1, \end{cases} \tag{3.9}$$

a strictly increasing between 0 and 1.

Notice that we have assumed that $a^2>k$; this is equivalent to the assumption $\mu>\frac{1}{2}$ in Sec. II E. If $a^2<k$, a similar but more complicated construction can be made.

For $\hat{g}$ constructed as above, the condition (3.7) is satisfied;

$$\sum_n |G(s+\alpha n)|^2 = 1,$$

which implies that the corresponding frame is tight. We thus have proved the following theorem.

Theorem: Let the function $\hat{g}$, with compact support $[l,kl]$ $(l>0, k>1)$, be constructed according to (3.8), with $a>k^{1/2}$ where q is a function satisfying (3.9). Then the set of vectors

$$\{|a^n, 2\pi m a^n/(k-1)l; g\rangle,$$
$$|-a^n, 2\pi m a^n/(k-1)l; g\rangle;\ m,n\in\mathbf{Z}\}$$

(i.e., the set of functions

$$|a|^{-n/2} g[a^{-n}x + 2\pi m/(k-1)l],$$
$$|a|^{-n/2} g[-a^{-n}x + 2\pi m/(k-1)l])$$

is a quasiorthogonal frame in $L^2(\mathbf{R})$, with $A=B=1$. This means the following: If f is any function in $L^2(\mathbf{R})$ and if we define coefficients $f_{mn}^{(\pm)}$ $(m,n\in\mathbf{Z})$ by

$$f_{mn}^{(+)} = |a|^{-n/2} \int dx\, g\left[a^{-n}x + \frac{2\pi m}{(k-1)l}\right]^* f(x),$$

$$f_{mn}^{(-)} = |a|^{-n/2} \int dx\, g\left[-a^{-n}x + \frac{2\pi m}{(k-1)l}\right]^* f(x),$$

then

$$f(x) = \sum_m \sum_n f_{mm}^{(+)} |a|^{-n/2} g\left[a^{-n}x + \frac{2\pi m}{(k-1)l}\right] + \sum_m \sum_n f_{mm}^{(-)} |a|^{-n/2} g\left[-a^{-n}x + \frac{2\pi m}{(k-1)l}\right],$$

where the sum converges in $L^2(\mathbf{R})$.

Let us give some specific examples.

Example 1: We take $l=1$, $k=3$, $a=\sqrt{3}$. Define

$$\hat{g}(w) = \begin{cases} 0, & \text{for } w<1, \\ \sin[\pi \log w/\log 3], & \text{for } 1<w<3, \\ 0, & \text{for } w>3. \end{cases}$$

The corresponding g cannot be calculated in closed analytic form. A graph of $\operatorname{Re} g$, $\operatorname{Im} g$ is given in Fig. 2. The corresponding function q is the same as in the example in Sec. II E:

$$q(x) = \begin{cases} 0, & \text{for } x<0, \\ \sin^2(\pi x/2), & \text{for } 0<x<1, \\ 1, & \text{for } x>1. \end{cases}$$

Example 2: We take $l=1$, $k=4$, $a=2$. Define

$$\hat{g}(w) = \begin{cases} 0, & \text{for } w<1, \\ 2\sqrt{2}[\log w/\log 2]^2, & \text{for } 1<w<\sqrt{2}, \\ [1-8(1-\log w/\log 2)^4]^{1/2}, & \text{for } \sqrt{2}<w<2\sqrt{2}, \\ 2\sqrt{2}[2-(\log w/\log 2)]^2, & \text{for } 2\sqrt{2}<w<4, \\ 0, & \text{for } w>4. \end{cases}$$

The corresponding function q is

$$q(x) = \begin{cases} 0, & \text{for } x<0, \\ 8x^4, & \text{for } 0<x<\frac{1}{2}, \\ 1-8(1-x)^4, & \text{for } \frac{1}{2}<x<1, \\ 1, & \text{for } x>1. \end{cases}$$

Graphs of $\operatorname{Re} g$, $\operatorname{Im} g$ are given in Fig. 3.

Because of the correspondence, noted above, between tight frames for the $ax+b$ group and the Weyl–Heisenberg group, all the examples given in the Appendix for the Weyl–Heisenberg group can easily be transposed to the present case.

D. A closer look at the necessary and sufficient condition

The necessary and sufficient condition that a band-limited function g, concentrated on positive frequencies, has to satisfy in order to generate a tight frame, is given by (3.7). This can be rewritten as

$$\sum_n f(a^n w) = \text{const}, \tag{3.10}$$

where $f(w) = |\hat{g}(w)|^2$.

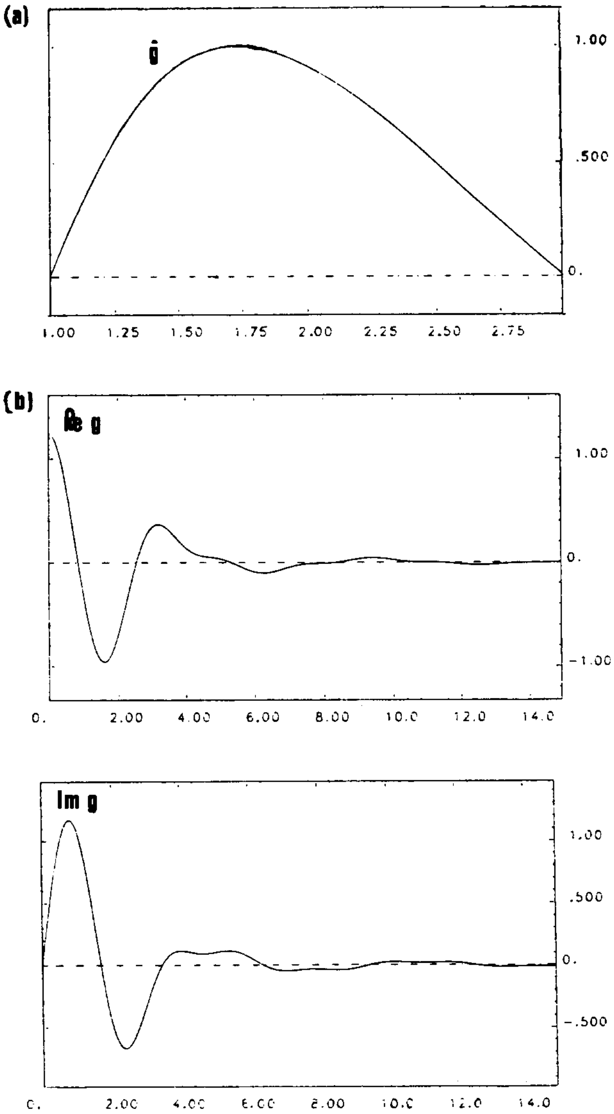

FIG. 2. (a) The function $\hat{g}(w) = \sin(\pi \log w/\log 3)\chi_{[1,3]}(w)$. (b) Real and imaginary parts of the inverse Fourier transform g of $\hat{g}$.

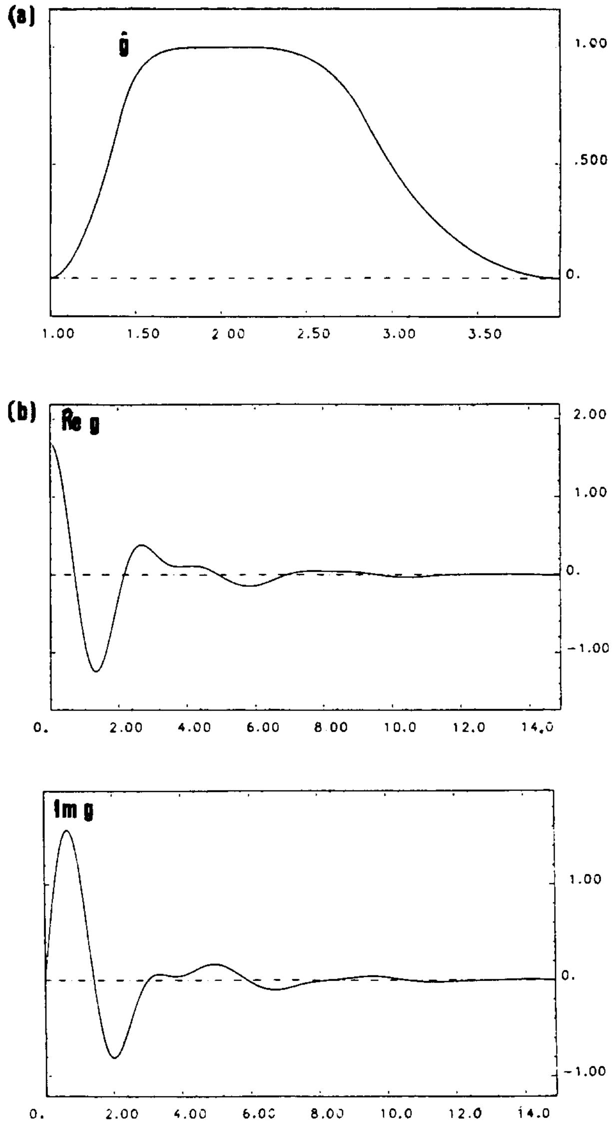

FIG. 3. (a) The function $\hat{g}(w)$ defined in example 2 of Sec. III D. (b) Real and imaginary parts of the inverse Fourier transform g or $\hat{g}$.

In this subsection we shall study the class of functions f: $\mathbf{R}_+ \to \mathbf{C}$, satisfying (3.10); for the purpose of this subsection only, we shall not require f to be positive or to have compact support. This study will be completely analogous to our study in Sec. II F of the functions satisfying condition (2.7); since the arguments run along exactly the same lines, we shall not go into as much detail here. The main difference is that we shall work with the Mellin transform of f rather than with its Fourier transform; this, of course, is due to the difference between (3.10), where the constant enters multiplicatively, and (2.7) where it enters additively.

The Mellin transform F of f: $\mathbf{R}_+ \to \mathbf{C}$ is defined by

$$F(s) = \int_0^\infty w^{s-1} f(w)\,dw\,;$$

the inversion formula is

$$f(w) = \frac{1}{2\pi i}\int ds\, t^{-s} F(s)\,,$$

where the integral is taken from $c - i\infty$ to $c + i\infty$ and where $c > 0$ has to be chosen so that the integral converges.

For the sake of convenience we shall restrict ourselves to functions $f \in \mathscr{C}^{\sim}$, where $\mathscr{C}^{\sim}$ is defined below.

Definition:

$$\mathscr{C}^{\sim} = \{f\colon \mathbf{R}_+ \to \mathbf{C};\ f \text{ is measurable and there exists } C > 0 \text{ and } k > 1 \text{ such that } |f(w)| < C(1 + |\log w|)^{-k}\}\,.$$

For functions $f \in \mathscr{C}^{\sim}$ the series $\Sigma_n f(a^n t)$ is absolutely convergent, uniformly on $(1,a]$; the Mellin transform F of f is well defined for purely imaginary arguments, and the inversion formula applies, with $c = 0$.

Notice that if we define g: $\mathbf{R} \to \mathbf{C}$ by $g(x) = f(e^x)$, we immediately find

$$f \in \mathscr{C}^{\sim} \Leftrightarrow g \in \mathscr{C}$$

(where $\mathscr{C}$ is defined in Sec. II F) and

$$F(ik) = (2\pi)^{1/2}\hat{g}(k) \quad (k \in \mathbf{R})\,.$$

This enables us to translate the results of Sec. II F to the present situation.

We have, for $f\in\mathscr{C}^{\sim}$,

$$\sum_n f(a^n w) = \frac{1}{\log a}\sum_n F\left(\frac{2\pi i n}{\log a}\right) w^{2\pi i n/\log a},$$

which leads to the following theorem.

Theorem: Take $f\in\mathscr{C}^{\sim}$. Then $\Sigma_n f(a^n w)$ is independent of w if and only if $F(2\pi i n/\log a) = 0$ for all nonzero integers n.

This then motivates the following definition.

Definition:

$$\mathscr{I}_a^{\sim} = \{f\colon \mathbb{R}_+ \to \mathbb{C}, \text{ and } F(2\pi i n/\log a) = 0, \text{ for } n\in\mathbb{Z},\ n\neq 0\}.$$

The following theorem lists a few properties of $\mathscr{I}_a^{\sim}$.

Theorem: (1) $\mathscr{I}_a^{\sim}$ is an ideal in $\mathscr{C}^{\sim}$ under "Mellin convolution," i.e., if $f\in\mathscr{I}_a^{\sim}$, $g\in\mathscr{C}^{\sim}$, then the function $f \underset{M}{*} g$ defined by

$$(f \underset{M}{*} g)(w) = \int_0^\infty \frac{du}{u} f(u) g\left(\frac{t}{u}\right),$$

belongs to $\mathscr{I}_a^{\sim}$.

(2) $\mathscr{I}_a^{\sim}$ is invariant under dilations: if $f\in\mathscr{I}_a^{\sim}$, then, for every $u\neq 0$, the function $w\to f(uw)$ also belongs to $\mathscr{I}_a^{\sim}$.

(3) If $f\in\mathscr{I}_a$, then the integral of $w^{-1}f(w)$ can be replaced by a discrete sum:

$$\int_0^\infty dw\, w^{-1} f(w) = \log a \sum_n f(a^n)$$

$$= \log a \sum_n f(a^n t), \quad \text{for all } t.$$

Notice that the "Mellin convolution" defined above is exactly the convolution in the sense of Ref. 19, with respect to the multiplicative group of nonzero real numbers. It is obvious that the Mellin transform of a "Mellin convolution" of two functions is given, up to a constant factor, by the product of the two Mellin transforms.

ACKNOWLEDGMENT

One of us (I. D.) would like to thank the Centre de Physique Theorique 2, CNRS-Marseille, where part of this work was done, for its hospitality and support.

APPENDIX: MORE EXAMPLES

We give a few more examples of functions h supported in $[-L/2, L/2]$ and satisfying the condition

$$\sum_{n\in\mathbb{Z}} |h(x + n\mu L)|^2 = \text{const}. \tag{A1}$$

We start by showing how one can extend the construction given in Sec. II E, for the case $\mu \geqslant \frac{1}{2}$, to more general μ.

For $2^{-(k+1)} \leqslant \mu \leqslant 2^{-k}$, with $k\geqslant 0$, one can always define $\mu^{\sim} = 2^k\mu$. Obviously $\frac{1}{2}\leqslant\mu^{\sim}\leqslant 1$. If we replace μ by $\mu^{\sim}$ in the construction of Sec. II E, we obtain a function h satisfying

$$\sum_{n\in\mathbb{Z}} |h(x + n2^k\mu L)|^2 = \text{const} = C.$$

Hence

$$\sum_{n\in\mathbb{Z}} |h(x + n\mu L)|^2 = 2^k C.$$

Functions h constructed in this way thus obviously satisfy condition (A1). It is clear from the construction that the tight frame $\{|n\mu L, m2\pi/L; h\rangle\}$ generated by h is in this case a superposition of the tight frame $\{|n\mu^{\sim} L, m2\pi/L; h\rangle\}$ and translated copies of this frame.

There also exist, of course, functions h satisfying (A1), for $\mu < \frac{1}{2}$, which cannot be reduced to the case $\frac{1}{2}\leqslant\mu\leqslant 1$. We give here an example of such a function, for the case $\mu = \frac{1}{3}$.

Let g be again a C^{2k} function, strictly increasing, such that

$$g(x) = \begin{cases} 0, & \text{for } x\leqslant 0, \\ 1, & \text{for } x\geqslant 1. \end{cases}$$

Define then

$$h(x) = \begin{cases} 0, & \text{for } x\leqslant -L/2, \\ \{g[(6x+3L)/4L]\}^{1/2}, & \text{for } -L/2\leqslant x\leqslant L/6, \\ \{1 + g(1/2) - g[(6x+L)/4L] \\ \quad - g[(6x-L)/4L]\}^{1/2}, \\ \quad \text{for } L/6\leqslant x\leqslant L/2, \\ 0, & \text{for } x\geqslant L/2. \end{cases}$$

Obviously h has support $[-L/2, L/2]$. One can check that h is a C^k function satisfying

$$\sum_{n\in\mathbb{Z}} \left| h\left(x + \frac{nL}{3}\right)\right|^2 = 1 + g\left(\frac{1}{2}\right).$$

This example can also be adapted to cover the case $\frac{1}{2} > \mu > \frac{1}{3}$ (instead of only $\mu = \frac{1}{3}$).

Finally, note that another class of examples, for the special cases $\mu = 1/2(k+1)$, with k a positive integer, can be constructed with the help of spline functions. Choose a knot sequence $t = (t_j)_{j\in\mathbb{Z}}$ with equidistant knots, $t_{j+1} - t_j = d > 0$ for all j. Let $B_{j,2k+2,t}$ be the jth B-spline of order $2k+2$ for the knot sequence t. (For the definition of B-splines, see, e.g., Ref. 20.) Then the $B_{j,2k+2,t}$ are all translated copies of $B_{0,2k+2,t}$:

$$B_j(x) = B_0(x - jd).$$

The B_j are (positive) C^{2k} functions, with support $[t_j, t_{j+2(k+1)}] = [t_j, t_j + 2(k+1)d]$. They have, moreover, the property that

$$\sum_{j\in\mathbb{Z}} B_j(x) = 1, \quad \text{for all } x$$

(only a finite number of terms contribute for any x). It is now easy to check that the function h, defined by

$$h(x) = \{B_{0,k,t}[t_0 + (k+1)d(2x+L)/L]\}^{1/2},$$

is a C^k function with support $[-L/2, L/2]$, satisfying condition (A1) with $\mu = 1/2(k+1)$.

[1]R. M. Young, *An Introduction to Non-harmonic Fourier Series* (Academic, New York, 1980).

[2]R. J. Duffin and A. C. Schaeffer, Trans. Am. Math. Soc. **72**, 341 (1952).

[3] I. Daubechies and A. Grossmann "Frames in the Bargmann space of entire functions" (to be published).
[4] A. Grossmann and J. Morlet, SIAM J. Math. Ann. **15**, 723 (1984); A. Grossmann and J. Morlet, "Decomposition of functions into wavelets of constant shape, and related transforms," to appear in *Mathematics and Physics, Lectures on Recent Results* (World Scientific, Singapore, 1985); P. Goupillaud, A. Grossmann, and J. Morlet, Geoexploration **23**, 85 (1984).
[5] M. Bertero, C. De Mol, and G. Viano, "The stability of inverse problems," in *Inverse Scattering Problems in Optics*, edited by H. P. Baltes (Springer, Berlin, 1980).
[6] (a) M. Duflo and C. C. Moore, J. Funct. Anal. **21**, 208 (1976); A. L. Carey, Bull. Austr. Math. Soc. **15**, 1 (1976); (b) S. A. Gaal, *Linear Analysis and Representation Theory* (Springer, Berlin, 1973); (c) A. Grossmann, J. Morlet, and T. Paul, J. Math. Phys. **26**, 2473 (1985).
[7] J. von Neumann, *Mathematical Foundation of Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1985) (English translation).
[8] V. Bargmann, P. Butera, L. Girardello, and J. R. Klauder, Rep. Math. Phys. **2**, 221 (1971).
[9] A. M. Perelomov, Teor. Mat. Fiz. **6**, 213 (1971).
[10] (a) H. Bacry, A. Grossmann, and J. Zak, Phys. Rev. B **12**, 1118 (1975); (b) R. Balian, C. R. Acad. Sci. Paris, Ser. 2 **292**, 1357 (1981).
[11] E. W. Aslaksen and J. R. Klauder, J. Math. Phys. **10**, 2267 (1969).
[12] T. Paul, J. Math. Phys. **25**, 3252 (1984); A. Grossmann and T. Paul, "Wave functions on subgroups of the group of affine canonical transformations," in *Lecture Notes in Physics*, Vol. 211 (Springer, Berlin, 1984).
[13] (a) Y. Meyer, "La transformation en ondellettes et les nouveaux paraproduits," to be published; (b) M. Frasier and B. Jawerth, "Decomposition of Besov spaces," to be published.
[14] Y. Meyer, "La transformation en ondelettes: de la recherche pétrolière à la géométrie des espaces de Banach," to be published; P. G. Lemarié and Y. Meyer, "Ondelettes et bases hilbertiennes," to be published.
[15] A. Grossmann and T. Paul (in preparation).
[16] J. Klauder and B. S. Skagerstam, *Coherent States. Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1985).
[17] J. Zak, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1972), Vol. 27.
[18] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[19] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis, II* (Springer, Berlin, 1970).
[20] C. de Boor, *A Practical Guide to Splines* (Springer, Berlin, 1978).

The Thomas precession and velocity-space curvature

Shahar Ben-Menahem[a)]
Department of Physics, Boston University, Boston, Massachusetts 02215

(Received 27 September 1985; accepted for publication 30 December 1985)

The motion of a physical system acted upon by external torqueless forces causes the relativistic Thomas precession of the system's spin vector, relative to an inertial frame. A time-dependent force that returns the system to its initial velocity is considered. The precession accumulates to become a finite rotation of the final spin vector, relative to its initial value. This rotation is commonly explained as the Wigner rotation due to the sequence of pure boosts caused by the force. An alternative interpretation is presented here: The rotation is due to the change of the spin vector as it is parallel-transported around the closed trajectory described by the system in hyperbolic three-velocity space. As an application, the angle of precession for a planar motion is shown to be equal to the area enclosed by the trajectory in velocity space.

I. INTRODUCTION

Consider a physical system moving through Minkowskian space-time. Let this system be acted upon by external torqueless forces, that is, forces that give rise to no torques about the instantaneous center-of-mass (COM) of the system, as measured in its instantaneous rest frame.[1-3] Let us further assume that the time-dependent forces return the system to its initial velocity after some finite time,[4] and let this velocity vanish in $\mathscr{O}$, an initial rest frame of the system.

The Pauli–Lubanski spin vector S_μ is defined to be $(0,S_i)$ in the system's rest frame, with S_i the angular-momentum three-vector[5] about the COM. In the motion described above, S_i undergoes a rotation (in $\mathscr{O}$) from its initial to final value. The continuous precession that results in this rotation is known as the Thomas precession.

The effect of the torqueless forces on S_μ, between two infinitesimally close instants of proper system time, is to multiply it by an infinitesimal-pure-boost matrix. The product of these matrices for the entire trajectory corresponds to a pure spatial rotation (the Wigner rotation), and this is the common explanation of the precession.[6] Note that the rotation applies also when the initial and final velocities differ, since we can multiply S_μ by a finite pure-boost matrix to return to a state of rest in $\mathscr{O}$, and then measure the rotation.

After reviewing the derivation of the precession using boosts in Sec. II, we present our alternative interpretation for it in Sec. III. A three-dimensional curved velocity space V is defined as the surface $v_\mu v^\mu = 1$ embedded in Minkowskian four-velocity space. We prove that $S_i^{(V)}$, the projection of S to V, is parallel-transported around the closed trajectory in V, and that its change thereupon is just the accumulated Thomas precession. Finally, in Sec. IV this method is used to relate the angle of precession for planar motion, to the area enclosed by the physical system's trajectory in V.

II. INFINITESIMAL BOOSTS

The motion[7] may be viewed as a succession of instantaneous rest frames $R(s)$, where s is proper time. In such a frame, the effect of the torqueless forces acting at time $s = \tau$ during an infinitesimal interval ds, is to leave S_i unchanged:

$$\frac{d}{ds} S_i(s)\bigg|_{s=\tau} = 0 \quad \text{in } R(\tau). \tag{1}$$

The component S_0, however, is changed, as we now show.

Let v^μ be the four-velocity; since $S_0(s)$ vanishes in $R(s)$,

$$\frac{d}{ds}(v^\mu S_\mu) = 0 \quad \text{in any frame,} \tag{2}$$

or, if a^μ is the four-acceleration,

$$v^\mu \frac{dS_\mu}{ds} + a^\mu S_\mu = 0. \tag{3}$$

At $s = \tau$ and in the frame $R(\tau)$, $v^\mu(\tau) = (1,0)$ and $a^\mu(\tau) = (0,a^i)$, so (3) becomes

$$\frac{d}{ds} S_0(s)\bigg|_{s=\tau} = -a^i S_i(\tau) \quad \text{in } R(\tau), \tag{4}$$

the promised change in S_0. This may be written as

$$S_0(\tau + ds) = -S_i(\tau) dv^i \quad \text{in } R(\tau), \tag{4'}$$

where $dv^i = a^i\, ds$ is the relative velocity of $R(\tau + ds)$ with respect to $R(\tau)$. Thus the forces cause S_μ to change as follows from τ to $\tau + ds$:

$$S_\mu \text{ in } R(\tau)\text{: } (0,S_i) \rightarrow (-S_i\, dv^i, S_i). \tag{5}$$

This is just the action of an infinitesimal boost of velocity $(-dv^i)$ upon S_μ (see Ref. 8). The effect of the force on S_μ during its finite duration, as measured in $\mathscr{O}$, will thus be an infinite product of infinitesimal boosts, which return the system to a state of rest in $\mathscr{O}$. But such a product is a pure spatial rotation (Wigner's rotation). The *differential* precession can be computed by referring S_μ, at any proper time, to a rest frame $B(s)$, obtained from $\mathscr{O}$ by a pure boost. Let $b^\mu_{(i)}(s)$ be the unit vector along the ith spatial axis of $B(s)$ [note that $B(s)$ and $R(s)$ differ by a rotation]. The precession is then found to be[6]

$$\frac{d}{ds}\left(b^\mu_{(i)}(s) S_\mu(s)\right) = \epsilon_{ijk} \Omega_{(j)}(s) \left(b^\mu_{(k)}(s) S_\mu(s)\right), \tag{6a}$$

$$\Omega_{(i)}(s) = \epsilon_{ijk} a^j(s) v^k(s) \left[1/(1 + v^0(s))\right]. \tag{6b}$$

In these formulas the v^μ are the components of the four-velocity in $\mathscr{O}$, and the a^μ are those of the four-acceleration in $\mathscr{O}$. Here $\Omega_{(i)}$ is not a tensor, whereas $b^\mu_{(i)}$ is a tensor in its upper index.

[a)] Current address: The Weizmann Institute of Science, Rehovot 76100, Israel.

 0022-2488/86/051284-03$02.50

III. PARALLEL TRANSPORT

Our alternative interpretation of the precession involves three-velocity space. Since we assumed that the forces return the system to its initial velocity, it describes a closed trajectory in that space. Each point along this trajectory then represents one of the instantaneous rest frames [e.g., $B(s)$ or $R(s)$].

Intuitively, the physical picture of the infinitesimal boosts between these frames, such that $dS_i/ds|_{s=\tau}=0$ in $R(\tau)$ [Eq. (1)], resembles parallel transport in differential geometry. We now proceed to make this analogy rigorous.

We define three-velocity space to be a three-dimensional manifold V, rather than a vector space. We want to embed it in a Minkowskian four-velocity space, and demand that boosts and rotations be isometries of V. Thus it must be the maximally symmetric, hyperbolic subspace defined by the constraint

$$v_\mu v^\mu = 1, \tag{7}$$

which is the condition satisfied by physical four-velocities. The spin vector in V is obtained from S_μ by a standard projection: for coordinates ζ^i on V, the projected vector is

$$S_i^{(V)} \equiv \frac{\partial v^\mu}{\partial \zeta^i} S_\mu . \tag{8}$$

When $S^{(V)}$ is parallel-transported along the trajectory, its components satisfy $(d/ds)S_i^{(V)}=0$ in a locally flat (Cartesian) coordinate system, i.e., one where the connection vanishes locally. Such a coordinate system is, at a point of proper time s,

$$\zeta^i = (v^i \text{ in } R(s)). \tag{9}$$

(We will continue to include in equations, where necessary, the Minkowski frame in which they hold, which should not be confused with the coordinate system on V.)

We will now prove that (9) are such coordinates. The metric on V is

$$g_{ij} = \frac{\partial v^\mu}{\partial \zeta^i}\frac{\partial v_\mu}{\partial \zeta^j}, \tag{10}$$

and working in the Minkowski frame $R(s)$, we have, from (7) and (9),

$$\begin{aligned} v^0 &= 1 + \tfrac{1}{2}\zeta^i\zeta^i + O(\zeta^m\zeta^n\zeta^k\zeta^l), \\ \frac{\partial v^\mu}{\partial \zeta^i} &= \delta_{\mu i} + \delta_{\mu 0}\zeta^i + O(\zeta^m\zeta^n\zeta^k). \end{aligned} \tag{11}$$

Hence by (10),

$$g_{ij} = \delta_{ij} + O(\zeta^m\zeta^n), \tag{12}$$

and the affine connection vanishes at $\zeta^m=0$. This proves that the ζ^i are indeed locally flat.

From (8) and (11), the parallel transport condition $(d/ds)S_i^{(V)}=0$ becomes

$$\frac{d}{ds}(S_i + \zeta^i S_0)\Big|_{s=\tau} = 0 \quad \text{in } R(\tau), \tag{13}$$

where the coordinates (9) are constructed at point $s=\tau$ along the trajectory. Since $S_0(\tau)=0$ in $R(\tau)$, (13) becomes

$$\frac{d}{ds}S_i\Big|_{s=\tau} = 0 \quad \text{in } R(\tau), \tag{13'}$$

which is the equation governing the action of the torqueless forces on S_μ [see (1) above].

In order to complete the parallel-transport picture of the Thomas precession, we extend the locally flat coordinates (9) at $s=0$ to cover the entire manifold[9] V. Here we define $s=0$ as the proper time at the initial point A of the trajectory, and $R(0)=\mathscr{O}$. At A, $\zeta^i=0$ and thus, by (8) and (11),

$$S_i^{(V)}(0) = (S_i(0) \text{ in } \mathscr{O}). \tag{14}$$

Since the trajectory is closed by assumption, its final point (at $s=\sigma$) will also be A. Working again in the Minkowski frame $\mathscr{O}$ and the same coordinate system ζ^i, we find

$$S_i^{(V)}(\sigma) = (S_i(\sigma) \text{ in } \mathscr{O}). \tag{15}$$

From (14) and (15) we see that $S_i^{(V)}$, in the coordinates that are flat at A, undergoes the same rotation from $s=0$ to $s=\sigma$ as does S_μ in the Minkowski frame $\mathscr{O}$. This rotation, as said in the previous section, is the accumulated Thomas precession.

IV. THE CASE OF PLANAR MOTION

Let the forces be such that there exist Minkowski frames in which one spatial component of v^μ, say v^3, vanishes along the trajectory (i.e., planar motion). We will then use a theorem in differential geometry to give a novel interpretation to the angle, through which S_μ processes during the accelerated motion.[10]

Consider the two-dimensional submanifold W of V, defined by $v^3=0$. From Gauss' integral curvature theorem it follows[11] that since W is orientable, the parallel transport of $S_i^{(W)}$ around the trajectory results in a precession through the counterclockwise angle,

$$\theta_\Sigma = -\int_\Sigma K\sqrt{\det g}\, d\zeta^1 d\zeta^2, \tag{16}$$

where ζ^i is any right-handed[12] coordinate system on W, g is the metric of W, and K is the Gaussian curvature. The integral (16) is over the region $\Sigma \subset W$ enclosed by the trajectory. Actually, (16) is only valid for a trajectory that goes around Σ in a counterclockwise sense. For the clockwise case, there is an extra minus sign; if the trajectory crosses itself, this divides Σ into two or more regions Σ_i, each clockwise or counterclockwise. The total precession is the algebraic sum of θ_{Σ_i}. We may therefore restrict our attention to a trajectory that goes around Σ in a constantly counterclockwise sense.

W is defined by the embedding

$$\begin{aligned} ds^2 &= (dv^0)^2 - (dv^1)^2 - (dv^2)^2, \\ (v^0)^2 &- (v^1)^2 - (v^2)^2 = 1, \end{aligned} \tag{17}$$

and is therefore maximally symmetric, since both rotations among v^1, v^2, and boosts that mix v^0,v^1,v^2 are isometries of W. Therefore[13] it is a space of constant Gaussian curvature, and in fact it has $K=-1$. Thus, from (16), the counterclockwise angle of precession is positive for a counterclockwise trajectory, and equal to the invariant area of the region Σ enclosed by the trajectory.

ACKNOWLEDGMENT

This research is supported in part by the Department of Energy under Contract No. ACO283ER40118.

[1] The concepts of COM and rest frame for a spatially extended relativistic system are somewhat subtle, and not uniquely defined. See Refs. 2 and 3.

[2] A. Schild, in *Relativity Theory and Astrophysics, 1. Relativity and Cosmology, Lectures in Applied Mathematics*, Vol. 8, edited by J. Ehlers (Am. Math. Soc., Providence, RI, 1967), pp. 17–20.

[3] H. P. Robertson and T. W. Noonan, *Relativity and Cosmology* (Saunders, Philadelphia, 1968), §5.4 and 5.5, pp. 139–144.

[4] It will be seen below that this assumption involves no loss of generality.

[5] We use Latin letters for spatial indices, and Greek for Minkowski indices. Our Minkowski metric shall be $(1, -1, -1, -1)$, and we use the normal convention $\epsilon_{123} = 1$, $\epsilon_{ijk} = -\epsilon_{jik}$, and $\epsilon_{ijk} = \epsilon_{jki}$.

[6] Reference 3, §3.11, pp. 66–69. These authors derive the precession as an infinitesimal Wigner rotation, and do not treat the spin vector specifically. A reference that introduces the spin precession in particular, in the context of atomic physics, where it was discovered, is J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962), §11.5, pp. 364–369. Our formula is a more formal version of that appearing in Robertson and Noonan [see our Eq. (6)].

[7] From this point on, all quantities are implicitly understood to be those of the physical system: velocity, acceleration, spin, etc.

[8] Alternatively, one may boost to a frame $R(\tau + ds)$, to find

$$(S_\mu(\tau) \text{ in } R(\tau)) = (S_\mu(\tau + ds) \text{ in } R(\tau + ds)).$$

Since the physical space axes and clocks, used to operationally define the frames $R(s)$, indeed undergo this boost [of velocity $(+dv^i)$] due to the forces, it follows that the Thomas precession is unobservable via internal measurements in the physical system.

[9] This is possible because V is topologically equivalent to R^3.

[10] Since the problem is now effectively that of two spatial dimensions, the Wigner rotations are characterized by a single angle.

[11] Reference 3, §7.1 and §8.12. Gauss' theorem applies to geodesic triangles, but the area enclosed by the trajectory can be divided into an infinite number of such triangles.

[12] That is, (ξ^1, ξ^2, v^3) is a right-handed system in V.

[13] On maximally symmetric spaces, see, e.g., Ref. 3, Chap. 13.

Simulation of classical particle trajectories in a complex two-dimensional space

G. H. Derrick
School of Physics, University of Sydney, Sydney, New South Wales 2006, Australia

(Received 3 June 1985; accepted for publication 11 December 1985)

Classical particles of arbitrary rest mass and spin are modeled in a two-dimensional space C_2 (which has two complex coordinates ξ^A, $A = 1, 2$) in the following way. It is first shown that a preferred set of trajectories $\xi^A = \zeta^A(s)$, designated geodesics, can be introduced from a variation principle that makes stationary the real variable s. The latter plays the role of proper time and serves to parametrize the trajectories. The geodesics so defined are then shown to be associated with a nonlinear representation of the Poincaré group. A set of Poincaré vectors and tensors are constructed from ζ^A and its proper time derivatives and these simulate the properties of the position, momentum, angular momentum, and internal angular momentum variables of a classical massive particle of arbitrary spin.

I. INTRODUCTION

Space and time coordinates are derived initially from operations on meter sticks and clocks, or at a more sophisticated level, from atomic and molecular radiation processes. However, at the nuclear and subnuclear level, the concept of the space-time continuum could well require modification, possibly a radical change in topology, or perhaps quantization in some sense. Rzewuski[1] suggested that vectors in Minkowski space M_4 should be derived from a more basic four-dimensional space of Dirac-like complex spinors. His ideas were given a firm basis in the twistor theory of Penrose and co-workers,[2–7] where the substructure behind M_4 is a complex projective three-space whose points are O(2, 4) spinors, i.e., "twistors." These correspond to null straight lines in M_4 and can be used to model particles of zero rest mass. A space-time event is given as the intersection of null lines, necessitating two or more twistors for its representation. In n-twistor particle theory, $n \geqslant 2$, both the external and internal symmetries of massive particles can be modeled.[8–12]

A different approach is taken by Nash,[13] who replaces M_4 as the basic manifold by a 16-dimensional real space, which is in a rough sense a square root of M_4. The vectors in this space may be constructed from the eight complex components of two independent twistors. Nash finds quadratic forms in his 16 coordinates, which model the usual position, momentum, and angular momentum variables of a free particle of arbitrary mass and spin. Similar in spirit is the work of Derrick,[14] who shows that the behavior of these same variables can be simulated by certain bilinear forms in the components of a vector in an eight-dimensional complex space. Once again two independent twistors may be used to construct this eight-component vector.

Thus in this previous work the basic space from which M_4 is derived needs to have at least eight complex or equivalently 16 real coordinates if one wishes to describe particles of nonzero rest mass. Such high dimensionality would seem to make it implausible that any of these spaces could be a fundamental entity.

In this paper we show how M_4 can be related to a two-dimensional complex space C_2 having only two complex coordinates[15] ξ^A. We exhibit geometric quantities in C_2 that model particles of nonzero rest mass and arbitrary spin and possess the full Poincaré invariance associated with M_4. For the subgroup of proper, orthochronous Lorentz transformations, SO (1, 3), the invariance is effected by requiring that $\xi = \begin{bmatrix} \xi^1 \\ \xi^2 \end{bmatrix}$ transforms according to the $D^{(1/2)0}$ representation of the covering group SL(2, $\mathbb{C}$). Thus to the element exp $(g\omega) \in$ SO (1, 3), there corresponds the transformation[15]

$$\xi' = \exp[-(i/4)\omega_{\kappa\lambda}\sigma^{\kappa\lambda}]\xi\,. \quad (1)$$

Here, $\omega = [\omega_{\kappa\lambda}]$ is an arbitrary real antisymmetric 4×4 matrix, $g = [g^{\kappa\lambda}] = [g_{\kappa\lambda}] = \text{diag.}(1, -1, -1, -1)$ is the Minkowski metric, while the spin coefficients are given in terms of the Pauli matrices by

$$(\sigma^{23}, \sigma^{31}, \sigma^{12}) = -i(\sigma^{01}, \sigma^{02}, \sigma^{03}) = \boldsymbol{\sigma}\,. \quad (2)$$

In contrast, we cannot achieve invariance with respect to the translation elements of the Poincaré group by any linear change of coordinates analogous to (1). There is simply no faithful two-dimensional matrix representation of the Poincaré group. For the same reason any attempt to identify the Minkowski coordinates[15] x^λ with four independent real functions of the ξ^A will not succeed.

Instead we proceed as follows. In Sec. II we introduce a set of preferred trajectories in C_2, $\xi^A = \zeta^A(\tau)$, τ being a parameter. These are analogous to straight lines in M_4, or to geodesics in a Riemannian space, and are derived from a variation principle for the "proper time" along the trajectories.

In Sec. III we exhibit a transformation law for $\zeta^A(\tau)$, which gives a nonlinear representation of the full Poincaré group. Finally, in Sec. IV we identify certain real quantities constructed from $\zeta^A(\tau)$ and its τ derivatives, which exhibit many of the properties of the position, momentum, and angular momentum variables of a Minkowski space particle of arbitrary mass and spin.

II. GEODESICS IN C_2

The straight line geodesics joining two points x_1^λ and x_2^λ in M_4 may be derived from the variation principle

 0022-2488/86/051287-06$02.50

$$\delta s_{12} = 0\,, \tag{3}$$

where

$$s_{12} = \int_1^2 L_M \, d\tau\,, \tag{4}$$

with

$$L_M = \left[g_{\lambda\mu} \frac{dz^\lambda}{d\tau} \frac{dz^\mu}{d\tau} \right]^{1/2}, \tag{5}$$

is interpreted as the proper time along any trajectory $x^\lambda = z^\lambda(\tau)$ that links the two points. Note that s_{12} depends only on the trajectory and not on the choice of parameter τ.

Let us try an analogous prescription to select a preferred set of trajectories in C_2, which we will again designate *geodesics*, though C_2 has no conventional metric. Along any trajectory $\zeta^A = \zeta^A(\tau)$, we wish to define some real Lagrangian L_C that is an SL (2, $\mathbb{C}$) scalar function of the real and imaginary parts of $\zeta^A(\tau)$ and the τ-derivatives of these quantities. Given any two spinors

$$p = \begin{bmatrix} p^1 \\ p^2 \end{bmatrix}, \quad q = \begin{bmatrix} q^1 \\ q^2 \end{bmatrix},$$

which belong to $D^{(1/2)0}$, we can form the SL (2, $\mathbb{C}$) scalar[15]

$$p \cdot q = p^T \epsilon q = p^1 q^2 - p^2 q^1 = -q \cdot p\,, \tag{6}$$

where

$$\epsilon = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{7}$$

Thus from $\zeta = \left[\begin{smallmatrix} \zeta^1(\tau) \\ \zeta^2(\tau) \end{smallmatrix}\right]$ and its derivatives we can form the scalars

$$\eta_{01} = \zeta \cdot \frac{d\zeta}{d\tau}\,, \tag{8}$$

$$\eta_{02} = \zeta \cdot \frac{d^2\zeta}{d\tau^2}\,, \tag{9}$$

$$\eta_{12} = \frac{d\zeta}{d\tau} \cdot \frac{d^2\zeta}{d\tau^2}\,, \tag{10}$$

and so on for higher derivatives.

What we seek, then, is a variational principle of the form (3), but with (4) replaced by

$$s_{12} = \int_1^2 L_C \, d\tau\,, \tag{11}$$

with L_C some real function of the real and imaginary parts of $\eta_{01}, \eta_{02}, \eta_{12}, \ldots$. If we demand that the C_2 analog of proper time given by (11) should depend only on the trajectory linking the two points ζ_1^A, ζ_2^A and not on the choice of parameter τ, then the allowed form of L_C is severely limited. Some Lagrangians L_C that satisfy this constraint are derived in Appendix A. The simplest candidates for L_C are Re (η_{01}), Im (η_{01}), $|\eta_{01}|$, $[\mathrm{Im}\ (\eta_{01}^* \eta_{02})]^{1/3}$, and $|\eta_{12}|^{1/3}$. The first four give trivial theories. In this paper we make the choice

$$L_C = |\eta_{12}|^{1/3}\,. \tag{12}$$

Equations (3) and (12) then lead to the *third*-order equations

$$\frac{d^3\zeta}{ds^3} = 0\,, \tag{13}$$

$$|\eta_{12}| = 1\,, \tag{14}$$

where, after the variation, we have chosen the proper time $s = \int^\tau L_C \, d\tau$ measured from some arbitrary point on the trajectory, as the parameter. From (13) and the definitions (8)–(10) follow immediately the relations

$$\frac{d\eta_{01}}{ds} = \eta_{02}, \quad \frac{d\eta_{02}}{ds} = \eta_{12}, \quad \frac{d\eta_{12}}{ds} = 0\,. \tag{15}$$

We have two constant spinors along the trajectory:

$$u = \frac{1}{2} \eta_{12}^* \frac{d^2\zeta}{ds^2}\,, \tag{16}$$

$$w = \frac{d\zeta}{ds} - \eta_{12}^* \eta_{02} \frac{d^2\zeta}{ds^2}\,. \tag{17}$$

These vectors satisfy the identity $w \cdot u = \frac{1}{2}$, and are consequently linearly independent. Any vector in C_2 may therefore be written as a linear combination of u and w, and in particular we have the identity

$$\zeta = 2[\eta_{12}^* (\eta_{02})^2 - \eta_{01}]u + \eta_{12}^* \eta_{02} w\,. \tag{18}$$

From u and w we can define a constant vierbein of orthonormal SO (1, 3) vectors[14]:

$$V^\lambda = w^\dagger \sigma^\lambda w + u^\dagger \sigma^\lambda u\,, \tag{19}$$

$$C^\lambda = w^\dagger \sigma^\lambda w - u^\dagger \sigma^\lambda u\,, \tag{20}$$

$$A^\lambda = w^\dagger \sigma^\lambda u + u^\dagger \sigma^\lambda w\,, \tag{21}$$

$$B^\lambda = i(w^\dagger \sigma^\lambda u - u^\dagger \sigma^\lambda w)\,, \tag{22}$$

V^λ being future pointing timelike and A^λ, B^λ, and C^λ being spacelike.

Equations (13) and (14) define a ten real parameter family of trajectories that possess some unusual features on account of their satisfying equations whose order is 3 rather than 2. The most notable of these is that one obtains an infinity of geodesics linking any two points ζ_1^A, ζ_2^A, which differ from one another in the initial "slope" $d\zeta/ds$ at ζ_1^A.

III. A NONLINEAR REPRESENTATION OF THE POINCARÉ GROUP

A. Introduction

The full Poincaré group has four disjoint connected pieces, corresponding to the four pieces of the homogeneous Lorentz group O(1, 3). Its general element may be written $\pi = (L, a) \equiv (I, a)\ (L, 0)$, with the multiplication law

$$\pi_1 \pi_2 = (L_1, a_1)\ (L_2, a_2) = (L_1 L_2, L_1 a_2 + a_1)\,. \tag{23}$$

Here $L \in \mathrm{O}\ (1, 3)$ is a 4×4 matrix satisfying $LgL^T g = I$, with I the 4×4 unit matrix, and a a four component column vector whose elements a^λ represent time and space translations. When L is restricted to proper orthochronous Lorentz transformations it may be parametrized $L = \exp\ (g\omega)$, as in (1). If $\pi = (I, a)\ (\exp\ [g\omega], 0)$ ranges over the proper orthochronous Poincaré group, i.e., the normal subgroup containing the identity, then $\pi(g, 0)$, $\pi(-g, 0)$, and $\pi(g, 0)$ $(-g, 0)$ range over the other three pieces of the Poincaré group. The transformation law (1) only applies to the subgroup SO (1, 3) elements $(\exp\ [g\omega], 0)$, and our problem is to find some way of extending (1) to the whole Poincaré group. In particular we need a prescription for translations (I, a), and for the improper elements space reflection $(g, 0)$, and time reversal $(-g, 0)$.

As noted in Sec. I, there is no faithful 2×2 matrix representation of the full Poincaré group. We can, however, find a *nonlinear* representation in C_2 that is associated with transformations between *geodesics* rather than between *points*. Let Ξ represent a geodesic with equation $\xi^A = \zeta^A(s)$, and consider a mapping $\Xi' = D(\Xi)$ from Ξ on to another geodesic Ξ' whose equation is $\xi^A = \zeta'^A(s)$. We could, for example, define $D(\Xi)$ by specifying that the $\zeta'^A(s)$ are some prescribed functions of the real and imaginary parts of $\zeta^B(s)$, $d\zeta^B/ds$, and $d^2\zeta^B(s)/ds^2$, subject to consistency with (13) and (14). Below we shall demonstrate a set of such mappings whose elements $D(\pi, \Xi)$ depend on the Poincaré elements π in such a way as to satisfy the representation conditions

$$D[\pi_1, D(\pi_2, \Xi)] = D(\pi_1\pi_2, \Xi), \tag{24}$$

$$D[(I,0), \Xi] = \Xi, \tag{25}$$

and further, imply (1) when restricted to SO (1, 3). For clarity the form of $D(\pi, \Xi)$ is quoted without proof in what follows, with the lengthy derivations postponed to Appendix B.

B. Representation of translations

For the translation $\pi = (I, a)$ we have

$$\zeta'(s) = \zeta(s) - a_\lambda(V^\lambda + i\alpha C^\lambda)\frac{d\zeta(s)}{ds} + [\gamma a_\lambda(A^\lambda - iB^\lambda) + \tfrac{1}{2}a_\lambda a_\mu(V^\lambda + i\alpha C^\lambda)(V^\mu + i\alpha C^\mu)] \times\frac{d^2\zeta(s)}{ds^2}, \tag{26}$$

where α and γ are arbitrary real constants and the vierbein V^λ, C^λ, A^λ, B^λ, which is independent of s, is defined by (19)–(22). It readily follows from (26) that the constant spinors u and w introduced in (16) and (17) are translation invariant, and hence the vierbein is also. By direct calculation we verifiy that the sequence of translations, a_2^λ followed by a_1^λ, gives the same transformation as the single translation $a_1^\lambda + a_2^\lambda$. Combining (26) with (1) we obtain a representation for all elements of the proper orthochronous Poincaré group. That the representation conditions (24) actually hold for all such elements is shown in Appendix B.

C. Representation of the improper Lorentz transformations

Appropriate transformations for the space reflection $(g, 0)$ and for the time reversal $(-g, 0)$, which satisfy the representation condition (24), are, respectively,

$$\zeta'(s) = \rho^*\epsilon u^* - [(\rho^*)^2 + 2\nu]\eta_{12}\epsilon w^*, \tag{27}$$

$$\zeta'(s') = i(\rho^2 - 2\nu^*)\eta_{12}^*\epsilon u^* + i\rho\epsilon w^*. \tag{28}$$

Here u^* and w^* are the complex conjugates, without transpose, of the constant spinors of (16) and (17), ϵ is the matrix (7), while

$$\rho = \eta_{12}^*\eta_{02}, \tag{29}$$

$$\nu = \eta_{12}^*\eta_{01} - \tfrac{1}{2}\rho^2, \tag{30}$$

Im (ρ) and ν are independent of s. In the time reversed trajectory given by (28) the right-hand side is to be evaluated at $s = -s'$.

D. Charge conjugation

A charge conjugation operation, which is compatible with (1) and (26)–(28), is

$$\mathscr{C}\zeta(s) = i\rho^*u + i(\rho^2 - 2\nu)^*\eta_{12}^*w. \tag{31}$$

If the geodesic $\xi = \zeta(s)$ transforms according to (1) and (26)–(28), then the geodesic $\xi = \mathscr{C}\zeta(s)$ is transformed in the same way.

IV. PARTICLE VARIABLES

Suppose a classical particle of rest mass m and internal angular momentum (spin) $\hbar S$ moves along the M_4 trajectory $x^\lambda = z^\lambda(s)$. The position vector $z^\lambda(s)$, linear momentum p^λ, angular momentum $j^{\kappa\lambda}$, spin angular momentum tensor $S^{\kappa\lambda}$, and the Pauli–Lubanski vector w^λ satisfy the following relations (not all independent):

$$mc\frac{dz^\lambda}{ds} = p^\lambda, \tag{32}$$

$$\frac{d}{ds}(p^\lambda, j^{\kappa\lambda}, S^{\kappa\lambda}, w^\lambda) = 0, \tag{33}$$

$$p^\lambda p_\lambda = (mc)^2, \quad p^\lambda w_\lambda = 0, \tag{34}$$

$$w^\lambda w_\lambda = \tfrac{1}{2}S^{\kappa\lambda}S_{\kappa\lambda} = -\hbar^2 S^2, \tag{35}$$

$$j^{\kappa\lambda} = z^\kappa p^\lambda - z^\lambda p^\kappa + S^{\kappa\lambda}, \tag{36}$$

$$mc\, S^{\kappa\lambda} = \epsilon^{\kappa\lambda\mu\nu}w_\mu p_\nu. \tag{37}$$

Here $\epsilon^{\kappa\lambda\mu\nu}$ is the Levi-Civita permutation symbol with the convention $\epsilon^{0123} = 1$.

Under the Poincaré transformation (L, a), these variables transform according to

$$(z^\lambda)' = L^\lambda_{\ \mu}z^\mu - a^\lambda, \tag{38}$$

$$(p^\lambda)' = \eta L^\lambda_{\ \mu}p^\mu, \tag{39}$$

$$(w^\lambda)' = (\det L)L^\lambda_{\ \mu}w^\mu, \tag{40}$$

$$(j^{\kappa\lambda})' = \eta[L^\kappa_{\ \mu}L^\lambda_{\ \nu}j^{\mu\nu} - (a^\kappa L^\lambda_{\ \mu} - a^\lambda L^\kappa_{\ \mu})p^\mu], \tag{41}$$

$$(S^{\kappa\lambda})' = \eta L^\kappa_{\ \mu}L^\lambda_{\ \nu}S^{\mu\nu}, \tag{42}$$

where η is ± 1 according as $L^0_{\ 0} > 0$ or $L^0_{\ 0} < 0$, and $\det L$ is the determinant of L.

The main result of this paper is that one can construct quantities satisfying all the relations (32)–(42) from ζ, $d\zeta/ds$, $d^2\zeta/ds^2$ along the C_2 geodesics defined in Sec. II. As shown in Appendix B, a mapping that achieves this is

$$p^\lambda = mcV^\lambda, \tag{43}$$

$$z^\lambda(s) = \mathrm{Re}[\rho(V^\lambda + iC^\lambda/\alpha) - \nu(A^\lambda + iB^\lambda)/\gamma], \tag{44}$$

$$w^\lambda = \hbar SC^\lambda, \tag{45}$$

$$j^{\kappa\lambda} = \mathrm{Re}[-i(mc/\gamma)(\nu\sigma^{\kappa\lambda}_{ww} - \nu^*\sigma^{\kappa\lambda}_{uu}) - \{2\hbar S - (mc/\alpha)(\rho - \rho^*)\}\sigma^{\kappa\lambda}_{uw}], \tag{46}$$

$$S^{\kappa\lambda} = \mathrm{Re}[\,-2\hbar S\sigma^{\kappa\lambda}_{uw}\,]\,. \tag{47}$$

Here, $\sigma^{\kappa\lambda}_{ww} = w^T\epsilon\sigma^{\kappa\lambda}w$, $\sigma^{\kappa\lambda}_{uu} = u^T\epsilon\sigma^{\kappa\lambda}u$, $\sigma^{\kappa\lambda}_{uw} = u^T\epsilon\sigma^{\kappa\lambda}w$ with the spin coefficients given by (2). Note that $\epsilon\sigma^{\kappa\lambda}$ is a symmetric matrix.

In summary, (43)–(47) represent a many to one mapping of C_2 geodesics on to particle trajectories in M_4. Poincaré symmetry is preserved, with the transformations (1) and (26)–(28) in C_2 going over to (38)–(42) in M_4. This mapping does not, however, lead to any simple correspondence between the *points* of C_2 and M_4.

APPENDIX A: PARAMETER INVARIANT LAGRANGIANS IN C_2

Suppose that in the parametric equations $\xi = \zeta(\tau)$ of a C_2 trajectory one replaces τ by a new parameter τ', which is some function of τ. Then

$$\frac{d\zeta}{d\tau'} = \tau_1\,\frac{d\zeta}{d\tau}\,,$$

$$\frac{d^2\zeta}{(d\tau')^2} = \tau_1^2\,\frac{d^2\zeta}{d\tau^2} + \tau_2\,\frac{d\zeta}{d\tau},\ldots,$$

where $\tau_n = d^n\zeta/(d\tau')^n$, $n = 1, 2,\ldots$. Whence, in the notation of Sec. II,

$$\eta'_{01} = \tau_1\eta_{01}\,, \tag{A1}$$

$$\eta'_{02} = \tau_1^2\eta_{02} + \tau_2\eta_{01}\,, \tag{A2}$$

$$\eta'_{12} = \tau_1^3\eta_{12}\,, \tag{A3}$$

and so on, where the primes denote that τ' derivatives have been used in the definitions (8)–(10). Equation (A2) implies that

$$(\eta^*_{01}\eta_{02} - \eta_{01}\eta^*_{02})' = \tau_1^3(\eta^*_{01}\eta_{02} - \eta_{01}\eta^*_{02})\,. \tag{A4}$$

Thus (A1), (A3), and (A4) show that any real homogeneous function L_C of degree 1 in the variables η_{01}, η^*_{01}, $(\eta_{12})^{1/3}$, $(\eta^*_{12})^{1/3}$, and $(\eta^*_{01}\eta_{02} - \eta_{01}\eta^*_{02})^{1/3}$ leads to $L_C\,d\tau$ being parameter invariant. Some particular examples of such L_C are $\mathrm{Re}(\eta_{01})$, $\mathrm{Im}(\eta_{01})$, $|\eta_{01}|$, $[\mathrm{Im}(\eta^*_{01}\eta_{02})]^{1/3}$, and $|\eta_{12}|^{1/3}$. Of these, the latter three are also invariant under the phase change, $\zeta\to\zeta\exp(i\phi)$ with ϕ any real constant. More complicated L_C of parameter invariant form may be constructed similarly from higher order derivatives of ζ.

APPENDIX B: DERIVATION OF THE REPRESENTATION OF THE POINCARÉ GROUP

Our aim is to construct functions of ζ and its proper time derivatives, which behave like the classical particle variables $z^\lambda, p^\lambda, w^\lambda, j^{\kappa\lambda}$, and $S^{\kappa\lambda}$ introduced in Sec. IV. It is required that these variables satisfy the identities (34)–(37) and evolve correctly with proper time s according to (32) and (33) as a consequence of the evolution of ζ given by (13) and (14). Further, we seek for ζ, a transformation law under Poincaré transformations that endows these postulated particle variables with the appropriate Minkowski space vector or tensor properties. We solve these problems in what follows by relating the C_2 geodesics to points in an invariant subspace of an eight-dimensional matrix representation of the Poincaré group.

As already noted, there is no faithful linear representation of the Poincaré group in two dimensions that is an extension of the $D^{(\frac{1}{2})0}$ representation of the SO (1, 3) subgroup. Nevertheless the proper orthochronous Poincaré subgroup possesses a four-dimensional representation D^4, that associated with twistors, which on restriction to SO (1, 3) decomposes according to $D^4 = D^{(1/2)0} \oplus (D^{(1/2)0})^*$ (see Refs. 2–7 and 14). Correspondingly the full Poincaré group, including the improper elements, has an eight-dimensional representation D^8, which, for proper orthochronous Poincaré transformations, reduces to $D^4 \oplus (D^4)^*$, and for the subgroup SO(1, 3) reduces further[14] to $D^8 = 2D^{(1/2)0} \oplus 2(D^{(1/2)0})^*$. From previous work on modeling particles of arbitrary mass and spin,[1–14] it is clear that we need at least two twistors or equivalently four $D^{(1/2)0}$ spinors for a satisfactory treatment. In this paper, we adopt D^8 as our basic representation and construct a column vector ψ of eight complex components, which transforms according to D^8. The vector ψ will be formed from functions of the C_2 coordinates ζ^A and their proper time derivatives along the geodesics introduced in Sec. II. Finally the particle variables are given as real bilinear forms in ψ and ψ^* according to the prescription of Ref. 14.

First let us recall the main properties of D^8. We start from the eight-dimensional spinor representation of SO(2, 6) whose 35 generators $M^{ab} \equiv -M^{ba}$ satisfy the commutators

$$[M^{ab}, M^{cd}] = i[\eta^{ad}M^{bc} + \eta^{bc}M^{ad} - \eta^{ac}M^{bd} - \eta^{bd}M^{ac}]\,,$$

where $[\eta^{ab}]$ is the SO(2, 6) metric diag $[1, -1, -1, -1, -1, 1, -1, -1]$ $(a, b, c, d = 0, 1, 2, 3, 5, 6, 7, 8)$. The 8×8 matrices representing M^{ab} are given explicitly in Ref. 14 in terms of three copies of the Pauli matrices, and real, orthogonal, symmetric matrices β and C are given for which

$$(M^{ab})^\dagger = \beta M^{ab}\beta\,,$$

$$(M^{ab})^T = -CM^{ab}C\,.$$

The representation $D^8(\pi)$ of the proper orthochronous Poincaré element $\pi = (\exp[g\omega], a)$ is

$$D^8(\pi) = \exp(-ia_\lambda P^\lambda/\hbar)\exp(-\tfrac{1}{2}i\omega_{\lambda\mu}J^{\lambda\mu})\,, \tag{B1}$$

where the generators of translations and of homogeneous Lorentz transformations are, respectively,

$$P^\lambda = \hbar\kappa(M^{\lambda 5} - M^{\lambda 6})\,,$$

$$J^{\lambda\mu} = \hbar M^{\lambda\mu}\,.$$

Here κ is an arbitrary positive real constant of dimensions $(\text{length})^{-1}$, with different values yielding equivalent representations. Below $m = \hbar\kappa/c$ is identified as the particle mass. Further, the representation D^8 can be extended to the improper Poincaré elements and to charge conjugation.

Suppose now that we have an eight-component column vector ψ that is transformed according to $D^8(\pi)$ given by (B1):

$$\psi' = D^8(\pi)\psi\,. \tag{B2}$$

The 35 real bilinear forms $m^{ab} = \psi^\dagger\beta M^{ab}\psi$ will then transform according to various real representations of the Poincaré group. In particular $m^{57} - m^{67}$, $m^{58} - m^{68}$, and m^{78} are scalars, being unchanged by (B2). Further scalars are $\psi^\dagger\beta\psi$

and $\psi^T C\psi$, the latter being complex. We thus have six real scalars, consistent with the result of Tod[8] that six real-valued Poincaré invariants may be formed from two twistors.

The simplest model for particlelike variables comes from restricting ψ to lie in the invariant subspace

$$m^{57} - m^{67} = 1, \quad m^{58} - m^{68} = 0, \qquad \text{(B3)}$$
$$m^{78} = 0, \quad \psi^T C\psi = 0, \quad \psi^\dagger \beta\psi = -2S,$$

which has ten real dimensions. Here S is a real constant. In this model the position, momentum, angular momentum, and the Pauli–Lubanski spin vector, of a classical particle of rest mass $m = \hbar\kappa/c$ and spin $\hbar S$ are, respectively,

$$z^\lambda = \kappa^{-1} m^{7\lambda}, \quad p^\lambda = \hbar\kappa(m^{\lambda 5} - m^{\lambda 6}), \qquad \text{(B4)}$$
$$j^{\lambda\mu} = \hbar m^{\lambda\mu}, \quad w^\lambda = \hbar m^{\lambda 8}.$$

In Ref. 14 it is shown that these quantities obey (34)–(37) and transform correctly both for proper and for improper Poincaré transformations. Further, if ψ satisfies the evolution equation

$$i\frac{d\psi}{ds} = -\kappa(M^{57} - M^{67})\psi, \qquad \text{(B5)}$$

then the variables (B4) evolve appropriately, satisfying (32) and (33) for mass $m = \hbar\kappa/c$.

Our goal is now to establish a correspondence between D^8 vectors ψ and the C_2 geodesics $\xi = \zeta(s)$. One prescription, certainly not unique, is as follows. Let χ_i, $i = 1, 2, ..., 12$ stand for the real and imaginary parts of the components of ζ, $d\zeta/ds$, and $d^2\zeta/ds^2$. For a given value of the phase constant η_{12} only ten of the χ_i are independent. Likewise any vector ψ constrained to satisfy (B3) depends on only ten real parameters. Let us seek an invertible functional relationship

$$\psi = \psi(\chi_i), \quad \chi_i = \chi_i(\psi, \psi^*), \qquad \text{(B6)}$$

where the dependence on the other parameters, κ, S, and η_{12} is suppressed. We impose two constraints.

(i) The SO (1, 3) transformation properties must be consistent, i.e., (1) must imply (B2) when a_λ is taken to be zero in (B1).

(ii) The evolution equations must be compatible, i.e., (13) and (14) must imply (B5) and vice versa.

Any vector ψ belonging to D^8 may be expressed in partitioned form in terms of four $D^{(1/2)0}$ spinors u, w, v, x:

$$\psi = \begin{bmatrix} u \\ (i\epsilon w)^* \\ v \\ (i\epsilon x)^* \end{bmatrix}. \qquad \text{(B7)}$$

Using the notation of (6), the constraint equations (B3) assume the form[16]

$$w \cdot u = \tfrac{1}{2},$$
$$\mathrm{Im}(w \cdot v) = \mathrm{Im}(u \cdot x) = -\tfrac{1}{2}S, \qquad \text{(B8)}$$
$$u \cdot v + (w \cdot x)^* = 0,$$

and the evolution equations (B5) become

$$\frac{du}{ds} = 0, \quad \frac{dw}{ds} = 0,$$
$$\frac{dv}{ds} = -\kappa u, \quad \frac{dx}{ds} = -\kappa w. \qquad \text{(B9)}$$

Thus u and w are constant spinors, and linearly independent on account of $w \cdot u = \frac{1}{2}$. An obvious choice for this pair of constant spinors is given by (16) and (17), and we make this identification, as anticipated by our notation. It remains to find v and x. These spinors may be written as linear combinations of u and w with complex number coefficients, and the latter then determined in terms of the geodesic variables by substitution in (B8) and (B9). The solution so obtained contains two arbitrary real constants, denoted α and γ:

$$v = \{-iS - \mathrm{Re}[\kappa(1 - i/\alpha)\rho]\}u + (\kappa/\gamma)\nu w,$$
$$x = (\kappa/\gamma)\nu^* u + \{iS - \mathrm{Re}[\kappa(1 + i/\alpha)\rho]\}w, \qquad \text{(B10)}$$

where ρ and ν are as in (29) and (30). Taken together, (16), (17) and (B10) give ψ as a function of the real and imaginary parts of ζ and its first two derivatives along a geodesic.

The inverse relations to (B10), which express ζ in terms of the components of ψ and the constant phase factor η_{12}, are

$$\zeta = \rho w + (\rho^2 - 2\nu)\eta_{12} u, \qquad \text{(B11)}$$

where ρ and ν have the same meaning as before and are given by

$$\kappa\rho = (1 - i\alpha)\mathrm{Re}(u \cdot x) - (1 + i\alpha)\mathrm{Re}(w \cdot v),$$
$$\kappa\nu = -2\gamma u \cdot v. \qquad \text{(B12)}$$

We can now determine the Poincaré transformation properties of ζ from the known relations (B2) for ψ. Written in terms of the constituent two-spinors u, w, v, x, a translation a^λ effects the changes

$$u' = u, \quad w' = w,$$
$$v' = v - \kappa a_\lambda \epsilon(\sigma^\lambda w)^*, \quad x' = x + \kappa a_\lambda \epsilon(\sigma^\lambda u)^*. \qquad \text{(B13)}$$

Thus the pairs (w, v) and (u, x) transform like twistors. Substitution of (B13) into (B11) then yields (26).

The representations of the improper Lorentz transformations and of charge conjugation are found similarly by following the prescription given in Ref. 14 for transforming ψ under these operations. In partitioned form, the effect of space reflection, time reversal, and charge conjugation is to replace (u, w, v, x), respectively, by $i\epsilon(u^*, -w^*, -v^*, x^*)$, $\epsilon(-w^*, u^*, -x^*, v^*)$, and $i(w, u, x, v)$. Applying these transformations to (B11) then yields (27), (28), and (31).

Finally the expressions for the particle variables (43) to (47) are obtained by substituting (B10) into (B4).

[1]J. Rzewuski, Nuovo Cimento 9, 942 (1958).
[2]R. Penrose, J. Math. Phys. 8, 345 (1967); 10, 38 (1969).
[3]R. Penrose, Int. J. Theor. Phys. 1, 61 (1968).

[4] R. Penrose and M. A. H. MacCallum, Phys. Reports C **6**, 243 (1973).
[5] R. Penrose, in *Quantum Gravity*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1975), p. 268.
[6] L. P. Hughston, "Twistors and Particles," in *Lecture Notes in Physics*, Vol. 97 (Springer, Berlin, 1979).
[7] A. Qadir, Phys. Rep. C **39**, 133 (1978); J. Math. Phys. **21**, 514 (1980).
[8] P. Tod, Rep. Math. Phys. **11**, 339 (1977); PhD. dissertation, Mathematical Institute, University of Oxford, U.K., 1975.
[9] Z. Perjés, Phys. Rev. D **11**, 2031 (1975).
[10] J. Rzewuski, J. Math. Phys. **23**, 1573 (1982); **25**, 2136 (1984).
[11] B. Lukács, Z. Perjés, A. Sebestyén, E. T. Newman, and J. Porter, J. Math. Phys. **23**, 2108 (1982).
[12] A. Bette, J. Math. Phys. **25**, 2456 (1984).
[13] P. L. Nash, J. Math. Phys. **21**, 1024, 2534 (1980); **22**, 631, 983 (1981).
[14] G. H. Derrick, Phys. Lett. A **92**, 374 (1982); Int. J. Theor. Phys. **23**, 359 (1984).
[15] *Alphabet conventions*: Capital Latin letters $A, B, C... = 1, 2$. Greek lowercase letters, $\kappa, \lambda, \mu,... = 0, 1, 2, 3$. *Conjugation operations*: A superscript $*$, T, $\dagger$ applied to a vector or matrix denotes, respectively, the complex conjugate, transpose, Hermitian conjugate. *Pauli matrices*:

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

[16] To obtain this form the substitutions $\rho_1 \to \rho_3, \rho_3 \to -\rho_1$ should be made in Eq. (24) of the second paper of Ref. 14. This amount to a unitary transformation of the representation given there.

Bäcklund transformation and the Painlevé property

John Weiss

Institute for Theoretical Physics, University of California, Santa Barbara, California 93106 and Institute for Pure and Applied Physical Science,[a] *University of California, San Diego, La Jolla, California 92093*[b]

(Received 20 June 1985; accepted for publication 14 January 1986)

When a differential equation possesses the Painlevé property it is possible (for specific equations) to define a Bäcklund transformation (by truncating an expansion about the "singular" manifold at the constant level term). From the Bäcklund transformation, it is then possible to derive the Lax pair, modified equations and Miura transformations associated with the "completely integrable" system under consideration. In this paper completely integrable systems are considered for which Bäcklund transformations (as defined above) may not be directly defined. These systems are of two classes. The first class consists of equations of Toda lattice type (e.g., sine–Gordon, Bullough–Dodd equations). We find that these equations can be realized as the "minus-one" equation of sequences of integrable systems. Although the "Bäcklund transformation" may or may not exist for the "minus-one" equation, it is shown, for specific sequences, that the Bäcklund transformation does exist for the "positive" equations of the sequence. This, in turn, allows the derivation of Lax pairs and the recursion operation for the entire sequence. The second class of equations consists of sequences of "Harry Dym" type. These equations have branch point singularities, and, thus, do not directly possess the Painlevé property. Yet, by a process similar to the "uniformization" of algebraic curves, their solutions may be parametrically" represented by "meromorphic" functions. For specific systems, this is shown to provide a natural extension of the Painlevé property.

I. INTRODUCTION

Informally, the Painlevé property requires that solutions (of analytic differential equations) that arise from "good" (analytic, noncharacteristic) data be "meromorphic" (see Refs. 1–10). As described in previous works, when an equation has the Painlevé property, we may construct (auto) Bäcklund transformations by truncating an expansion of the solution about the "movable" singularity manifold at the "constant" level term.

For instance and later reference, the Korteweg–de Vries (KdV) sequence[4]

$$U_t + \frac{\partial}{\partial x} b^{n+1}(U) = 0, \tag{1.1}$$

$$\frac{\partial}{\partial x} b^{n+1} = b^n_{xxx} + 2Ub^n_x + U_x b^n, \tag{1.2}$$

for $n = 0,1,2,\ldots$, has the Bäcklund transformation (BT)

$$U = 4\frac{\partial^2}{\partial x^2}\ln\phi + U_2, \tag{1.3}$$

$$U_2 = -\frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2, \tag{1.4}$$

$$\frac{\phi_t}{\phi_x} + b^n(\{\phi;x\}) = 0, \tag{1.5}$$

where

$$\{\phi;x\} = \frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2 \tag{1.6}$$

is the Schwarzian derivative that is invariant under the Möbius group

$$\phi = (a\psi + b)/(c\psi + a). \tag{1.7}$$

Furthermore,

$$U_3 = \{\psi;x\} \tag{1.8}$$

is a solution of Eq. (1.1) and Eq. (1.5) is invariant under the transformation

$$\phi_x = \psi_x^{-1}. \tag{1.9}$$

We note that

$$b^0 = 1, \quad b^1 = U, \quad b^2 = U_{xx} + \tfrac{3}{2}U^2,$$
$$b^3 = U_{xxxx} + 5UU_{xx} + \tfrac{5}{2}U_x^2 + \tfrac{5}{2}U^3. \tag{1.10}$$

In terms of the "modified" variable

$$V = \phi_{xx}/\phi_x, \tag{1.11}$$

the Lenard formula (1.2) factors into

$$Db^{n+1} = (D - V)D(D + V)b^n, \tag{1.12}$$

where

$$D = \frac{\partial}{\partial x},$$
$$U = V_x - \tfrac{1}{2}V^2 \tag{1.13}$$
$$b^n(U) = b^n(V_x - \tfrac{1}{2}V^2).$$

From Eq. (1.5) this obtains the modified KdV (MKdV) sequence

$$V_t + M_v b^n(V_x - \tfrac{1}{2}V^2) = 0, \tag{1.14}$$

where

$$M_v = D(D - V). \tag{1.15}$$

The KdV/MKdV sequence consists of scalar equations, local in D and first-order in $\partial/\partial t$ (evolution equation). Excepting equations that can be directly transformed into linear systems (i.e., Burgers sequence), the results of this paper argue that a scalar evolution equation with the Painlevé

---

[a] Present address.

[b] Mailing address: Division of Applied Sciences, Harvard University, Cambridge, Massachusetts 02138.

 0022-2488/86/051293-13$02.50

property belongs to either the KdV/MKdV sequence or to the Caudrey–Dodd–Gibbon (CDG)/MCDG sequence.

The CDG sequence is the double sequence[4]

$$U_t + \theta_1 G_n(U) = 0, \quad (1.16)$$

$$A_t + \theta_2 H_n(A) = 0, \quad (1.17)$$

where

$$G_{n+2} = J_1(U)\theta_1(U)G_n,$$
$$H_{n+2} = J_2(A)\theta_2(A)H_n, \quad (1.18)$$
$$G_0 = 1, \quad H_0 = 1,$$
$$G_1 = U_{xx} + \tfrac{1}{4}U^2, \quad H_1 = A_{xx} + 4A^2. \quad (1.19)$$

In terms of the "modified" variables

$$U = W_x - \tfrac{1}{2}W^2, \quad A = V_x - \tfrac{1}{2}V^2, \quad (1.20)$$

$$W = \psi_{xx}/\psi_x, \quad V = \phi_{xx}/\phi_x, \quad (1.21)$$

$$\theta_1 = (D - W)D(D + W),$$

$$J_1 = D^{-1}\left\{\left(D - \frac{W}{2}\right)\left(D + \frac{W}{2}\right)D\left(D - \frac{W}{2}\right)\left(D + \frac{W}{2}\right)\right\}D^{-1}, \quad (1.22)$$

$$\theta_2 = (D - V)D(D + V),$$

$$J_2 = D^{-1}\{(D - 2V)(D - V)D(D + V)(D + 2V)\}D^{-1}. \quad (1.23)$$

The sequence (1.16), (1.17) has the Bäcklund transformation

$$U = 12\frac{\partial^2}{\partial x^2}\ln\phi + U_2,$$

$$A = \frac{3}{2}\frac{\partial^2}{\partial x^2}\ln\psi + A_2, \quad (1.24)$$

where

$$U_2 = -2(\phi_{xxx}/\phi_x),$$

$$A_2 = -\frac{1}{2}\left(\frac{\psi_{xxx}}{\psi_x} - \frac{3}{4}\frac{\psi_{xx}^2}{\psi_x^2}\right), \quad (1.25)$$

$$U_3 = \{\psi;z\}, \quad A_3 = \{\phi;x\} \quad (1.26)$$

are solutions of Eqs. (1.16) and (1.17) and

$$\phi_t/\phi_x + H_n(\{\phi;x\}) = 0,$$

$$\psi_t/\psi_x + G_n(\{\psi;x\}) = 0, \quad (1.27)$$

possess the symmetry

$$\psi_x = \phi_x^{-2}. \quad (1.28)$$

The "modified" sequence is

$$W_t + M_w G_n(W_x - \tfrac{1}{2}W^2) = 0,$$

$$V_t + M_v H_n(V_x - \tfrac{1}{2}V^2) = 0, \quad (1.29)$$

where

$$M_w = D(D - W), \quad M_v = D(D - V).$$

In term of the preceding (and somewhat more) we show in Sec. II that the sine–Gordon equation, which has a Bäcklund transformation, is the minus-one equation of the KdV sequence, while the Bullough–Dodd equation, which is shown not to have a Bäcklund transformation, is the minus-one equation of the CDG sequence. The minus-one equation of the Boussinesq sequence is shown to be equivalent to the equation for the two-dimensional, three-component periodic Toda lattice (of which the Bullough–Dodd equation is a scalar reduction). The minus-one equation of the Hirota–Satsuma sequence is also found to be an equation of Toda lattice type. In view of these results, we propose a method for constructing recursion operators of equation sequences from the annihilators of "minus-one" functionals.

In Sec. III various sequences of "Harry Dym" type equations, which have highly branched (non-Painlevé) singularities, are shown to have a "uniformization" in terms of the KdV sequence. Besides allowing the implicit definition of Bäcklund transformations, this procedure ("uniformization" of integrable systems) provides a natural extension of the Painlevé property. We note various connections with the classical theory of uniformization of algebraic curves (i.e., Schwarzian derivatives, Lax pairs).

In the appendices we present results not directly related to the discussion of Secs. II and III.

In Appendix A we find the Lax pair for the Caudrey–Dodd–Gibbon equation directly from the Bäcklund transformation and without the previously found resummation of terms. Surprisingly, the two methods for finding the Lax pair are nonequivalent.

In Appendix B we consider the factorizations of scalar, linear operators depending on one dependent variable and how this relates to the scalar reductions of the two-dimensional Toda lattice equations.

Finally, we note that the "Caudrey–Dodd–Gibbon" equation[11] has also been studied by Sawada and Kotera[12] (e.g., "Sawada-Kotera" or "Caudrey–Dodd–Gibbon–Kotera–Sawada" equation). Also, the first nontrivial integrals of the "Bullough–Dodd" equation were found in Ref. 13. In Refs. 14 and 15 this equation was then shown to be completely integrable.

II. MINUS-ONE FUNCTIONALS AND THE TWO-DIMENSIONAL TODA LATTICE

To begin, consider the sine–Gordon equation

$$U_{xt} = \sin(U), \quad (2.1)$$

which is equivalent to the equation

$$VV_{xt} - V_x V_t = \tfrac{1}{2}(V^3 - V), \quad (2.2)$$

where

$$V = e^{iu}. \quad (2.3)$$

Equation (2.2) has the Painlevé property[1] and a Bäcklund transformation[5]

$$V = -4\frac{\partial^2}{\partial x \partial t}\ln\phi + V_2, \quad (2.4)$$

where

$$V_2 = \phi_{xt}^2/\phi_x\phi_t, \quad (2.5)$$

$$\Omega_1 = \{\phi;t\} + 2Z_{tt}/Z = \alpha,$$

$$\Omega_2 = \{\phi;x\} + 2W_{xx}/W = \beta, \quad (2.6)$$

$$\alpha\beta = \tfrac{1}{4}.$$

On the other hand, the Bullough–Dodd equation[13–15]

$$U_{xt} = ae^u - be^{-2u} \quad (2.7)$$

is equivalent to the equation

$$VV_{xt} - V_x V_t = -aV + bV^4, \tag{2.8}$$

where $V = e^{-u}$. Equation (2.8) has the Painlevé property with singularities of the form

$$V = \phi^{-1} \sum_{j=0}^{\infty} V_j \phi^j, \tag{2.9}$$

and resonances at $j = -1,2$. The Bäcklund transformation for Eq. (2.8)

$$V = (V_0/\phi) + V \tag{2.10}$$

obtains, with $b = 1$,

$$V_0^2 = \phi_x \phi_t,$$
$$V_1 = -\frac{1}{2}\frac{\phi_{xt}}{V_0} = -\frac{1}{2}\frac{\phi_{xt}}{\phi_x\phi_t} V_0, \tag{2.11}$$

and the following overdetermined system of equations for ϕ:

$$\phi_x \frac{\partial}{\partial x}\Omega_1 + \phi_t \frac{\partial}{\partial t}\Omega_2 = 4a(\phi_x\phi_t)^{1/2},$$
$$\Omega_1\Omega_2 = 0, \tag{2.12}$$

where Ω_1 and Ω_2 are defined by (2.6). From the identity

$$\phi_x \frac{\partial}{\partial x}\Omega_1 = \phi_t \frac{\partial}{\partial t}\Omega_2, \tag{2.13}$$

Eqs. (2.12) have only the trivial solution and, as a consequence, the Bäcklund transformation (2.10) does not exist. This corresponds to the general result that Eq. (2.7) is known not to have a Bäcklund transformation,[14,16] although it does have a Lax pair[14,15] and is completely integrable.

To proceed further we note the following direct formulation of Eq. (2.2) (sine–Gordon equation) in terms of the Schwarzian derivative. With

$$V = \phi_x, \tag{2.14}$$

we find that

$$\frac{\partial}{\partial t}\{\phi;x\} = -\frac{\partial}{\partial x}\left(\frac{1}{\phi_x}\right). \tag{2.15}$$

From the symmetry $V \to 1/V$ of (2.2) we also have

$$V = 1/\psi_x = \phi_x, \tag{2.16}$$

$$\frac{\partial}{\partial t}\{\psi;x\} = -\frac{\partial}{\partial x}\left(\frac{1}{\psi_x}\right). \tag{2.17}$$

The symmetry (2.16) is identical to (1.9) for the Schwarzian formulation (1.5) of the KdV sequence. With reference to the KdV sequence (1.1) Eq. (2.15) is identically

$$U_t + \frac{\partial}{\partial x} b^{-1}(U) = 0, \tag{2.18}$$

where

$$U = \{\phi;x\}, \quad b^{-1} = 1/\phi_x. \tag{2.19}$$

From Eqs. (1.11)–(1.13), the minus-one functional $b^{-1}(U)$, with $U = \{\phi;x\}$, satisfies the condition

$$(D - (\phi_{xx}/\phi_x))D(D + (\phi_{xx}/\phi_x))b^{-1}(U) = 0, \tag{2.20}$$

or

$$b^{-1} = \frac{a}{\phi_x} + b\frac{\phi}{\phi_x} + c\frac{\phi^2}{\phi_x}, \tag{2.21}$$

which obtains Eqs. (2.18) and (2.19) with

$$a = 1, \quad b = c = 0.$$

In other words, the sine–Gordon equation is a specialization of the minus-one KdV equation. A comparison of the respective formulations (2.5), (2.6) and (2.14), (2.15), where the variable ϕ is *not* identified as the same in each, yields

$$\left(\frac{V_{2x}}{V_2} + \frac{V_x}{V}\right) = 2\sigma\left(\frac{V^{1/2}}{V_2^{1/2}} - \frac{V_2^{1/2}}{V^{1/2}}\right), \tag{2.22}$$

where, in Eq. (2.6),

$$\beta = -2\sigma^2.$$

Equation (2.22) is the classical BT for the sine–Gordon equation.[5]

Now, for the Bullough–Dodd equation (2.7), we let

$$e^u = \phi_x, \tag{2.23}$$

and find the equation

$$\frac{\partial}{\partial t}\{\phi;x\} = -\frac{3}{2} b \frac{\partial}{\partial x}\left(\frac{1}{\phi_x^2}\right). \tag{2.24}$$

The substitution

$$e^{-2u} = \psi_x, \tag{2.25}$$

gives us

$$\frac{\partial}{\partial t}\{\psi;x\} = -6a\frac{\partial}{\partial x}\psi_x^{-1/2}. \tag{2.26}$$

With reference to the CDG sequence (1.16), (1.17), we have, for Eqs. (2.24) and (2.26), respectively,

$$A_t + \theta_2 H_{-2}(A) = 0, \tag{2.27}$$
$$U_t + \theta_1 G_{-2}(U) = 0, \tag{2.28}$$

where

$$A = \{\phi;x\}, \quad U = \{\psi;x\}. \tag{2.29}$$

From (1.18)–(1.23) with

$$V = \phi_{xx}/\phi_x, \quad W = \psi_{xx}/\psi_x, \tag{2.30}$$

$$\theta_2 H_{-2}(A) = \frac{\partial}{\partial x}\left\{\frac{1}{\phi_x^2}(a' + b'\phi + c'\phi^2 + d'\phi^3 + e'\phi^4)\right\}, \tag{2.31}$$

$$\theta_1 G_{-2}(U) = \frac{\partial}{\partial x}\{\psi_x^{-1/2}(a' + b'\psi)\}, \tag{2.32}$$

where (a',b',c',d',e') are numerical constants. Thus, Eqs. (2.24) and (2.26) are specializations of the minus (two) CDG equations. From equations (1.25) and (1.26), we conclude that

$$A_3 = \{\phi;x\},$$
$$A_2 = -\frac{1}{2}\left(\frac{\psi_{xxx}}{\psi_x} - \frac{3}{4}\frac{\psi_{xxx}^2}{\psi_x^2}\right), \tag{2.33}$$
$$U_3 = \{\psi;x\},$$
$$U_2 = -2(\phi_{xxx}/\phi_x) \tag{2.34}$$

are solutions of Eqs. (2.27) and (2.28), respectively, and Eqs. (2.24) and (2.26) are connected by the transformation

$$\psi_x = \phi_x^{-2}. \tag{2.35}$$

However, without the invariance under the Möbius group

$$\phi = (a\phi' + b)/(c\phi' + d) \tag{2.36}$$

for Eqs. (2.24) and (2.26), the Bäcklund transformation (1.24) does not exist for Eqs. (2.27) and (2.28).

From the results of Ref. 7 the CDG sequence is a consistent reduction of the Boussinesq sequence. There, the modified Boussinesq sequence is found to be

$$\begin{pmatrix}\theta\\ Z\end{pmatrix}_t = L^n\Omega_2\begin{pmatrix} Z_x + \theta Z\\ -S - \frac{3}{2}Z^2\end{pmatrix}, \tag{2.37}$$

where

$$S = \theta_x - \tfrac{1}{2}\theta^2, \quad \Omega_2 = \begin{pmatrix} D & 0\\ 0 & \frac{1}{3}D\end{pmatrix}, \tag{2.38}$$

$$L = \Omega_2 B^*\Omega_1^{-1}B, \quad \Omega_1^{-1} = \begin{pmatrix} 0 & D^{-1}\\ D^{-1} & 0\end{pmatrix}, \tag{2.39}$$

$$B = \tfrac{1}{3}\begin{pmatrix} D-\theta & 3(D-Z)\\ -(D-2Z)(D-\theta) & D^2 - 3DZ + 3Z^2 + 2(\theta_x - \frac{1}{2}\theta^2)\end{pmatrix}, \tag{2.40}$$

and B^* is the adjoint to B. Letting

$$\theta = \phi_{xx}/\phi_x, \quad Z = \beta_{xx}/\beta_x, \tag{2.41}$$

we find that

$$\begin{pmatrix}\theta\\ Z\end{pmatrix}_t = L^{-2}\Omega_2\begin{pmatrix} Z_x + \theta Z\\ -S - \frac{3}{2}Z^2\end{pmatrix} = L^{-1}\begin{pmatrix}0\\0\end{pmatrix} \tag{2.42}$$

is the equation

$$\frac{\partial}{\partial t}\begin{pmatrix}\phi_{xx}/\phi_x\\ \beta_{xx}/\beta_x\end{pmatrix} = \begin{pmatrix} a\phi_x^{-1/2}\beta_x^{-3/2} - b\phi_x^{-1/2}\beta_x^{3/2} + c\phi_x\\ a\phi_x^{-1/2}\beta_x^{-3/2} + b\phi_x^{-1/2}\beta_x^{3/2}\end{pmatrix}. \tag{2.43}$$

Let

$$c = 2, \quad b = 1, \quad a = -1, \tag{2.44}$$

$$\phi_x = \phi_{1x}/\phi_{3x}, \quad \beta_x = 1/\phi_{2x},$$
$$W_i = \phi_{ixx}/\phi_{ix}, \quad \phi_{1x}\phi_{2x}\phi_{3x} = 1, \tag{2.45}$$

and $\Sigma_1^3 W_i = 0$. This gives us, from Eq. (2.43), the three component Toda equation

$$\frac{\partial}{\partial t}\hat{W} = \begin{pmatrix}\phi_{1x}/\phi_{3x} - \phi_{2x}/\phi_{1x}\\ \phi_{2x}/\phi_{1x} - \phi_{3x}/\phi_{2x}\\ \phi_{3x}/\phi_{2x} - \phi_{1x}/\phi_{3x}\end{pmatrix}. \tag{2.46}$$

With

$$\phi_{ix} = e^{\theta i}, \tag{2.47}$$

this is

$$\theta_{ixt} = e^{\theta i - \theta i - 1} - e^{\theta i + 1 - \theta i}, \tag{2.48}$$

for $i = 1,2,3,...$ (mod N), where

$$\sum^N \theta_{ix} = 0, \quad N = 3.$$

Thus, the three component Toda lattice is the minus-one equation of the Boussinesq sequence. A Bäcklund transformation is known to exist for Eq. (2.48).[14] However, whether a Bäcklund transformation can be constructed for one of the equivalent forms of Eq. (2.48) by the Painlevé method is nearly a moot point. For instance, in Eq. (2.43), let

$$U = \ln(\beta_x^{3/2}\phi_x^{1/2}), \quad \theta = \ln(\beta_x^{3/2}\phi_x^{-1/2}), \tag{2.49}$$

and find the equation

$$U_{xt} = 2ae^{-u} + be^{\theta} + (c/2)e^{u-\theta}, \tag{2.50}$$

$$\theta_{xt} = 2ae^{-u} + 2be^{\theta} - (c/2)e^{u-\theta}. \tag{2.51}$$

With

$$W = e^u, \quad V = e^{-\theta},$$

we find the system

$$V(WW_{xt} - W_xW_t) = 2aVW + bW^2 + (c/2)V^2W^3,$$
$$W(VV_{xt} - V_xV_t) = -aV^2 - 2bVW + (c/2)V^3W^2. \tag{2.52}$$

Equations (2.52) have singularities of the form

$$W = \phi^{-1}\sum W_j\phi^j, \quad V = \phi^{-1}\sum V_j\phi^j, \tag{2.53}$$

with resonances at

$$j = -1,0,1,2. \tag{2.54}$$

The Bäcklund transformation

$$W = W_0\phi^{-1} + W_1, \quad V = V_0\phi^{-1} + V_1, \tag{2.55}$$

taking into account (2.54), produces a system of nine equations for five functions, $(\phi, W_0, V_0, W_1, V_1)$. An analysis of this system indicates that the BT (2.55) determines a reduction of Eq. (2.52):

$$V = \lambda^2 W, \tag{2.56}$$

where

$$\lambda^2 = -b/a. \tag{2.57}$$

The reduced system is Eq. (2.8), the Bullough–Dodd equation, for which a Bäcklund transformation of the form (2.10), (2.55) does not exist. This result is similar to that of Ref. 7, where the BT for the modified nonlinear Shrödinger

(NLS) equations determined a reduction to Burgers equation. The BT's defined by other forms of the singularities for equations (2.52) and equivalent systems have not been investigated, these typically being highly overdetermined and implicit systems of equations.

The situation here contrasts sharply with the analysis for the (positive) Boussinesq sequence. For the (positive) Boussinesq sequence the system of equations produced by the (Painlevé) BT is *not* overdetermined. This is the result of the distribution of the resonances and the linearity of the highest derivative for the positive sequence. Therefore, from the point of view of Painlevé analysis and the calculation of Bäcklund transformations it is of interest to identify when a system is a defined by a negative functional of a sequence of equations. The results of Ref. 17 demonstrate that all the (N-component) two-dimensional, periodic Toda lattice equations can be identified with the minus-one functionals of equation sequences. Therefore, by suitably developing the recursion operators for these sequences it should be possible to recursively define the (Painlevé) BT's. On the other hand, a different approach to Bäcklund transformations and the Painlevé property for the Toda lattice is presented in Ref. 18.

In terms of the variables W_i defined by Eqs. (2.45) for the (modified) Toda equation (2.46), the recursion operator for the Boussinesq sequence assumes a considerably more symmetric form. That is, with

$$\hat{W} = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix}, \tag{2.58}$$

the modified Boussinesq sequence is

$$\hat{W}_t = L^n \hat{W}_x, \tag{2.59}$$

where

$$L = \Omega J,$$

$$\Omega = \begin{pmatrix} D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{pmatrix}, \tag{2.60}$$

$$D = \frac{\partial}{\partial x},$$

and for $i,j = 1,2,3$,

$$J = \{J_{ij}\}, \tag{2.61}$$

$$J_{ii} = -16W_iD - 8W_{ix} - 8W_iD^{-1}A_i - 8A_iD^{-1}W_i, \tag{2.62}$$

$$A_i = (W_{i+1} - W_{i-1})_x - (W_{i+1} - W_{i-1})^2 - 6W_{i+1}W_{i-1}, \tag{2.63}$$

where $i = 1,2,3,\ldots(\text{mod } 3)$. And

$$J_{12} = 8D^2 - 16W_3D - 8B_3 - 8W_1D^{-1}A_2 - 8A_1D^{-1}W_2, \tag{2.64}$$

$$J_{13} = -8D^2 - 16W_2D - 8C_2 - 8W_1D^{-1}A_3 - 8A_1D^{-1}W_3, \tag{2.65}$$

$$J_{23} = 8D^2 - 16W_1D - 8B_1 - 8W_2D^{-1}A_3 - 8A_2D^{-1}W_3, \tag{2.66}$$

$$J_{21} = -J_{21}^*, \quad J_{31} = -J_{13}^*, \quad J_{32} = -J_{23}^*, \tag{2.67}$$

$$\begin{aligned} B_3 &= W_{3x} + (W_1 - W_3)^2 + 3W_1W_3, \\ B_1 &= W_{1x} + (W_1 - W_3)^2 + 3W_1W_3, \\ C_2 &= W_{2x} - W_2^2 + W_1W_3. \end{aligned} \tag{2.68}$$

It is to be remarked that

$$J^* = -J, \tag{2.69}$$

and, by (2.45),

$$\sum_1^3 W_i = 0 \tag{2.70}$$

implies that

$$\sum_{i=1}^3 J_{ij} = 0, \quad \text{for } j = 1,2,3, \tag{2.71}$$

$$\sum_{j=1}^3 J_{ij} = 0, \quad \text{for } i = 1,2,3. \tag{2.72}$$

From Eqs. (2.46) and (2.60), the modified Toda lattice equations are

$$L \circ \hat{W}_t = 0, \tag{2.73}$$

or the recursion operator annihilates the right side of the modified Toda lattice equation (2.46). This suggests that explicit formulas for recursion operators of the Toda lattice equations can be constructed from a suitable system of annihilators expressed in terms of the variables $\{W_i\}$, (2.45). It is not difficult to find formulas for the annihilators [for any N in Eq. (2.48)]. Yet is is nontrivial to verify that the resulting expression is the recursion operator for a sequence. After verification of this procedure it then is necessary to find a transformation analogous to (2.45) for the Boussinesq three-component sequence in which (1) [unlike the Toda formulation (2.46)] all variables allow simultaneous singularities, and (2) a component of the transformed system is invariantly formulated (in terms of the Schwarzian derivative) thereby allowing an analysis similar to that for the Boussinesq sequence.[7] It would be most interesting to resolve the question of Schwarzian formulation through construction of Bäcklund transformations. It is our view that the Schwarzian derivative arises naturally from the essential dependence of the singularities on one preferred (spacelike) independent variable (x), and not, say, from the order of the monodromny group of the associated Lax operator. In this connection the Schwarzian derivative expresses the differential invariance of an equation when subject to the natural (unique) group of conformal transformations preserving the complex sphere (C^1). When the structure of an equation's singularities depends essentially on more than one complex independent variable various generalizations of the Schwarzian derivative are indicated.

Finally, the minus-one equation of the Hirota–Satsuma[5,19] sequence can be shown to be equivalent to the system

$$\begin{aligned} A_{xt} &= ae^{A-B} + be^{-A-B}, \\ B_{xt} &= -ae^{A-B} + be^{-A-B} - e^B \end{aligned} \tag{2.74}$$

of Toda type. Equation (2.74) is reduction of the four-component Toda lattice presented in Ref. 14. With reference to Eq. (2.43) the minus-one Boussinesq equation can be written as

$$C_{xt} = ae^{C-D} + be^{-C-D},$$
$$D_{xt} = -ae^{C-D} + be^{-C-D} + ce^{2D}, \quad (2.75)$$

where

$$C = \ln(\beta_x^{3/2}), \quad D = \ln(\phi_x^{1/2}). \quad (2.76)$$

III. UNIFORMIZATION OF THE HARRY DYM SEQUENCE

With reference to Eqs. (1.1)–(1.9), the KdV equation

$$U_t + \frac{\partial}{\partial x}(U_{xx} + \tfrac{3}{2}U^2) = 0 \quad (3.1)$$

has the Bäcklund transformation

$$U = 4\frac{\partial^2}{\partial x^2}\ln\phi + U_2, \quad (3.2)$$

where

$$U_2 = -\frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2, \quad (3.3)$$

and

$$\phi_t/\phi_x + \{\phi;x\} = \lambda. \quad (3.4)$$

The Schwarzian derivative

$$\{\phi;x\} = \frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2 \quad (3.5)$$

is invariant under the Möbius group

$$\phi = (a\psi + b)/(c\psi + d), \quad (3.6)$$

and it can be shown[4] that Eq. (3.4) is invariant under (3.6) and the transformation

$$\phi_x = \beta_x^{-1}. \quad (3.7)$$

Furthermore,

$$\{\phi;x\} = \{\psi;x\} + \{f;\psi\}\psi_x^2, \quad (3.8)$$
$$\{\phi;x\} = h'^2\{\psi;z\} + \{h;x\},$$

where

(i) $\phi = f(\psi)$

is an arbitrary change of dependent variable, and

(ii) $z = h(x)$

is an arbitrary change in independent variable. By the transformation properties of the Schwarzian derivative,

$$\{\phi;x\} = -\phi_x^2\{x;\phi\}. \quad (3.9)$$

Under the change of variables

$$x \to \phi, \quad t \to t, \quad \phi \to x, \quad (3.10)$$

$$\phi_x = 1/x_\phi, \quad (3.11)$$

and

$$x_t = -\phi_t/\phi_x.$$

Equation (3.4) becomes[2]

$$x_\phi^2 x_t + \lambda x_\phi^2 + \{x;\phi\} = 0, \quad (3.12)$$

which is invariant under the change of independent variable (3.6). Equation (3.12) is equivalent to the Harry Dym equation.[2] Equation (3.12) is also invariant under the transformation (implicit change of independent variable) (3.7) or

$$x_\phi x_\beta = 1. \quad (3.13)$$

From (3.13)

$$\phi_\beta = x_\beta/x_\phi = x_\beta^2, \quad (3.14)$$
$$\beta_\phi = x_\phi^2,$$

and using (3.8) or directly from (3.12)

$$\beta_t + \beta_\phi^{-1/2}\{\beta;\phi\} = 0. \quad (3.15)$$

By the involution (3.13) or directly,

$$\phi_t + \phi_\beta^{-1/2}\{\phi;\beta\} = 0. \quad (3.16)$$

Therefore, under the inversion

$$(\phi,\beta) \to (\beta,\phi), \quad (3.17)$$

Eqs. (3.15) and (3.16) are invariant. However, unlike Eq. (3.12), Eq. (3.15) is not invariant under the Möbius group (3.6). The equivalent Bäcklund transformation for (3.15) is defined by

$$h_\phi = \phi^2\beta_\phi, \quad \phi = -1/\psi, \quad (3.17')$$

where it can be shown that $h = h(\psi)$ satisfies

$$h_t + h_\psi^{-1/2}\{h;\psi\} = 0. \quad (3.17'')$$

By composition of the BT (3.17) and (3.17′) various sequences of solutions can be constructed. Equation (3.15) may also be written as

$$\beta_t = 2\frac{\partial^2}{\partial\phi^2}(\beta_\phi^{-1/2}), \quad (3.18)$$

and

$$W = \beta_\phi = x_\phi^2 \quad (3.19)$$

obtains the Harry Dym equation

$$W_t = 2\frac{\partial^3}{\partial\phi^3}W^{-1/2}. \quad (3.20)$$

Equations (3.12), (3.15), and (3.20) all allow movable branch point singularities (even logarithmic singularities) and thus do not directly possess the Painlevé property. For instance, Eq. (3.12) has singularities of the form

$$x = x_0\epsilon^{1/3} + x_1\epsilon + x_2\epsilon^{5/3} + \cdots, \quad (3.21)$$

where

$$\epsilon = \epsilon(\phi,t)$$

represents the singularity manifold. For this reason, it is not possible to directly calculate BT's for, say, Eq. (3.15) by an expansion about the singular manifold. However, solutions of Eqs. (3.15) and (3.16) can be represented implicitly through solutions of Eq. (3.4) that satisfy the condition (3.7). From the rational solutions of Eq. (3.4),[4] the expressions

(i) $\phi = 1/x, \quad \beta = x^3/3 + 4t, \quad (3.22)$

(ii) $\phi = \dfrac{1}{x^3/3 + 4t}, \quad \beta = \dfrac{x^6 + 60tx^3 - 720t^2}{5x} \quad (3.23)$

implicitly define solutions of (3.15) and (3.16). In the above, (ϕ,β) are meromorphic functions of a "uniformizing" variable x (and t). In general, Eq. (3.15) [or (3.16)] has a uniform representation

$$\phi = f(x,t), \quad \beta = g(x,t), \quad (3.24)$$

where f and g satisfy Eq. (3.4) and

$$f_x g_x = 1. \quad (3.25)$$

Since Eqs. (3.15) and (3.16) are invariant under inversion (3.17) of independent/dependent variables, the uniformizing functions (f,g) satisfy the same differential equation. For an arbitrary differential equation it would be necessary to make a simultaneous substitution of dependent and independent variables subject to the requirement that the "uniformizing" substitutions be "meromorphic" functions of the uniform variable(s). The resulting set of differential equations (possessing the Painlevé property) may then be studied directly by use of the Bäcklund transformations.

The KdV sequence [Eq. (1.5)] is

$$\phi_t/\phi_x + b^n(\{\phi;x\}) = 0, \quad (3.26)$$

where

$$D_x b^{n+1} = (D_x^3 + 2UD_x + U_x)b^n, \quad b^n = b^n(U), \quad (3.27)$$

$$D_x = \frac{\partial}{\partial x}, \quad b^0 = 1, \quad b^1 = U, \quad b^2 = U_{xx} + \tfrac{3}{2}U^2. \quad (3.27')$$

With

$$V = \phi_{xx}/\phi_x, \quad U = V_x - \tfrac{1}{2}V^2 = \{\phi;x\}, \quad (3.28)$$

the Lenard formula (3.27) is

$$D_x b^{n+1} = (D_x - V)D_x(D_x + V)b^n \quad (3.29)$$

or

$$b^{n+1} = L_v \circ b^n, \quad (3.30)$$

where

$$L_v = D_x^{-1}(D_x - V)D_x(D_x + V).$$

From the identification (3.28),

$$L_v = D_x^{-1}\phi_x D_x \phi_x^{-1} D_x \phi_x^{-1} D_x \phi_x, \quad (3.31)$$

and, by (3.26).

$$\phi_t/\phi_x + L_v^n \circ 1 = 0, \quad (3.32)$$

where $L_v^n = L_v \circ L_v^{n-1}$ and (3.32) are invariant under (3.6) and (3.7). Equations (3.32) can be written as "Hamiltonian" systems

$$\phi_t + M_v^n \circ \phi_x = 0, \quad (3.33)$$

where

$$M_v = \Omega_1 J_1, \quad (3.34)$$

$$\Omega_1 = \phi_x D_x^{-1} \phi_x, \quad J_1 = D_x \phi_x^{-1} D_x \phi_x^{-1} D_x. \quad (3.35)$$

Under the change of variable (3.10) and (3.11), and using

$$D_\phi = x_\phi D_x = \phi_x^{-1} D_x, \quad (3.36)$$

we find

$$x_t = L_\phi^n \circ 1, \quad (3.37)$$

where

$$L_\phi = \Omega_2 J_2, \quad (3.38)$$

$$\Omega_2 = D_\phi^{-1}, \quad J_2 = x_\phi^{-1} D_\phi^3 x_\phi^{-1}. \quad (3.39)$$

Equations (3.37) are Hamiltonian and invariant under (3.6) and (3.13). From (3.13) and (3.14),

$$x_\phi = \beta_\phi^{1/2}, \quad (3.40)$$

and

$$\beta_t = \Phi_\phi^n \circ 1, \quad (3.41)$$

$$\Phi_\phi = \Omega_3 J_3, \quad (3.42)$$

$$\Omega_3 = D_\phi, \quad J_3 = D_\phi \beta_\phi^{-1/2} D_\phi^{-1} \beta_\phi^{-1/2} D_\phi. \quad (3.43)$$

Under the inversion (3.17), the identity

$$\beta_\phi \Phi_\beta = \Phi_\phi \beta_\phi \quad (3.44)$$

implies

$$\phi_t = \Phi_\beta^n \circ 1, \quad (3.45)$$

or the sequence (3.41) is invariant under inversion (3.17) [and also under (3.17′)]. The Harry Dym sequence

$$W = \beta_\phi, \quad W_t = H_\phi^n \circ 1, \quad (3.46)$$

where

$$H_\phi = \Omega_4 J_4, \quad (3.47)$$

$$\Omega_4 = D_\phi^3, \quad J_4 = W^{-1/2} D_\phi^{-1} W^{-1/2}, \quad (3.48)$$

is Hamiltonian and, by the above identification, is uniformly represented by the KdV sequence (3.33).

A similar procedure for the Caudrey–Dodd–Gibbon sequence (1.27) obtains the sequence

$$x_t = M_\phi^n \circ 1, \quad (3.49)$$

where

$$M_\phi = J_2 \theta_2, \quad J_2 = D_\phi^{-1} x_\phi^{-2} D_\phi^5 x_\phi^{-2} D_\phi^{-1}, \quad \theta_2 = x_\phi^{-1} D_\phi^3 x_\phi^{-1}, \quad (3.50)$$

and J_2, θ_2 are identified by Eqs. (1.23). We note the relative simplicity of the recursion operators (3.38) and (3.50) for the "automorphic" sequences (3.37) and (3.49) [the term "automorphic" indicating the invariance of these equations under the Möbius group (3.6)].

The classical theory of uniformization by automorphic functions is concerned with algebraic functions (Riemann surfaces) defined by polynomial in two (complex) variables[20,21]

$$P(z,w) = 0, \quad (3.51)$$

where P is irreducible (not the product of two polynomials of lower degree). Equation (3.51) defines (in general) w as the multiple-valued function of z (or z as a multiple-valued function of w); for instance, when

$$z^2 + w^2 = 1, \quad w = \pm(1 - z^2)^{1/2}. \quad (3.52)$$

Yet is is also possible to represent the solutions of (3.52) as

$$\text{(i)} \quad z = \sin t, \quad w = \cos t, \quad (3.53)$$

or

$$\text{(ii)} \quad z = 2t/(1 + t^2), \quad w = (1 - t^2)/(1 + t^2), \quad (3.54)$$

where in both cases (z,w) are single-valued functions t [in (3.54) rational functions of t]. Equations (3.53) and (3.54) define uniformizations of Eq. (3.52), i.e., uniform, single-valued.

A general result[20] is an algebraic function (3.51) can be uniformized by means of (i) rational functions if the genus of (3.51) is zero ($p = 0$), (ii) elliptic functions if the genus is one ($p = 1$), and (iii) Fuchsian functions of the first kind when the genus is greater than one ($p > 1$).

A Fuchsian function is a meromorphic function that is (automorphic) invariant under a "discrete" subgroup of the Möbius group. The discrete (properly discontinuous) subgroup is required to preserve (carry into itself) a circle (the "principal" circle) and map interior points (of the circle) into interior points and exterior points into exterior points. That is, if M is the discrete subgroup

$$M \subset M_{0b}$$

and the principal circle is $|t| = 1$,

$$t' = (at + b)/(ct + d) \in M,$$

then $f(t') = f(t)$, where

$$|t| < 1 \Leftrightarrow |t'| < 1,$$
$$|t| = 1 \Leftrightarrow |t'| = 1,$$

and $f(t)$ is meromorphic in $|t| < 1$.

Finally, a Fuchsian function of the first kind is a Fuchsian function for which the limit points of the associated group are dense on the principal circle, and hence, the principal circle is a natural boundary for the function. For a discussion of the above, the reader is advised to consult Refs. 20 and 21. In general, the theory of uniformization, especially as formulated by Teichmuller, Alfors, and Bers[22–24] determines the existence of various forms of uniformizations for general classes of Riemann surfaces. Yet, actual uniformizations for Eqs. (3.51) are known in only a few special cases.[25] That is, representations of the group and the associated automorphic functions have been calculated for only a finite number of equations of the form (3.51). Of special interest in this regard is Whittaker's conjecture, our account of which is taken from Ref. 25.

Whitaker considered the hyperelliptic equations

$$w^2 = (z - e_1)(z - e_2) \cdots (z - e_{2p+2}) = f(z), \tag{3.55}$$

where the $2p + 2$ $\{e_j\}$ are distinct complex numbers. It is first shown that the group Γ of Eq. (3.55) is a subgroup of index 2 in a larger group Γ^*, in which the fundamental region for Eq. (3.55) has genus zero. Therefore, by a general result, any automorphic function attached to Γ^* is a rational function of one such automorphic function $z(t)$. The Schwarzian derivative of an automorphic function is itself an automorphic function and therefore must be a rational function of z. That is,

$$\{t;z\} = -z_t^{-2}\{z;t\} = R(z) \tag{3.56}$$

is automorphic (in t) and it can be shown that

$$R(z) = \frac{3}{8}\left[\sum_{n=1}^{2p+2} \frac{1}{(z - e_n)^2} - \frac{g(z)}{f(z)}\right], \tag{3.57}$$

where $f(z)$ is defined by (3.55) and $g(z)$ is of the form

$$g(z) = 2(p+1)z^{2p} - 2p\sum_{n=1}^{2p+2} e_n z^{2p-1} + \sum_{k=0}^{2p-2} C_k z^{2p-2-k}. \tag{3.58}$$

The $2p - 1$ coefficients $\{C_k\}$ are the accessory parameters and are determined by the group Γ^*.

If $R(z)$ is determined, then it is known that by the substitution

$$t = (\eta_1/\eta_2)(z), \tag{3.59}$$

Eq. (3.56) becomes the linear differential equation

$$\frac{d^2\eta}{dz^2} + \frac{1}{2}R(z)\eta = 0, \tag{3.60}$$

and Γ^* is the monodromny group of (3.60). Therefore to find Γ^* (and Γ) and $z(t)$ (the uniformization) it is necessary to determine $R(z)$ (the accessory parameters) $\{C_k\}$. Motivated by certain scaling arguments, Whittaker conjectured that

$$R(z) = \frac{3}{8}\left[\left(\frac{f'(z)}{f(z)}\right)^2 - \frac{(2p+2)f''(z)}{(2p+1)f(z)}\right]. \tag{3.61}$$

It can be shown that if the $\{C_k\}$ are polynomial functions of $\{e_n\}$ then (3.61) is correct for the general hyperelliptic equation (3.55).

As is indicated by the above, the Schwarzian derivative, through its connection with the automorphic functions, arises in many problems related to the study of Riemann surfaces.[26,27] For instance, the space of moduli (accessory parameters) of Riemann surfaces, the so called universal Teichmuller space, is equivalent to a Banach space of bounded Schwarzian derivatives of univalent functions.[23] In a different direction, Burchnall and Chaundy[28] have shown that a commuting pair of linear differential operators

$$PQ = QP, \tag{3.62}$$

where $\deg(P) = m$, $\deg(Q) = n$, satisfy an algebraic identity

$$f(P,Q) = 0, \tag{3.63}$$

with constant coefficients, where f is of degree n in P and of degree m in Q. They remark that (3.62) and (3.63) provide a form of uniformization for the algebraic curve $f(z,w) = 0$. Since the work of Burchnall and Chaundy is the forerunner of later developments in the theory of integrable systems (Lax pairs), there may be connections between the classical (and recent) theory of uniformization (Riemann surfaces, automorphic functions) and the theory of integrable systems (isomonodromny deformations, finite zone potentials).

Now, with regard to the uniformization of the Harry Dym sequences (3.37), (3.41), and (3.46) by the (Schwarzian) KdV sequence (3.32), we remark that, since ϕ is meromorphic as a function of x and x is (loosely speaking) "automorphic" with regard to ϕ [Eqs. (3.37) being invariant under the Möbius group], the formulation is not standard in terms of the theory of uniformization (where the meromorphic and automorphic dependence refer to the same variable). We can, of course, claim that ϕ is automorphic with regard to the identity group acting on the independent variable x but this somehow seems to miss the point. Instead, it seems worthwhile to enquire whether Eqs. (3.37) allow solutions that are automorphic (invariant under a discrete subgroup of the Möbius group acting on ϕ) without requiring meromorphicity.

Toward this goal, consider Eq. (3.12), where without loss of generality, $\lambda = 0$,

$$x_t + x_\phi^{-2}\{x;\phi\} = 0, \tag{3.64}$$

which is equivalent to

$$x_t = D_\phi^{-1} x_\phi^{-1} \frac{\partial^3}{\partial\phi^3} x_\phi^{-1}, \tag{3.65}$$

where

$$D_\phi = \frac{\partial}{\partial\phi}.$$

The first two conserved quantities of Eq. (3.64) are

$$C_1 = \oint x_\phi^2 \, d\phi, \qquad C_2 = \oint \frac{x_{\phi\phi}^2}{x_\phi^3} \, d\phi. \tag{3.66}$$

From the recursion operator for the sequence (3.37),

$$L_\phi = \Omega_2 J_2, \quad \Omega_2 = D_\phi^{-1}, \quad J_2 = x_\phi^{-1} D_\phi^3 x_\phi^{-1}, \tag{3.67}$$

the functional gradients of the conserved quantities b^j satisfy the equation

$$b^{j+1}(x_\phi) = M_\phi b^j(x_\phi), \tag{3.68}$$

where

$$M_\phi = J_2 \Omega_2, \quad b^0 = 0. \tag{3.69}$$

The Lax pair for Eqs. (3.64) and (3.65) is

$$Y_{\phi\phi} = \lambda x_\phi^2 Y, \qquad Y_t = -4\lambda x_\phi^{-1} Y_\phi + 2\lambda (x_\phi^{-1})_\phi Y. \tag{3.70}$$

Equations (3.70) have a curious property in that if $Y(\phi)$ is a solution, then

$$W(\epsilon) = \epsilon Y(1/\epsilon) \tag{3.71}$$

is a solution of

$$W_{\epsilon\epsilon} = \lambda x_\epsilon^2 W, \qquad W_t = -4\lambda x_\epsilon^{-1} W_\epsilon + 2\lambda (x_\epsilon^{-1})_\epsilon W, \tag{3.72}$$

where

$$\epsilon = 1/\phi. \tag{3.73}$$

Therefore, from (3.71), (3.73), and the invariance of Eqs. (3.70) under scaling and translation,

$$\phi \to a\phi, \quad \phi \to \phi + b, \tag{3.74}$$

it is possible to observe the effect of the Möbius group on the eigenfunctions of (3.70).

Next we recall that equations (3.64) and (3.65) have movable singularities of the form

$$\text{(i)} \quad x = \psi^{1/3} \sum_{j=0}^{\infty} x_j \psi^{j/3}, \tag{3.75}$$

with resonances at

$$j = -3,2,4$$

and

$$\psi = \psi(\phi,t), \quad \psi_\phi \neq 0.$$

There are also singularities of the form

$$\text{(ii)} \quad x = x_0\phi^{-1} + x_1 + x_2\phi + x_3\phi^2 + x_4\phi^3 + x_5 t\phi^4 + \cdots, \tag{3.76}$$

and when $m > 1$

$$\text{(iii)} \quad x = x_0\phi^{-m} + \cdots + x_{3m} t\phi^{2m} + \cdots, \tag{3.77}$$

where the $\{x_j\}$ are constants. Furthermore, it can be shown that the singularities (3.77) have an expansion of the form

$$x = \phi^{-m} \sum_{k=0}^{\infty} P_k(\phi) t^k \phi^{3km}, \tag{3.78}$$

where

$$P_k(\phi) = P_{0k} + P_{1k}\phi + \cdots + P_{3m-1,k}\phi^{3m-1}. \tag{3.79}$$

The locations of the singularities (3.76) and (3.77) do not depend on t but are movable, since the locations are not fixed by the equation. By the invariance of Eq. (3.64) under translation in ϕ, the singularities may be located at any point $\phi = \phi_0$, where ϕ_0 is constant. Also the singularities (3.76) and (3.77) refer to the behavior of ϕ as $x \to \infty$.

To obtain an automorphic solution it is consistent to set

$$x(\phi,0) = x_0(\phi),$$

where, say,

$$z_0 = -\phi^{-1} + \phi. \tag{3.80}$$

By the invariance of Eqs. (3.64) and (3.80),

$$x(\phi,t) = x(\epsilon,t), \tag{3.81}$$

where

$$\phi = -1/\epsilon. \tag{3.82}$$

With (3.80), the scattering operator in the Lax pair (3.70) has an irregular singular point at $\phi = 0$ and at $\phi = \infty$. Therefore, it is not (entirely) obvious how to implement (3.70) to find the solution (3.81). However, we claim that, with automorphic x, it is natural to study solutions of (3.70) on the boundary of a "fundamental domain" of the automorphism group. In this case, the boundary of the fundamental domain is, by (3.82),

$$|\phi| = 1. \tag{3.83}$$

Using (3.80) and extending the path of integration along the contour (3.83), it is found, by the residue theorem applied to (3.66), that

$$C_1 = C_2 \equiv 0. \tag{3.84}$$

In the same way all the conserved qualities of Eq. (3.64) can be evaluated.

On the other hand, the KdV data, defined by (3.80), is

$$U_0 = \{\phi;x\} = -6/(x^2+4)^2, \tag{3.85}$$

which is within the class of initial data for which the inverse scattering on the real line can be solved. That is

$$\phi = V_1/V_2, \tag{3.86}$$

where (V_1,V_2) satisfy

$$V_{xx} = (\lambda - U_0/2)V. \tag{3.87}$$

We note that the irregular singularities of (3.70) refer to the behavior of ϕ (and the KdV data) as $x \to \infty$. We shall defer further consideration of this problem at this time, except to note that (3.87) with U_0 defined by (3.85) (1) does not support solitons, and (2) can be solved in terms of hypergeometric functions when $\lambda = 0$.

As an example of a system with finite-degree branch points, the equation

$$U_t + U^3 U_x + U_{xxx} = 0 \tag{3.88}$$

has singularities of the form

$$U = \psi^{-2/3} \sum_{j=0}^{\infty} U_j \psi^{j/3}, \tag{3.89}$$

with resonances at $j = -3, 8, 10$ and is known to be nonintegrable[29] (has only three conserved quantities). Since the traveling wave solution

$$U + U(x + Ct) \tag{3.90}$$

can be found (as a hyperelliptic function) the presence of a nontrivial time dependence probably introduces additional singularities along the systems characteristics. On the other hand, in Ref. 30 (and the references cited therein) many examples of integrable systems with branch point behavior are found.

ACKNOWLEDGMENTS

At the Institute for Theoretical Physics this research was supported in part by the National Science Foundation under Grant No. PHY82-17853, supplemented by funds from the National Aeronautics and Space Administration; and the Air Force Office of Scientific Research Grant No. AFOSR 84-0128 at the Institute for Pure and Applied Physical Science.

APPENDIX A: THE CAUDREY–DODD–GIBBON EQUATION RECONSIDERED

In Ref. 4 we have studied the Bäcklund transformation

$$U = \frac{\partial^2}{\partial x^2} \ln \phi + U_2 \tag{A1}$$

for the Caudrey–Dodd–Gibbon equation

$$U_t + \frac{\partial}{\partial x}(U_{xxxx} + 30 U U_{xx} + 60 U^3) = 0. \tag{A2}$$

The resulting overdetermined system of equations for (ϕ, U_2) is

$$\text{(i)}\ \frac{\phi_t}{\phi_x} + \frac{\partial^2}{\partial x^2}\{\phi;x\} + 4\{\phi;x\}^2 + 5(\theta_{xx} + \theta^2 + 2\{\phi;x\}\theta) = 0, \tag{A3}$$

$$\text{(ii)}\ \theta\theta_{xx} - \theta_x^2/2 + \tfrac{2}{3}\theta^3 + \{\phi;x\}\theta^2 = \lambda, \tag{A4}$$

$$\text{(iii)}\ 6U_2 + \tfrac{3}{2}(\phi_{xx}^2/\phi_x^2) = \theta - \{\phi;x\}, \tag{A5}$$

or

$$6U_2 + \phi_{xxx}/\phi_x = \theta. \tag{A6}$$

In Ref. 4, due to a certain complexity in the analysis, we set

$$\theta = \lambda = 0 \tag{A7}$$

and found the Lax pair for Eq. (A2) in the instance when the spectral parameter vanishes. Later, the complete Lax pair was derived through a certain resummation of terms defined by (A1). In effect, the Lax pair was not derived directly from the BT (A1). In this appendix we will derive the Lax pair directly from the BT [i.e., Eqs. (A3)–(A6)], thereby showing that a resummation of terms is unnecessary. We present this result for the sake of completeness since, as shown earlier, the minus-one equation of the Caudrey–Dodd–Gibbon sequence does not have a Bäcklund transformation. The point here is that, in spite of this result, everything goes through without restriction for the CDG equation.

The first step is to "solve" Eq. (A4). Therefore, let

$$\lambda = -2\sigma^2, \tag{A8}$$

$$\theta = \sigma(\beta/\beta_x), \tag{A9}$$

and find that Eq. (A4) is

$$\{\phi;x\} + \tfrac{2}{3}\sigma(\beta/\beta_x) = \{\beta;x\}. \tag{A10}$$

From the formula for the Schwarzian derivative, if

$$\phi = V_1/V_2, \tag{A11}$$

$$\Phi_x = V_2 V_{1x} - V_1 V_{2x}, \tag{A12}$$

then

$$\{\phi;x\} = \{\Phi;x\} + 2(V_{2x}V_{1xx} - V_{1x}V_{2xx})/\Phi_x. \tag{A13}$$

Comparing the identity (A13) with (A10) suggests that

$$\beta_x = \Phi_x = V_2 V_{1x} - V_1 V_{2x}, \tag{A14}$$

and

$$(\sigma/3)\beta = V_{1x}V_{2xx} - V_{2x}V_{1xx}. \tag{A15}$$

From (A14) and (A15)

$$V_{1x}V_{2xxx} - V_{2x}V_{1xxx} = (\sigma/3)(V_2 V_{1x} - V_1 V_{2x}), \tag{A16}$$

which we solve by requiring that (V_1, V_2) satisfy

$$V_{xxx} = aV_x + (\sigma/3)V. \tag{A17}$$

From (A14)–(A17)

$$\beta_{xxx} = a\beta_x - (\sigma/3)\beta. \tag{A18}$$

Now, let

$$V_t = cV_{xx} + dV_x + eV. \tag{A19}$$

Then, substituting for (θ, ϕ) [using (A9)–(A11) and (A14)–(A18)] in Eq. (A3) and requiring that $V_{txxx} = V_{xxxt}$, gives us

$$c = 3(a_x + \sigma), \quad d = -a_{xx} - a^2, \quad e = -2\sigma a, \tag{A20}$$

and

$$a_t + \frac{\partial}{\partial x}\left(a_{xxxx} - 5aa_{xx} + \frac{5}{3}a^3\right). \tag{A21}$$

Equation (A21) is the CDG equation (A2) with the identification

$$a = -6U \tag{A22}$$

and Eqs. (A17) and (A19) are the Lax pair for Eq. (A21).

Finally, from equations (A6), (A9), and (A14),

$$a = -6U_2 + 6\frac{\partial^2}{\partial x^2}\ln V_2$$

$$= -6U + 6\frac{\partial^2}{\partial x^2}\ln V_1 \qquad (A23)$$

[by (A1) and (A11)]. It is curious that, in the resummation used earlier[4] to calculate the Lax pair, ϕ was directly an eigenfunction, while here it is a ratio of eigenfunctions (A11).

By a further analysis it can be shown that, with (A14), (A15), (A17), and (A19),

$$\beta_t = 3(a_x - \sigma)\beta_{xx} - (a_{xx} + a^2)\beta_x + 2\sigma a\beta, \qquad (A24)$$

which is Eq. (A19) with $\sigma \to -\sigma$. From (A18) and (A24), β is an eigenfunction of the Lax pair (with $\sigma \to -\sigma$).

APPENDIX B: FACTORIZATION OF SCALAR OPERATORS AND THE SCHWARZIAN DERIVATIVE

It is well known that the Schrödinger operator

$$D^2 + \tfrac{1}{2}U = (D - \tfrac{1}{2}V)(D + \tfrac{1}{2}V), \qquad (B1)$$

where

$$U = V_x - \tfrac{1}{2}V^2, \quad D = \frac{\partial}{\partial x}, \qquad (B2)$$

with

$$V = \phi_{xx}/\phi_x, \qquad (B3)$$

$$U = \{\phi;x\}. \qquad (B4)$$

From Eqs. (1.11) and (1.12), the third-order operator in the Lenard formula is

$$D^3 + 2UD + U_x = (D - V)D(D + V), \qquad (B5)$$

where (V,U) are defined by (B3) and (B4).

We claim that the operator sequence

$$L_{n+1} = \prod_{j=0}^{n}\left[D + \left(j - \frac{n}{2}\right)V\right] = \left(D - \frac{n}{2}V\right)\cdots\left(D + \frac{n}{2}V\right), \qquad (B6)$$

where $n = 0,1,2,3,\ldots$, and

$$V = \phi_{xx}/\phi_x, \qquad (B7)$$

defines operators whose coefficients, when expanded, can be expressed entirely in terms of the Schwarzian derivative

$$U = V_x - \tfrac{1}{2}V^2 = \{\phi;x\} \qquad (B8)$$

and its derivatives of order $(N-2)$ or less. From (B6)

$$\deg(L_n) = n \qquad (B9)$$

and the operator (B1) is L_2 and (B5) is L_3.

The proof of the above uses the following result of Lavie[31]: A differential expression that is invariant under the Möbius group

$$\phi = (a\psi + b)/(c\psi + d) \qquad (B10)$$

is a functional of the Schwarzian derivative. In particular, a polynomial differential invariant of order n is a polynomial in the Schwarzian derivative and suitable derivatives of the Schwarzian derivatives of order $n - 3$ or less.

The Möbius transformation (B10) can be expressed as the composition of three operations (i) scaling

$$\phi = a\psi, \qquad (B11)$$

(ii) translation

$$\phi = \psi + b, \qquad (B12)$$

(iii) inversion

$$\phi = -1/\psi. \qquad (B13)$$

From (B7) the coefficients of (B6) are invariant under (B11) and (B12). Therefore, it remains to show that (B6) is invariant under (B13).

Now

$$D + mV = D + m(\phi_{xx}/\phi_x) = \phi_x^{-m}D\phi_x^m, \qquad (B14)$$

and from (B14)

$$L_{n+1} = \phi_x^{n/2+1}(\phi_x^{-1}D)^{n+1}\phi_x^{n/2}. \qquad (B15)$$

Under the change of variable

$$\phi \to x, \quad x \to \phi, \qquad (B16)$$

$$\phi_x^{-1} = x_\phi, \quad \phi_x^{-1}D = \frac{\partial}{\partial\phi} = D_\phi. \qquad (B17)$$

The operator

$$L_{n+1} = x_\phi^{-n/2-1}D_\phi^{n+1}x_\phi^{-n/2}. \qquad (B18)$$

With the change of variable (B13)

$$D_\phi = \psi^2 D_\psi, \qquad (B19)$$

and

$$L_{n+1} = x_\psi^{-n/2-1}\psi^{-n-2}(\psi^2 D_\psi)^{n+1}\psi^{-n}x_\psi^{-n/2}. \qquad (B20)$$

Now

$$\psi^{-n-2}(\psi^2 D_\psi)^{n+1}\psi^{-n} = \psi^{-n}\{D_\psi\psi^2 D_\psi\psi^2_\psi \cdots \psi^2 D_\psi\}\psi^{-n}. \qquad (B21)$$

We will show that

$$(B21) = D_\psi^{n+1}, \qquad (B22)$$

demonstrating that (B18) [and (B15)] is invariant under (B13) and, by the previous remark, under (B10). Equation (B22) is trivial when $n = 0$. When $n = 1$, the identity

$$D_\psi\psi^2 D_\psi = \psi D_\psi^2\psi \qquad (B23)$$

demonstrates (B22).

Now, by induction on n, (B22) is equivalent to

$$\psi^n D_\psi^{n+1}\psi^n = D_\psi\psi^n D_\psi^{n-1}\psi^n D_\psi, \qquad (B24)$$

and directly

$$D_\psi\psi^n D_\psi^{n-1}\psi^n D_\psi = \psi^n D_\psi^{n+1}\psi^n + nR_n, \qquad (B25)$$

where

$$R_n = \psi^{n-1}D_\psi^{n-1}\psi^n D_\psi - \psi^n D_\psi^n\psi^{n-1}. \qquad (B26)$$

Now

$$R_n = \psi^{n-1}\{D_\psi^n\psi - \psi D_\psi^n - nD_\psi^{n-1}\}\psi^{n-1}, \qquad (B27)$$

and the identity

$$D_\psi^n\psi = \psi D_\psi^n + nD_\psi^{n-1} \qquad (B28)$$

establishes that

$$R_n = 0$$

and (B24) and (B22).

Therefore, by the result of Lavie, the operators have coefficients that are polynomial in the Schwarzian derivative (and its derivatives) when

$V = \phi_{xx}/\phi_x$.

For reference, the first few operators are

$$
\begin{aligned}
L_1 &= D, \\
L_2 &= D^2 + \tfrac{1}{2} U, \\
L_3 &= D^3 + 2UD + U_x, \\
L_4 &= D^4 + 5UD^2 + 5U_xD + \tfrac{3}{2}(U_{xx} + \tfrac{3}{2}U^2), \\
L_5 &= D^5 + 10UD^3 + 15U_xD^2 + 9(U_{xx} + 16U^2)D \\
&\quad + 2\frac{\partial}{\partial x}(U_{xx} + 4U^2), \qquad \text{(B29)} \\
L_6 &= D^6 + \tfrac{35}{2} UD^4 + 35U_xD^3 + (\tfrac{63}{2} U_{xx} + \tfrac{259}{4} U^2)D^2 \\
&\quad + \frac{\partial}{\partial x}(14U_{xx} + \tfrac{259}{4} U^2)D + U_{xxxx} \\
&\quad + \tfrac{31}{2} UU_{xx} + 13U_x^2 + \tfrac{45}{4} U^3, \\
L_7 &= D^7 + 28UD^5 - 20U_xD^4 + (84U_{xx} + 196U^2)D^3 \\
&\quad + (56U_{xxx} + 588UU_x)D^2 \\
&\quad + (20U_{xxxx} + 352UU_{xx} + 295U_x^2 + 288U^3)D \\
&\quad + \frac{\partial}{\partial x}\left(U_{xxxx} + 26UU_{xx} + \frac{33}{2}U_x^2 + 48U^3\right),
\end{aligned}
$$

where

$$U = V_x - \tfrac{1}{2} V^2 = \{\phi;x\}. \qquad \text{(B30)}$$

Note that under the scaling

$$x \to a^{-1}x, \qquad \text{(B31)}$$

from (B6) and (B7)

$$L_{n+1} \to a^{n+1}L_{n+1}, \qquad \text{(B32)}$$

and each term of the expanded operator has the same weight, where in accord with (B30),

$$U \to a^2U. \qquad \text{(B33)}$$

We also note the simplicity of the operator forms (B18).

Remark (1): The operators (L_1, L_3) define the dual Hamiltonian structure of the KdV equation.

Remark (2): The operator (L_2, L_3) do not define the Lax pair for the KdV equation.

Remark (3): From (B6), the coefficient of D^{n-1} in the expansion for L_n vanishes.

Remark (4): The L_{2k} operators are symmetric, the L_{2k+1} are antisymmetric.

Remark (5): For the CDG sequence the defining operators [Eqs. (1.22) and (1.23)] are

$$
\begin{aligned}
&\theta_1 = \theta_2 = L_3, \\
&J_1 = D^{-1}L_2DL_2D^{-1}, \qquad \text{(B34)} \\
&J_2 = D^{-1}L_5D^{-1}.
\end{aligned}
$$

Now consider the N-component, two-dimensional, periodic Toda lattice[14]

$$\theta_{jxt} = e^{\theta_j - \theta_{j-1}} - e^{\theta_{j+1} - \theta_j}, \qquad \text{(B35)}$$

where $j = 1,2,3,\ldots \bmod (N)$. The Nth order, scalar operator

$$L = \prod_{j=0}^{n} (D - \theta_{jx}), \qquad \text{(B36)}$$

where $\Sigma_1^n \theta_{jx} = 0$, is the scattering part of the Lax pair for (B35). It is not difficult to see that (1) when (B36) is L_2, Eq. (B35) reduces to the sine–Gordon system; (2) when (B36) is L_3, Eq. (B35) reduces to the Bullough–Dodd Eq. (2.7); (3) the identification of (B36) with L_n for $n > 2$ does not lead to a consistent scalar reduction of (B35). Therefore, subject to the identification of (B36) with (B6), the sine–Gordon and Bullough–Dodd equations represent the only consistent scalar reductions of Eqs. (B35). If the sequence (B6) is unique (in the sense that L_n is the unique factorization of an nth order scalar operator consistent with the Schwarzian formulation and not a product of operators of lower degrees) then the results of Sec. I show that, within a wide class of equations, the KdV and CDG sequences are the unique instances of scalar evolution equations (See Ref. 4 and especially Ref. 32).

Remark (6): It is easy to see, from (B6) and (B7), that a basis for the null space of the operator L_{n+1} is

$$\{\phi_x^{-n/2}\phi^j;\ \ j = 0,1,2,\ldots,n\}. \qquad \text{(B37)}$$

Therefore, the general null function, which satisfies

$$L_{n+1} f_n = 0, \qquad \text{(B38)}$$

is

$$f_n = \phi_x^{-n/2} \sum_{j=0}^{n} c_j \phi^j, \qquad \text{(B39)}$$

where the $\{c_j\}$ are constants.

In a previous work[4] we have investigated the class of partial differential equations

$$\phi_t/\phi_x + B(\{\phi;x\}) = 0 \qquad \text{(B40)}$$

that are formulated in terms of the Schwarzian derivative and have a transformation

$$\phi_x = \psi_x^m, \qquad \text{(B41)}$$

which preserves the formulation (B40) in terms of the Schwarzian derivative. The KdV sequence corresponds to $m = -1$ and the CDG sequence corresponds to $m = -2$. In terms of the "modified" variables

$$V = \phi_{xx}/\phi_x, \quad W = \psi_{xx}/\psi_x, \qquad \text{(B42)}$$

the symmetry (B41) corresponds to

$$V = mW. \qquad \text{(B43)}$$

In analogy with the recursion operators for the modified KdV and CDG sequence define the "recursion operators"

$$\Omega_{n+2} = D(D + V)D^{-1} \circ L_{n+1} \circ D^{-1}(D - V), \qquad \text{(B44)}$$

where

$$V = \phi_{xx}/\phi_x. \qquad \text{(B45)}$$

The "sequence" of modified equations are

$$V_t + \Omega_{n+2}^j \circ V_x = 0, \qquad \text{(B46)}$$

which corresponds to equations of the form (B40), where

$$\phi_t/\phi_x + M_{n+2}^j \circ 1 = 0, \qquad \text{(B47)}$$

where

$$
\begin{aligned}
&M_{n+2} = D^{-1} \circ L_{n+1} \circ D^{-1} \circ \theta, \\
&\theta = (D - V)D(D + V) = L_3.
\end{aligned} \qquad \text{(B48)}
$$

We note that Ω_4 is the modified KdV recursion operator and Ω_6 is the modified CDG operator. (Actually, $\Omega_4 = \Omega_2^2$, where Ω_2 is the MKdV recursion operator.) From the identity

$$(D + aV)D^{-1}(D + bV) = (D + bV)D^{-1}(D + aV), \tag{B49}$$

the definition of L_{n+1} (B6), and the form of the Eqs. (B46), a symmetry of the form (B43) that preserves the Schwarzian formulation of (B46) and (B47) requires that

$$m = -n/2. \tag{B50}$$

Again, if the operators L_n or the composite operators

$$\prod_j L_{kj}, \tag{B51}$$

where $\Sigma_j k_j = n$ are the unique operators of order n with a product form and Schwarzian coefficients, then the symmetry (B43) is allowed only if $n = 2$ or $n = 4$. That is,

$$m = -1, -2, \tag{B52}$$

which are the KdV and CDG sequences.

It may be of some interest to note that the "Harry Dym" sequence corresponding to (B47) is

$$x_t = M^j_{n+2} \circ 1, \tag{B53}$$

where

$$M_{n+2} = D_\phi^{-1} x_\phi^{-n/2} D_\phi^{n+1} x_\phi^{-n/2} D_\phi^{-1} x_\phi^{-1} D_\phi^3 x_\phi^{-1}.$$

Finally, it is not difficult to show that if Eq. (B40) has a transformation (B41), which preserves the Schwarzian formulation, then the resulting equations will have the Painlevé property only for the index pairs $(m, 1/m)$:

$$\text{(i)}\ (-1, -1), \quad \text{(ii)}\ (-2, -\tfrac{1}{2}). \tag{B54}$$

That is, only for the KdV and CDG sequences. To see this, we start with the "generic" singularity:

$$\varphi = \varphi_0/\varepsilon + \varphi_1 + \cdots, \tag{B55}$$

which is single-valued for Eqs. (B40) (see Ref. 4). Then, by applying (B41) to (B55) we obtain the result.

[1]J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[2]J. Weiss, J. Math. Phys. **24**, 1405 (1983).
[3]D. V. Chudnovsky, G. V. Chudnovsky, and M. Tabor, Phys. Lett. A **97**, 268 (1983).
[4]J. Weiss, J. Math. Phys. **25**, 13 (1984).
[5]J. Weiss, J. Math. Phys. **25**, 2226 (1984).
[6]J. D. Gibbon, P. Radmore, M. Tabor, and D. Wood, Stud. Appl. Math. **72**, 39 (1985).
[7]J. Weiss, J. Math. Phys. **26**, 258 (1985).
[8]J. Weiss, J. Math. Phys. **26**, 2174 (1985).
[9]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962) Vol. 2.
[10]H. Yoshida, Celest. Mech. **31**, 363, 381 (1983).
[11]P. J. Caudrey, R. K. Dodd, and J. D. Gibbon, Proc. R. Soc. London Ser. A **351**, 407 (1976).
[12]K. Sawada and T. Kotera, Prog. Theor, Phys. **51**, 1355 (1974).
[13]R. K. Dodd and R. K. Bullough, Proc. R. Soc. London Ser. A **351**, 499 (1976).
[14]A. P. Fordy and J. Gibbons, Commun. Math. Phys. **77**, 21 (1980).
[15]A. V. Mikhailov, Physica D **3**, 73 (1981).
[16]W. Rund, in *Lecture Notes in Mathematics*, Vol. 515 (Springer, New York, 1976).
[17]B. A. Kuperschmidt and G. Wilson, Commun. Math. Phys. **81**, 189 (1981).
[18]J. D. Gibbon and M. Tabor, J. Math. Phys. **26**, 1956 (1985).
[19]B. Fuchssteiner, Prog. Theor. Phys. **68**, 1082 (1982).
[20]L. R. Ford, *Automorphic Functions*(Chelsea, New York, 1951).
[21]J. Lehner, "Discontinuous Groups and Automorphic Functions," in *Mathematical Surveys* (Am. Math. Soc., Providence, RI, 1964).
[22]L. V. Alfors, Proc. Int. Congr. Math. Helsinki **1978**, 71.
[23]L. Bers, Bull. London Math. Soc. **4**, 257 (1972).
[24]L. Bers, Bull. Am. Math. Soc. **83**, 1083 (1977).
[25]R. A. Rankin, Proc. Edinburgh Math. Soc. **11**, 25 (1958).
[26]R. A. Rankin, J. Analyse Math. **6**, 149 (1958).
[27]D. A. Hejhal, Acta. Math. **135**, 1 (1975).
[28]J. L. Burchnall and T. W. Chaundy, Proc. R. Soc. London Ser. A **118**, 557 (1928).
[29]M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transformation* (SIAM, Philadelphia, 1981), p. 266.
[30]A. F. Rañada, A. Ramani, B. Dorizzi, and B. Grammaticos, J. Math. Phys. **26**, 708 (1985).
[31]M. Lavie, Can. J. Math. **21**, 235 (1969).
[32]J. Weiss, Phys. Lett. A **105**, 387 (1984).

General Lorentz transformations and applications

C. B. van Wyk
University of the Orange Free State, Bloemfontein, South Africa

(Received 10 July 1985; accepted for publication 30 December 1985)

It is known that the most general proper orthochronous vector Lorentz (transformation) operator can be generated by a skew-symmetric 4×4 matrix containing an antisymmetric tensor of the second rank. The corresponding Lorentz operator for the two-component spinor is presented and, as can be expected, it contains the same tensor as the vector operator. Since the Pauli matrices of the spinor operator have very simple multiplication properties, the behavior of the tensor under multiplication of spinor operators is easily obtained. By comparison the corresponding properties of the tensor in vector operators can be obtained without multiplying 4×4 matrices. The physical meaning of the tensor contained in a Lorentz operator is discussed. Apart from the usual or regular operator a singular operator is discussed. Still other types of Lorentz operators are possible.

I. INTRODUCTION

The general finite proper orthochronous Lorentz transformation operator for vectors has been described in terms of an antisymmetric second-rank tensor (two three-vectors).[1] In this paper we show how this tensor as well as other parameters can be found for a given Lorentz operator of this kind.

A practical obstacle in the handling of these vector operators is the amount of labor required to multiply 4×4 matrices.[2] This obstacle is surmounted by formulating the corresponding general operator for the two-component spinor that contains the same tensor. Since these 2×2 operators can be multiplied very simply, the properties of the tensor under the multiplication of spinor operators are easily obtained. Its properties under a similar multiplication of vector operators are then obtained by comparison without multiplying 4×4 matrices.

The tensor belonging to a product of elementary operators is constructed in several cases and attempts at interpreting it physically are made. A procedure related to this one has been applied to the relativistic composition of velocities and to the Thomas precession.[3]

With suitable adjustments the same procedure is used to generate a singular Lorentz operator. The possibility of still other types of operators is pointed out but not discussed.

One would not expect singular and ordinary Lorentz operators to occur in close association. However, we discuss cases where they do.

Finally we display the corresponding general Lorentz operator for the four-component spinor.

II. THE VECTOR FORMULATION

As in a previous work[1] we introduce the antisymmetric 4×4 matrix U and its dual U^D as follows:

$$U=\begin{pmatrix} 0 & -ia_3 & ia_2 & b_1\\ ia_3 & 0 & -ia_1 & b_2\\ -ia_2 & ia_1 & 0 & b_3\\ -b_1 & -b_2 & -b_3 & 0\end{pmatrix}, \tag{1}$$

$$U^D_{\mu\nu}=\tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}U_{\rho\sigma},$$

where $\mathbf{a}$ is a real pseudovector and $\mathbf{b}$ is a real vector. These matrices have the characteristic equation

$$\lambda^4-(a\cdot a-b\cdot b)\lambda^2-(\mathbf{a}\cdot\mathbf{b})^2=0, \tag{2}$$

and satisfy

$$\begin{aligned} U^{D^3}&=(a^2-b^2)U^D-i(\mathbf{a}\cdot\mathbf{b})U,\\ U^3&=(a^2-b^2)U-i(\mathbf{a}\cdot\mathbf{b})U^D,\\ UU^D&=U^DU=i\mathbf{a}\cdot\mathbf{b}I.\end{aligned} \tag{3}$$

Choose

$$\mathbf{a}\cdot\mathbf{a}-\mathbf{b}\cdot\mathbf{b}=1,\quad \mathbf{a}\cdot\mathbf{b}=0,\quad\text{or}\quad(\mathbf{a}+i\mathbf{b})^2=1, \tag{4}$$

then

$$U^3=U,\quad (U^D)^3=U^D,\quad UU^D=U^DU=0, \tag{5}$$

indicating that U and U^D have the eigenvalues $1,-1,0,0$ and common eigenvectors, which will be discussed in Sec. III.

Now

$$\begin{aligned}\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)&=\exp(iU\theta+U^D\phi)\\ &=\exp(iU\theta)\exp(U^D\phi)\\ &=\Lambda_1(\mathbf{a},\mathbf{b},\theta,0)\Lambda_1(\mathbf{a},\mathbf{b},0,\phi),\end{aligned} \tag{6}$$

with θ and ϕ real parameters, satisfies

$$\Lambda_1\Lambda_1^T=\Lambda_1^T\Lambda_1=I,\quad \det\Lambda_1=1, \tag{7}$$

with $(\Lambda_1)_{jk}$ and $(\Lambda_1)_{44}>0$ real while $(\Lambda_1)_{j4}$ and $(\Lambda_1)_{4j}$ are imaginary ($j,k=1,2,3$). Then the transformation

$$\begin{aligned}x'&=\Lambda_1 x\\ &=[I+iU\sin\theta+U^2(\cos\theta-1)+U^D\sinh\phi\\ &\quad+U^{D^2}(\cosh\phi-1)]x,\end{aligned} \tag{8}$$

with $x_4=ict$ yields

$$x'_jx'_j=x_kx_k\quad(j,k=1,2,3,4),$$

and can be regarded as a Lorentz transformation. Since Λ_1 contains six independent parameters (four of the a_j and b_j and θ and ϕ) it can be regarded as the most general, proper, finite, orthochronous vector Lorentz transformation operator.

Inspection shows that

$$\begin{aligned}\Lambda_1(\hat{\mathbf{a}},0,\theta,0)&=R_1(\hat{\mathbf{a}},\theta),\\ \Lambda_1(\hat{\mathbf{a}},0,0,\phi)&=L_1(\hat{\mathbf{a}},\phi),\end{aligned} \tag{9}$$

 0022-2488/86/051306-05$02.50

where $R_1(\hat{\mathbf{a}},\theta)$ is the operator for a pure rotation through an angle θ about the axis defined by the unit vector $\hat{\mathbf{a}}$, and $L_1(\hat{\mathbf{a}},\phi)$ is the operator for a pure Lorentz transformation relating to a relative velocity with direction $\hat{\mathbf{a}}$ and magnitude $c \tanh \phi$.

We notice that Λ_1 is the product of two commuting factors $\Lambda_1(\mathbf{a},\mathbf{b},\theta,0)$ and $\Lambda_1(\mathbf{a},\mathbf{b},0,\phi)$, and we now conclude that these factors are generalizations of a pure rotation and a pure Lorentz transformation operator, respectively.

The question arises how to find the $\mathbf{a},\mathbf{b},\theta$, and ϕ for any given matrix Λ_1, which is known to be of this type. In terms of

$$\operatorname{Tr}\Lambda_1 = 2(\cos\theta + \cosh\phi),$$
$$M = \tfrac{1}{2}(\Lambda_1 - \Lambda_1^T), \tag{10}$$

we have

$$\begin{aligned}
\cosh\phi &= \tfrac{1}{4}\operatorname{Tr}\Lambda_1 + [1 + \tfrac{1}{4}\operatorname{Tr}M^2 - \tfrac{1}{16}(\operatorname{Tr}\Lambda_1)^2]^{1/2},\\
\cos\theta &= \tfrac{1}{4}\operatorname{Tr}\Lambda_1 - [1 + \tfrac{1}{4}\operatorname{Tr}M^2 - \tfrac{1}{16}(\operatorname{Tr}\Lambda_1)^2]^{1/2},\\
a_j(\sin^2\theta + \sinh^2\phi) &= \tfrac{1}{2}\epsilon_{jkm}M_{km}\sin\theta + iM_{j4}\sinh\phi,\\
b_j(\sin^2\theta + \sinh^2\phi) &= \tfrac{1}{2}\epsilon_{jkm}M_{km}\sinh\phi - iM_{j4}\sin\theta.
\end{aligned} \tag{11}$$

In the special case where

$$\operatorname{Tr}M^2 = \tfrac{1}{2}(\operatorname{Tr}\Lambda_1)^2 - 2\operatorname{Tr}\Lambda_1, \tag{12}$$

it can be shown that

$$\begin{aligned}
\Lambda_1 &= \Lambda_1(\mathbf{a},\mathbf{b},\theta,0) \quad \text{if} \quad \operatorname{Tr}\Lambda_1 < 4,\\
\Lambda_1 &= \Lambda_1(\mathbf{a},\mathbf{b},0,\phi) \quad \text{if} \quad \operatorname{Tr}\Lambda_1 > 4.
\end{aligned} \tag{13}$$

If $\phi = 0$ and $\theta = \pi$, then Λ_1 is a symmetric matrix and the above procedure fails. It can be shown that in this case the top left-hand 3×3 submatrix takes the form

$$H = (1 - 2a^2)I + 2(\mathbf{a}\mathbf{a}^T + \mathbf{b}\mathbf{b}^T).$$

Now $\mathbf{a}$ and $\mathbf{b}$ are eigenvectors of H belonging to the eigenvalues 1 and -1, respectively. It is convenient to refer to $\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)$ as the standard form for the vector transformation operator.

III. THE MEANING OF THE TENSOR

In certain special cases like (9), the physical meaning of the parameters occurring in the Lorentz operator $\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)$ is quite clear. Another case is

$$\Lambda_1(\hat{\mathbf{a}},0,\theta,\phi) = R_1(\hat{\mathbf{a}},\theta)L_1(\hat{\mathbf{a}},\phi).$$

In an attempt to arrive at a general physical meaning for $\mathbf{a}$ and $\mathbf{b}$ we define the four-vectors

$$\begin{aligned}
A &= \begin{pmatrix}\hat{\mathbf{a}}\\ 0\end{pmatrix}, \quad D = \begin{pmatrix}\hat{\mathbf{a}}\wedge\mathbf{b}\\ ia\end{pmatrix},\\
B &= \begin{pmatrix}\hat{\mathbf{b}}\\ 0\end{pmatrix}, \quad E = \begin{pmatrix}\mathbf{a}\wedge\hat{\mathbf{b}}\\ ib\end{pmatrix},
\end{aligned} \tag{14}$$

using an obvious notation. They satisfy

$$A\cdot A = B\cdot B = E\cdot E = -D\cdot D = 1,$$
$$A\cdot B = A\cdot E = A\cdot D = B\cdot E = B\cdot D = E\cdot D = 0.$$

Also

$$UA = 0 = UD, \quad UB = iE, \quad UE = -iB,$$
$$U^DA = D, \quad U^DD = A, \quad U^DB = 0 = U^DE.$$

It follows that

$$\begin{aligned}
\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)(D\pm A) &= \exp(\pm\phi)(D\pm A),\\
\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)(B\pm iE) &= \exp(\pm i\theta)(B\pm iE),
\end{aligned} \tag{15}$$

whence A and D are left invariant by $\Lambda_1(\mathbf{a},\mathbf{b},\theta,0)$ and B and E by $\Lambda_1(\mathbf{a},\mathbf{b},0,\phi)$. Hence every Lorentz transformation for which $\theta = 0$ or $\phi = 0$ transforms only that part of space-time that is left invariant by the other. The two projection operators U^2 and U^{D^2} satisfy

$$U^2 + U^{D^2} = I,$$

and decompose space-time into two subspaces; $\mathbf{a}$ and $\mathbf{b}$ specify that decomposition.

The general transformation with $\theta \neq 0 \neq \phi$ leaves none of these vectors invariant. Nevertheless, the decomposition still occurs and is specified by $\mathbf{a}$ and $\mathbf{b}$.

A second way of attaching physical meaning to $\mathbf{a}$ and $\mathbf{b}$ will emerge in Sec. V, where they will be expressed in terms of axes of rotation, directions of relative velocities, etc.

Finally, in certain cases, $\mathbf{a}$ and $\mathbf{b}$ can be expressed as simple functions of the initial and final vectors of a Lorentz transformation.[4]

IV. THE OPERATOR FOR THE TWO-COMPONENT SPINOR

By means of the formalism described above we can study, for instance, the invariance properties of any product of Lorentz operators. Such a study would be hampered by the labor of multiplying 4×4 matrices. To minimize this labor, the discussion of such products was often limited to the infinitesimal cases.[2]

We now assume that for a given physical Lorentz transformation the general operator in the vector and spinor representations contain the same antisymmetric tensor and "angular" variables. Hence the properties of the tensor obtained from the spinor operator with its very simple multiplication rules can be used in the vector case and no multiplication of 4×4 matrices is required. To utilize this approach we require the general Lorentz operator for the two-component spinor.

In the theory of the two-component spinor, the coordinate-free pure rotation and pure Lorentz operators are, respectively, given by

$$\begin{aligned}
R_2(\hat{\mathbf{n}},\theta) &= \exp(\tfrac{1}{2}i\theta\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}) = c + i\boldsymbol{\sigma}\cdot\hat{\mathbf{n}}s,\\
L_2(\hat{\mathbf{u}},\phi) &= \exp(\tfrac{1}{2}\phi\boldsymbol{\sigma}\cdot\hat{\mathbf{u}}) = C + \boldsymbol{\sigma}\cdot\hat{\mathbf{u}}S,
\end{aligned} \tag{16}$$

where

$$c = \cos(\theta/2), \quad s = \sin(\theta/2),$$
$$C = \cosh(\phi/2), \quad S = \sinh(\phi/2),$$

and σ_j $(j = 1,2,3)$ are the 2×2 Pauli matrices while $\hat{\mathbf{n}}$ and $\hat{\mathbf{u}}$ are unit vectors. The property of the 2×2 matrices R_2 and L_2 that qualifies them as pure rotation and pure Lorentz operators is that

$$\det R_2 = \det L_2 = 1.$$

By borrowing ideas from the standard form for the vector transformation described above, we generalize (16) to the following general Lorentz operator for the two-component spinor:

$\Lambda_2(\mathbf{a},\mathbf{b},\theta,\phi)$

$$= \exp(i\tfrac{1}{2}\boldsymbol{\sigma}\cdot(\mathbf{a}+i\mathbf{b})\theta)\exp(\tfrac{1}{2}\boldsymbol{\sigma}\cdot(\mathbf{a}+i\mathbf{b})\phi)$$
$$= (c+is\boldsymbol{\sigma}\cdot(\mathbf{a}+i\mathbf{b}))(C+S\boldsymbol{\sigma}\cdot(\mathbf{a}+i\mathbf{b})). \quad (17)$$

The determinant of this matrix operator equals unity provided that

$$(\mathbf{a}+i\mathbf{b})\cdot(\mathbf{a}+i\mathbf{b}) = 1,$$

which is the same as (4) for the vector operator. Hence Λ_2 is a Lorentz operator that contains six independent parameters and can be regarded as the standard form for the spinor operator. It can be expressed as the product of two commuting factors:

$$\Lambda_2(\mathbf{a},\mathbf{b},\theta,\phi) = \Lambda_2(\mathbf{a},\mathbf{b},\theta,0)\Lambda_2(\mathbf{a},\mathbf{b},0,\phi).$$

V. APPLICATIONS

We illustrate the procedure described above by a few examples.

(i) The product of two pure Lorentz operators has been considered with certain specific practical objectives in mind.[3] We now consider more general objectives.

For any given $\hat{\mathbf{u}},\phi_1,\hat{\mathbf{v}},\phi_2$ we express the product of two pure Lorentz opertors in the standard form according to

$$L_2(\hat{\mathbf{u}},\phi_1)L_2(\hat{\mathbf{v}},\phi_2) = \Lambda_2(\mathbf{a},\mathbf{b},\theta,\phi);$$

then we have to determine $\mathbf{a},\mathbf{b},\theta$, and ϕ. Now (16), (17), and the elementary result

$$(\boldsymbol{\sigma}\cdot\mathbf{u})(\boldsymbol{\sigma}\cdot\mathbf{v}) = \mathbf{u}\cdot\mathbf{v} + i\boldsymbol{\sigma}\cdot(\mathbf{u}\wedge\mathbf{v})$$

for Pauli matrices must be used. We separate real and imaginary parts and equate coefficients of corresponding σ's. Then we have

$$\begin{aligned} &sS = 0,\\ &cC = C_1C_2 + \hat{\mathbf{u}}\cdot\hat{\mathbf{v}}S_1S_2,\\ &cS\mathbf{a} - sC\mathbf{b} = C_1S_2\hat{\mathbf{v}} + C_2S_1\hat{\mathbf{u}},\\ &sC\mathbf{a} + cS\mathbf{b} = S_1S_2\hat{\mathbf{u}}\wedge\hat{\mathbf{v}}, \end{aligned} \quad (18)$$

where

$$C_j = \cosh(\phi_j/2), \quad S_j = \sinh(\phi_j/2).$$

Hence $s=0$ or $S=0$ while cC obviously takes values greater than unity and therefore $s=0$, $c=1$.

Now Eqs. (18) become

$$\begin{aligned} &C = C_1C_2 + \hat{\mathbf{u}}\cdot\hat{\mathbf{v}}S_1S_2,\\ &S\mathbf{a} = C_1S_2\hat{\mathbf{v}} + C_2S_1\hat{\mathbf{u}},\\ &S\mathbf{b} = S_1S_2\hat{\mathbf{u}}\wedge\hat{\mathbf{v}}, \end{aligned} \quad (19)$$

from which the unknowns ϕ, $\mathbf{a}$, and $\mathbf{b}$ can be determined in terms of $\hat{\mathbf{u}},\phi_1,\hat{\mathbf{v}}$, and ϕ_2. From the assumption of Sec. IV we have at once for the vector formulation

$$L_1(\hat{\mathbf{u}},\phi_1)L_1(\hat{\mathbf{v}},\phi_2) = \Lambda_1(\mathbf{a},\mathbf{b},0,\phi), \quad (20)$$

with (19) also holding for (20). From (19) $\mathbf{a}\wedge\mathbf{b}$ can be found and then by (15) the space left invariant by the operator (20) is given by all linear combinations of

$$kB = \begin{pmatrix}\hat{\mathbf{u}}\wedge\hat{\mathbf{v}}\\ 0\end{pmatrix},$$

and

$$k'E = \begin{pmatrix}\hat{\mathbf{u}}\{-C_1S_2+S_1C_2(\hat{\mathbf{u}}\cdot\hat{\mathbf{v}})\} + \hat{\mathbf{v}}\{C_1S_2(\hat{\mathbf{u}}\cdot\hat{\mathbf{v}}) - S_1C_2\}\\ iS_1S_2\{1-(\hat{\mathbf{u}}\cdot\hat{\mathbf{v}})^2\}\end{pmatrix}$$

where k and k' are known constants but unimportant for our purposes.

In the following examples the products of pure vector operators have been expressed in standard vector form by first carrying out the spinor multiplication as was done above. Where appropriate the space left invariant by a given operator may be found according to Sec. III.

(ii) For the product

$$R_1(\hat{\mathbf{n}},\theta_1)L_1(\hat{\mathbf{u}},\phi_1) = \Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi), \quad (21)$$

it follows that

$$\begin{aligned} &cC = c_1C_1,\\ &sS = s_1S_1(\hat{\mathbf{n}}\cdot\hat{\mathbf{u}}),\\ &cS\mathbf{a} - sC\mathbf{b} = S_1(c_1\hat{\mathbf{u}} - s_1\hat{\mathbf{n}}\wedge\hat{\mathbf{u}}),\\ &sC\mathbf{a} + cS\mathbf{b} = s_1C_1\hat{\mathbf{n}}, \end{aligned} \quad (22)$$

and in the special case where

$$\hat{\mathbf{n}}\cdot\hat{\mathbf{u}} = 0, \quad c_1C_1 < 1,$$

we have $\phi = 0$ and

$$\begin{aligned} &c = c_1C_1,\\ &s\mathbf{a} = s_1C_1\hat{\mathbf{n}},\\ &s\mathbf{b} = S_1(s_1\hat{\mathbf{n}}\wedge\hat{\mathbf{u}} - c_1\hat{\mathbf{u}}), \end{aligned} \quad (23)$$

whence

$$R_1(\hat{\mathbf{n}},\theta_1)L_1(\hat{\mathbf{u}},\phi_1) = \Lambda_1(\mathbf{a},\mathbf{b},\theta,0).$$

If in (22)

$$\hat{\mathbf{n}}\wedge\hat{\mathbf{u}} = 0, \quad c_1C_1 > 1,$$

then $\theta = 0$ and

$$R_1(\hat{\mathbf{n}},\theta_1)L_1(\hat{\mathbf{u}},\phi_1) = \Lambda_1(\mathbf{a},\mathbf{b},0,\phi),$$

where

$$\begin{aligned} &C = c_1C_1,\\ &S\mathbf{a} = S_1(c_1\hat{\mathbf{u}} - s_1\hat{\mathbf{n}}\wedge\hat{\mathbf{u}}),\\ &S\mathbf{b} = s_1C_1\hat{\mathbf{n}}. \end{aligned} \quad (24)$$

(iii) If in (21)

$$\theta_1 = \pi, \quad \hat{\mathbf{u}}\cdot\hat{\mathbf{n}} = 0,$$

then from (22)

$$c = 0, \quad S = 0,$$
$$\mathbf{a} = C_1\hat{\mathbf{n}}, \quad \mathbf{b} = S_1\hat{\mathbf{n}}\wedge\hat{\mathbf{u}}.$$

Hence

$$R_1(\hat{\mathbf{n}},\pi)L_1(\hat{\mathbf{u}},\phi_1) = \Lambda_1(\mathbf{a},\mathbf{b},\pi,0),$$

if $\hat{\mathbf{u}}\cdot\hat{\mathbf{n}} = 0$. This is perhaps one of the simplest ways of illustrating the part played by $\mathbf{b}$.

(iv) The general operator can be decomposed into a product of pure operators according to

$$\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi) = R_1(\hat{\mathbf{n}},\theta_1)L_1(\hat{\mathbf{u}},\phi_1), \quad (25)$$

where

$$C_1 = (a^2C^2 - b^2c^2)^{1/2},$$

$$c_1 = cC/C_1,$$
$$S_1C_1\hat{\mathbf{u}} = \mathbf{a}SC - \mathbf{b}sc - \mathbf{a} \wedge \mathbf{b}(C^2 - c^2), \tag{26}$$
$$\hat{\mathbf{n}} = (\mathbf{a}t + \mathbf{b}T)/(a^2t^2 + b^2T^2)^{1/2},$$

and

$$t = \tan(\theta/2), \quad T = \tanh(\phi/2).$$

(v) The operator for transforming from the rest to the laboratory frame of reference is given by the similarity transformation

$$L_1^{-1}(\hat{\mathbf{v}},\phi_1)R_1(\hat{\mathbf{n}},\theta)L(\hat{\mathbf{v}},\phi_1) = \Lambda_1(\mathbf{a},\mathbf{b},\theta,0), \tag{27}$$

where

$$\mathbf{a} = (C_1^2 + S_1^2)\hat{\mathbf{n}} - 2S_1^2(\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})\hat{\mathbf{v}},$$
$$\mathbf{b} = 2S_1C_1\hat{\mathbf{n}} \wedge \hat{\mathbf{v}}. \tag{28}$$

A second similarity transformation is

$$L_1^{-1}(\hat{\mathbf{v}},\phi_1)L_1(\hat{\mathbf{u}},\phi)L_1(\hat{\mathbf{v}},\phi_1) = \Lambda_1(\mathbf{a},\mathbf{b},0,\phi), \tag{29}$$

with

$$\mathbf{a} = (C_1^2 + S_1^2)\hat{\mathbf{u}} - 2S_1^2(\hat{\mathbf{v}}\cdot\hat{\mathbf{u}})\hat{\mathbf{v}},$$
$$\mathbf{b} = 2S_1C_1\hat{\mathbf{u}} \wedge \hat{\mathbf{v}}.$$

A third one is obtained by adjusting (29) and then multiplying by (27):

$$L_1^{-1}(\hat{\mathbf{v}},\phi_1)R_1(\hat{\mathbf{n}},\theta)L_1(\hat{\mathbf{n}},\phi)L_1(\hat{\mathbf{v}},\phi_1)$$
$$= \Lambda_1(\mathbf{a},\mathbf{b},\theta,0)\Lambda_1(\mathbf{a},\mathbf{b},0,\phi)$$
$$= \Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi), \tag{30}$$

where $\mathbf{a}$ and $\mathbf{b}$ are given by (28). According to (27), (29), and (30), $\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)$ and its two factors can be transformed to pure operators or products thereof by means of similarity transformations involving only a pure Lorentz operator. To solve (28) for $\hat{\mathbf{n}}$, $\hat{\mathbf{v}}$, and ϕ_1 when $\mathbf{a}$ and $\mathbf{b}$ are given may be a formidable nonlinear problem.

VI. A SINGULAR OPERATOR

The above discussion of the usual or regular Lorentz operator was based on assumption (4) in the characteristic equation (2). Now assume that in (2) we have

$$\mathbf{a}\cdot\mathbf{a} - \mathbf{b}\cdot\mathbf{b} = 0, \quad \mathbf{a}\cdot\mathbf{b} = 0, \quad \text{or} \quad (\mathbf{a} + i\mathbf{b})^2 = 0. \tag{31}$$

Then

$$\lambda^4 = 0,$$

and, from (3),

$$U^3 = 0 = U^{D^3}. \tag{32}$$

It follows that U has four zero eigenvalues but only two linearly independent eigenvectors

$$A_s = \begin{pmatrix}\hat{\mathbf{a}}\\0\end{pmatrix}, \quad D_s = \begin{pmatrix}\hat{\mathbf{a}} \wedge \hat{\mathbf{b}}\\ i\end{pmatrix}, \tag{33}$$

the last one being a null vector. The Lorentz transformation operator generated by U is

$$\Lambda_1^s(\mathbf{a},\mathbf{b}) = \exp(iU)$$
$$= I + iU - \tfrac{1}{2}U^2, \tag{34}$$

after using (32). Since the magnitude of $\mathbf{a}$ or $\mathbf{b}$ is not prescribed, it seems superfluous to consider $\exp(iU\theta)$ instead of (34). This operator contains four arbitrary parameters and has four eigenvalues equal to unity, but only the two linearly independent eigenvectors (33).

We could consider the operator

$$\exp(U^D) = I + U^D + \tfrac{1}{2}U^{D^2}$$

arising from U^D. However, since $\mathbf{a}$ and $\mathbf{b}$ have the same magnitude this operator is essentially the same as (34).

Because of the unusual property (32) of U it is convenient to refer to Λ_1^s as a singular Lorentz operator. It is a generalization of the operator discussed by Hamermesh.[5]

For given Λ_1^s we have

$$iU = \tfrac{1}{2}(\Lambda_1^s - (\Lambda_1^s)^T),$$

from which $\mathbf{a}$ and $\mathbf{b}$ follow.

The corresponding two-component spinor operators Λ_2^s is obtained by starting from (17) and using (31). We find

$$\Lambda_2^s(\mathbf{a},\mathbf{b}) = \exp(\tfrac{1}{2}i\sigma\cdot(\mathbf{a} + i\mathbf{b}))$$
$$= I + \tfrac{1}{2}i\sigma\cdot(\mathbf{a} + i\mathbf{b})$$
$$= I + \tfrac{1}{2}\sigma\cdot(-\mathbf{b} + i\mathbf{a}). \tag{35}$$

We have now studied the Lorentz operators resulting from the choices (4) and (31) in (2). Still other choices are possible but we shall not discuss them.

VII. CLOSE ASSOCIATION OF Λ AND Λ*

Despite the widely differing properties of the regular and singular operators we have the following result. If

$$\hat{\mathbf{u}}\cdot\hat{\mathbf{n}} = 0, \quad c_1C_1 = 1, \quad c_1 \neq 1 \neq C_1,$$

then

$$R_1(\hat{\mathbf{n}},\theta_1)L_1(\hat{\mathbf{u}},\phi_1) = \Lambda_1^s(\mathbf{a},\mathbf{b}), \tag{36}$$

where

$$\mathbf{a} = 2s_1C_1\hat{\mathbf{n}},$$
$$\mathbf{b} = 2S_1(s_1\hat{\mathbf{n}} \wedge \hat{\mathbf{u}} - c_1\hat{\mathbf{u}}). \tag{37}$$

This follows at once by substituting (16) into (36) and using (35):

$$R_2(\hat{\mathbf{n}},\theta_1)L_2(\hat{\mathbf{u}},\phi_1)$$
$$= I + i\sigma\cdot[s_1C_1\hat{\mathbf{n}} + iS_1(s_1\hat{\mathbf{n}} \wedge \hat{\mathbf{u}} - c_1\hat{\mathbf{u}})]$$
$$= I + \tfrac{1}{2}i\sigma\cdot(\mathbf{a} + i\mathbf{b})$$
$$= \Lambda_2^s(\mathbf{a},\mathbf{b}).$$

We can also prove (36) directly by starting from (23) or (24) and then letting $c_1C_1 \to 1$.

Another example of the close association of the regular and singular operators is obtained when we seek the most general operator that leaves a given null vector

$$N = \begin{pmatrix}\mathbf{a} \wedge \mathbf{b}\\ i\end{pmatrix}$$

invariant. From (33)

$$\Lambda_1^s(\mathbf{a},\mathbf{b})N = N,$$

and the properties of a rotation

$$R_1(\mathbf{a} \wedge \mathbf{b},\theta_1)N = N,$$

where

$$\mathbf{a}\cdot\mathbf{b} - \mathbf{b}\cdot\mathbf{b} = 0, \quad \mathbf{a}\cdot\mathbf{b} = 0.$$

Hence the spinor operator we are looking for is given by the product

$$R_2(\hat{\mathbf{a}} \wedge \hat{\mathbf{b}},\theta_1)\Lambda_2^s(\mathbf{a},\mathbf{b}) = \Lambda_2(\boldsymbol{\alpha}, \boldsymbol{\beta},\theta_1,0),$$

where

$$2s_1\boldsymbol{\alpha} = c_1\mathbf{a} - s_1\mathbf{b} + 2s_1\hat{\mathbf{a}} \wedge \hat{\mathbf{b}},$$
$$2s_1\, \boldsymbol{\beta} = s_1\mathbf{a} + c_1\mathbf{b},$$

and

$$\boldsymbol{\alpha}\cdot\boldsymbol{\alpha} - \boldsymbol{\beta}\cdot\boldsymbol{\beta} = 1, \quad \boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 0.$$

Here we have used (16), (17), and (35). Consequently

$$\Lambda_1(\boldsymbol{\alpha}, \boldsymbol{\beta},\theta_1,0) = R_1(\hat{\mathbf{a}} \wedge \hat{\mathbf{b}},\theta_1)\Lambda_1^s(\mathbf{a},\mathbf{b})$$

is the required operator. It is a generalization of the one studied by Han and Kim.[6]

Third, we consider the product of two operators of the type $\Lambda_2(\mathbf{a},\mathbf{b},\pi,0)$. From (17) we have

$$\begin{aligned}\Lambda_2\Lambda_2' &= -(\boldsymbol{\sigma}\cdot(\mathbf{p}+i\mathbf{q}))(\boldsymbol{\sigma}\cdot(\mathbf{P}+i\mathbf{Q}))\\ &= -(\mathbf{p}+i\mathbf{q})\cdot(\mathbf{P}+i\mathbf{Q})\\ &\quad - i\boldsymbol{\sigma}\cdot[(\mathbf{p}+i\mathbf{q}) \wedge (\mathbf{P}+i\mathbf{Q})],\end{aligned} \tag{38}$$

where

$$(\mathbf{p}+i\mathbf{q})^2 = 1 = (\mathbf{P}+i\mathbf{Q})^2.$$

Now

$$\begin{aligned}&[(\mathbf{p}+i\mathbf{q}) \wedge (\mathbf{P}+i\mathbf{Q})]^2\\ &\quad = (\mathbf{p}+i\mathbf{q})^2(\mathbf{P}+i\mathbf{Q})^2 - ((\mathbf{p}+i\mathbf{q})\cdot(\mathbf{P}+i\mathbf{Q}))^2\\ &\quad = 0,\end{aligned}$$

if we choose $\mathbf{p}$, $\mathbf{q}$, $\mathbf{P}$, and $\mathbf{Q}$ in such a way that

$$(\mathbf{p}+i\mathbf{q})\cdot(\mathbf{P}+i\mathbf{Q}) = -1. \tag{39}$$

In this case (38) becomes

$$\Lambda_2\Lambda_2' = I + \tfrac{1}{2}i\boldsymbol{\sigma}\cdot(\mathbf{a}+i\mathbf{b}), \tag{40}$$

where

$$(\mathbf{a}+i\mathbf{b})^2 = 0.$$

If we compare (40) with (35) we have

$$\Lambda_2(\mathbf{p},\mathbf{q},\pi,0)\Lambda_2(\mathbf{P},\mathbf{Q},\pi,0) = \Lambda_2^s(\mathbf{a},\mathbf{b}),$$

where

$$-\tfrac{1}{2}(\mathbf{a}+i\mathbf{b}) = (\mathbf{p}+i\mathbf{q}) \wedge (\mathbf{P}+i\mathbf{Q}),$$

and provided that (39) holds. As before, subject to (39), we have for the vector operators

$$\Lambda_1(\mathbf{p},\mathbf{q},\pi,0)\Lambda_1(\mathbf{P},\mathbf{Q},\pi,0) = \Lambda_1^s(\mathbf{a},\mathbf{b}).$$

Hence the product of two regular operators may be a singular operator and conversely.

VIII. GENERAL OPERATOR FOR THE FOUR-COMPONENT SPINOR

The operator for the general Lorentz transformation of a four-component spinor will be discussed very briefly.

The pure rotation and pure Lorentz transformation for the four-component spinor are given by

$$\begin{aligned}\psi' &= \exp[i\boldsymbol{\sigma}'\cdot\hat{\mathbf{n}}(\theta/2)]\psi,\\ \psi' &= \exp[\boldsymbol{\alpha}\cdot\hat{\mathbf{u}}(\phi/2)]\psi,\end{aligned} \tag{41}$$

respectively, where the 4×4 Dirac matrices like $\boldsymbol{\sigma}'$ and $\boldsymbol{\alpha}$ will be defined below. These transformations have the invariance property

$$\bar{\psi}'\psi' = \psi'^{*}\beta\psi' = \psi^{*}\beta\psi = \bar{\psi}\psi. \tag{42}$$

Note that each of these transformations is already coordinate free but contains only three independent parameters.

Following previous works we generalize these pure transformations so that six independent parameters are involved. This is done by introducing the 4×4 matrices

$$\begin{aligned}V &= \boldsymbol{\sigma}'\cdot\mathbf{a} + i\boldsymbol{\alpha}\cdot\mathbf{b},\\ V^D &= \boldsymbol{\alpha}\cdot\mathbf{a} + i\boldsymbol{\sigma}'\cdot\mathbf{b} = -\gamma_5 V,\end{aligned} \tag{43}$$

where $\mathbf{a}$ and $\mathbf{b}$ are real three-vectors and

$$\boldsymbol{\sigma}' = \begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & \boldsymbol{\sigma}\end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix}0 & \boldsymbol{\sigma}\\ \boldsymbol{\sigma} & 0\end{pmatrix}, \quad \beta = \begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix},$$

where $\boldsymbol{\sigma}$ are the 2×2 Pauli matrices. Also

$$\gamma_5 = i\alpha_1\alpha_2\alpha_3, \quad \boldsymbol{\alpha} = -\gamma_5\boldsymbol{\sigma}' = -\boldsymbol{\sigma}'\gamma_5,$$
$$\gamma_5^2 = I, \quad \gamma_5 V = V\gamma_5, \quad \gamma_5 V^D = V^D\gamma_5.$$

Hence

$$V^2 = I = V^{D^2}, \quad VV^D = V^D V = -\gamma_5, \tag{44}$$

provided that

$$a^2 - b^2 = 1, \quad \mathbf{a}\cdot\mathbf{b} = 0. \tag{45}$$

The Hermitian conjugate of V and V^D are given by

$$V^{*} = \beta V\beta, \quad V^{D*} = -\beta V^D\beta. \tag{46}$$

Either of (41) can now be generalized to

$$\psi' = \Lambda_3(\mathbf{a},\mathbf{b},\theta,\phi))\psi, \tag{47}$$

where

$$\begin{aligned}\Lambda_3 &= \exp(iV(\theta/2))\exp(V^D(\phi/2))\\ &= (c + isV)(C + SV^D),\end{aligned} \tag{48}$$

and contains six independent parameters. From (43) and (46) it follows that

$$\Lambda_3^{*}\beta\Lambda_3 = \beta,$$

which ensures that (47) satisfies (42). Hence (48) is the required Lorentz operator for the four-component spinor.

[1]C. B. van Wyk, Am. J. Phys. **47**, 747 (1979).
[2]J. T. Cushing, Am. J. Phys. **35**, 858 (1967).
[3]C. B. van Wyk, Am. J. Phys. **52**, 853 (1984).
[4]C. B. van Wyk, J. Math. Phys. **27**, 1311 (1986).
[5]M. Hamermesh, *Group Theory* (Addison–Wesley, Reading, MA, 1962), p. 494.
[6]D. Han and Y. S. Kim, Am. J. Phys. **49**, 348 (1981).

Lorentz transformations in terms of initial and final vectors

C. B. van Wyk
University of the Orange Free State, Bloemfontein, South Africa

(Received 10 July 1985; accepted for publication 30 December 1985)

Given arbitrary initial vector(s) and their final vector(s) in a Lorentz transformation, the problem is to determine the operator of the transformation. The solution presented here consists of expressing the tensor contained in the Lorentz operator in terms of the given vectors and then the operator itself follows by the exponentiation of a generating matrix. This is possible in certain special cases. In the general case a number of simultaneous nonlinear equations have to be solved. The analytical solution of these equations is elusive while attempts at numerical solution indicate that supplementary information is required. The corresponding procedure for a singular operator is also presented.

I. INTRODUCTION

Various attempts have been made to express a rotation operator[1] and a Lorentz transformation operator in terms of the initial and final vectors of the transformation.[2] See Fradkin for further references.

The attempt described in this paper is based on the formulation of the general Lorentz transformation operator presented previously.[3] This formulation expresses the general Lorentz operator Λ_1 $(\mathbf{a},\mathbf{b},\theta,\phi)$ as a product of two Lorentz operators Λ_1 $(\mathbf{a},\mathbf{b},\theta,0)$ and Λ_1 $(\mathbf{a},\mathbf{b},0,\phi)$ and each of these factors is derived by exponentiation from a generating matrix that contains an antisymmetric tensor of the second rank with components $\mathbf{a}$ and $\mathbf{b}$.

This theory is applied to the subject of this paper by expressing the tensor and the generating matrices in a direct way in terms of two arbitrary given initial vectors and the resulting final vectors and then the required operators Λ_1 $(\mathbf{a},\mathbf{b},\theta,0)$ and Λ_1 $(\mathbf{a},\mathbf{b},0,\phi)$ follow quite simply by exponentiation.

Fradkin uses eigenvectors of the Lorentz operator Λ_1 $(\mathbf{a},\mathbf{b},\theta,0)$ as initial and final vectors to construct his operator and therefore his approach is rather specialized.

It is shown that to obtain the general Lorentz operator from two given but arbitrary initial vectors and the resulting final vectors, a system of simultaneous nonlinear equations must be solved. These equations are presented below but an analytic solution is still outstanding. Subsequently numerical methods were applied and it appears that two initial and two final vectors do not provide enough information to specify the operator. Ways to supplement the information are suggested.

The construction of the operator for a singular Lorentz transformation appears to be straightforward.

II. THE OPERATOR Λ_1 $(\mathbf{a},\mathbf{b},0,\phi)$

Let x be an arbitrary vector that is linked with x' by the Lorentz transformation[3]

$$x' = \Lambda_1\ (\mathbf{a},\mathbf{b},0,\phi)x, \tag{1}$$

where

$$\Lambda_1\ (\mathbf{a},\mathbf{b},0,\phi) = I + \sinh(\phi)U^D + ((\cosh\phi) - 1)U^{D^2}, \tag{2}$$

and

$$U=\begin{pmatrix} 0 & -ia_3 & ia_2 & b_1 \\ ia_3 & 0 & -ia_1 & b_2 \\ -ia_2 & ia_1 & 0 & b_3 \\ -b_1 & -b_2 & -b_3 & 0 \end{pmatrix},$$
$$U^D=\begin{pmatrix} 0 & b_3 & -b_2 & -ia_1 \\ -b_3 & 0 & b_1 & -ia_2 \\ b_2 & -b_1 & 0 & -ia_3 \\ ia_1 & ia_2 & ia_3 & 0 \end{pmatrix}. \tag{3}$$

As before $x_4 = ict$, $\mathbf{a}\cdot\mathbf{a} - \mathbf{b}\cdot\mathbf{b} = 1$, $\mathbf{a}\cdot\mathbf{b} = 0$.

For arbitrary x and the x' resulting from (1) we have to find $\mathbf{a}$, $\mathbf{b}$, and ϕ and therefore Λ_1 $(\mathbf{a},\mathbf{b},0,\phi)$. Since $U\,U^D = 0$, we can eliminate ϕ by multiplying (1) by U, and then the problem of finding $\mathbf{a}$ and $\mathbf{b}$ reduces to

$$U(x' - x) = 0. \tag{4}$$

In an obvious notation, this means that

$$i\mathbf{a}\wedge(\mathbf{x}'-\mathbf{x}) + \mathbf{b}(x_4' - x_4) = 0, \tag{5}$$

$$\mathbf{b}\cdot(\mathbf{x}' - \mathbf{x}) = 0. \tag{6}$$

For arbitrary $\mathbf{a}$, (5) provides a $\mathbf{b}$ that satisfies (6). Hence the data is insufficient to determine $\mathbf{a}$ and $\mathbf{b}$ uniquely and we need another arbitrary vector y that is linked to y' as in (1). From (6) and the corresponding equation for $y' - y$ we have at once

$$k\mathbf{b} = \mathbf{p}\wedge\mathbf{q}, \tag{7}$$

where

$$p = x' - x, \quad q = y' - y, \tag{8}$$

and k is a real number. From (5) and the corresponding equation for y we find

$$k\mathbf{a} = i(\mathbf{p}\,q_4 - \mathbf{q}\,p_4), \tag{9}$$

while

$$k^2 = (pq)^2 - p^2q^2 \tag{10}$$

follows from

$$\mathbf{a}\cdot\mathbf{a} - \mathbf{b}\cdot\mathbf{b} = 1.$$

As is usual the notation x^T for the transpose of x will be used only under special circumstances. We can reformulate (4), (7), and (9) as follows. If

 0022-2488/86/051311-04$02.50

$$Up = 0 = Uq \quad (p \neq q), \tag{11}$$

then

$$kU^D = pq^T - qp^T. \tag{12}$$

Having expressed **a** and **b** in terms of x, x', y, and y', we proceed to do the same for ϕ. From (1), (2), and (8), it follows that

$$\cosh\phi - 1 = (px)/(xU^{D^2}x) = k^2/N, \tag{13}$$

where, from (12)

$$k^2 x U^{D^2} x = (px)N,$$

and

$$N = 2(x'q)(xq) - q^2(px)$$
$$= 2[(px)(qy) - (py)(xq)]. \tag{14}$$

From (13), it follows that

$$\sinh^2\phi = (k^2/N^2)(k^2 + 2N)$$
$$= (k^2/N^2)(xy' - x'y)^2,$$

whence

$$\sinh\phi = k((x' + x)q)/N$$
$$= k(xy' - x'y)/N. \tag{15}$$

Substitution of (12), (13), and (15) into (2) yields

$$\Lambda_1\ (\mathbf{a},\mathbf{b},0,\phi) = I + (1/N)[(xy' - x'y)(pq^T - qp^T) + (pq^T - qp^T)^2]. \tag{16}$$

Now (16) is the operator of (1) expressed in terms of x, x', y, and y' as required. This operator leaves the spacelike vectors

$$B = \begin{pmatrix}\hat{\mathbf{b}}\\0\end{pmatrix}, \quad E = \begin{pmatrix}\mathbf{a}\wedge\hat{\mathbf{b}}\\ ib\end{pmatrix}$$

invariant and transforms the space defined by the vectors

$$A = \begin{pmatrix}\hat{\mathbf{a}}\\0\end{pmatrix}, \quad D = \begin{pmatrix}\hat{\mathbf{a}}\wedge\mathbf{b}\\ ia\end{pmatrix},$$

where the latter is a timelike vector.[3] Fradkin derives a similar operator by using invariance properties. There is, however, an important difference in that our x and y may be arbitrary vectors and need not be eigenvectors of some operator as required by Fradkin. Note that $x' - x$ and $y' - y$ belong to the timelike space of A and D despite the arbitrariness of x and y.

The case excluded in (11) occurs if

$$y = x + z,$$

where z is any linear combination of B and E. Then

$$y' = x' + z \quad \text{and} \quad p = q.$$

If x is in the plane of A and D, then

$$Ux = 0 \quad \text{and} \quad Ux' = 0, \tag{17}$$

and

$$kU^D = x'x^T - xx'^T,$$

and y is not required. Then

$$k^2 = (x'x)^2 - x^4,$$
$$\cosh\phi = (x'x)/x^2, \quad \sinh\phi = k/x^2, \tag{18}$$
$$\Lambda_1 = I + (1/M)[(x'x + x^2)(x'x^T - xx'^T) + (x'x^T - xx'^T)^2],$$

where

$$M = x^2(x'x + x^2).$$

Since Λ_1 $(\mathbf{a},\mathbf{b},0,\phi)$ does not have -1 as an eigenvalue, M vanishes only when x is a null vector.

In the pure Lorentz operator, $b = 0$, and it follows from (7) that

$$\mathbf{p} = \mathbf{q}, \quad \mathbf{p} = 0, \quad \text{or} \quad \mathbf{q} = 0.$$

If we choose $\mathbf{q} = 0$, $q_4 = i$, then from (9)

$$k\mathbf{a} = -\mathbf{p}. \tag{19}$$

Despite this choice we can still regard q as a four-vector and the theory of (12)–(16) remains valid. In this case the first forms of (14) and (15) are convenient.

Alternatively, note that (19) follows directly from the pure Lorentz transformation in the form

$$x' = L(\hat{\mathbf{u}},\phi)x,$$

and

$$\mathbf{x}' = \mathbf{x} + [-ix_4\sinh\phi + (\cosh\phi - 1)\hat{\mathbf{u}}\cdot\mathbf{x}]\hat{\mathbf{u}}.$$

This procedure suffers from a lack of symmetry but it requires only one vector x.

III. THE OPERATOR Λ_1 (a,b,θ,0)

This operator is treated in almost the same way as Λ_1 $(\mathbf{a},\mathbf{b},0,\phi)$. In this case (1) is replaced by

$$x' = \Lambda_1\ (\mathbf{a},\mathbf{b},\theta,0)x, \tag{20}$$

where x is arbitrary and

$$\Lambda_1\ (\mathbf{a},\mathbf{b},\theta,0) = I + i(\sin\theta)U + (\cos\theta - 1)U^2,$$

and again we need a second arbitrary initial vector y and the corresponding final vector y' satisfying (20). For given x, x', y, and y' we must find Λ_1 $(\mathbf{a},\mathbf{b},\theta,0)$. We now have

$$U^Dp = 0 = U^Dq, \tag{21}$$

with p and q given by (8). As in (11) and (12), we now conclude, on the basis of (21), that

$$iKU = pq^T - qp^T, \tag{22}$$

or

$$K\mathbf{a} = \mathbf{p}\wedge\mathbf{q},$$
$$K\mathbf{b} = -i(\mathbf{p}\,q_4 - \mathbf{q}\,p_4), \tag{23}$$

where

$$K^2 = p^2q^2 - (pq)^2 \tag{24}$$

follows from

$$\mathbf{a}\cdot\mathbf{a} - \mathbf{b}\cdot\mathbf{b} = 1.$$

Having found U in terms of x and x' we now turn to θ. From (20) it follows that

$$\cos\theta - 1 = (px)/xU^2x$$
$$= -K^2/N, \tag{25}$$

where N is given by (14). As for $\sinh\phi$ we have

$$\sin^2\theta = (K^2/N^2)(2N - K^2)$$
$$= (K^2/N^2)(xy' - x'y)^2,$$

whence

$$\sin\theta = K(xy' - x'y)/N.$$

Finally we have the required operator

$$\Lambda_1(\mathbf{a},\mathbf{b},\theta,0) = I + (1/N)[(xy' - x'y)(pq^T - qp^T) + (pq^T - qp^T)^2]. \quad (26)$$

When $\theta = \pi$ we have from (20)

$$\Lambda_1 = (\mathbf{a},\mathbf{b},\pi,0) = I - 2U^2.$$

If we expand

$$x = \alpha A + \delta D + \beta B + \gamma E,$$

and similarly for x', y, and y', where A, B, E, and D are the vectors defined in Sec. II and α, β, γ, and δ are real numbers, then it can be shown that

$$x'y = xy', \quad py = xq,$$

so that $\sin\theta$ vanishes indeed. Hence from (26),

$$\Lambda_1(\mathbf{a},\mathbf{b},\pi,0) = I + (1/N)(pq^T - qp^T)^2. \quad (27)$$

Formally (26) is the same as (16) for $\Lambda_1(\mathbf{a},\mathbf{b},0,\phi)$. Their basic difference is contained in (10) and (24), which indicate that the initial and final vectors of $\Lambda_1(\mathbf{a},\mathbf{b},0,\phi)$ define a timelike space while those of $\Lambda_1(\mathbf{a},\mathbf{b},\theta,0)$ define a spacelike space.

Should it happen that

$$Ux = 0, \quad \text{then} \quad Ux' = 0,$$

a situation similar to (17) and (18) arises, and y is not required.

In the special case of a pure rotation we have

$$b = 0, \quad p_4 = q_4 = 0,$$

in accordance with (23). It follows that for a rotation we still need two initial three-vectors $\mathbf{x}$ and $\mathbf{y}$ and the corresponding final vectors $\mathbf{x}'$ and $\mathbf{y}'$.

IV. THE OPERATOR $\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)$

In the discussion of $\Lambda_1(\mathbf{a},\mathbf{b},\theta,0)$ and $\Lambda_1(\mathbf{a},\mathbf{b},0,\phi)$ above, we succeeded in eliminating θ and ϕ, respectively, by multiplying the Lorentz transformation by U^D and U, respectively. After the determination of $\mathbf{a}$ and $\mathbf{b}$ it was easy to find θ or ϕ.

We now deal with the general case

$$\begin{aligned} x' &= \Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi)x \\ &= \Lambda_1(\mathbf{a},\mathbf{b},\theta,0)\Lambda_1(\mathbf{a},\mathbf{b},0,\phi)x \\ &= x + i(\sin\theta)Ux + (\cos\theta - 1)U^2x \\ &\quad + (\sinh\phi)U^Dx + (\cosh\phi - 1)U^{D^2}x. \end{aligned} \quad (28)$$

We eliminate ϕ by multiplying (28) and the corresponding equation in y by U. Then

$$Up = 0 = Uq,$$

where

$$\begin{aligned} p &= x' - x\cos\theta - i(\sin\theta)Ux, \\ q &= y' - y\cos\theta - i(\sin\theta)Uy. \end{aligned} \quad (29)$$

This situation is similar to the one in (11) and (12), and therefore

$$kU^D = pq^T - qp^T,$$

or

$$\begin{aligned} k\mathbf{b} &= \mathbf{p} \wedge \mathbf{q}, \\ k\mathbf{a} &= i(\mathbf{p}\,q_4 - \mathbf{q}\,p_4). \end{aligned} \quad (30)$$

Since

$$Ux = \begin{pmatrix} i\mathbf{a} \wedge \mathbf{x} + x_4\mathbf{b} \\ -\mathbf{b}\cdot\mathbf{x} \end{pmatrix},$$

it follows that Eqs. (29) express p and q in terms of the given x, x', y, and y' and the unknowns $\mathbf{a}$,$\mathbf{b}$, and θ. Elimination of p and q between (29) and (30) yields a set of nonlinear equations for $\mathbf{a}$, $\mathbf{b}$, and θ.

If we have solved for $\mathbf{a}$, $\mathbf{b}$, and θ, the multiplication of (27) by U^{D^2} and x^T yields ϕ in the form

$$\cosh\phi = x'U^{D^2}x/(xU^{D^2}x). \quad (31)$$

In the special case where $\theta = \pi$, Eqs. (29) reduce to

$$p = x' + x, \quad q = y' + y,$$

and then the nonlinearity vanishes and the solution for $\Lambda_1(\mathbf{a},\mathbf{b},\pi,\phi)$ follows the simple pattern of Sec. II.

Similarly, elimination of θ by the multiplication of (28) by U^D yields

$$U^DP = 0 = U^DQ,$$

where

$$\begin{aligned} P &= x' - x\cosh\phi - (\sinh\phi)U^Dx, \\ Q &= y' - y\cosh\phi - (\sinh\phi)U^Dy. \end{aligned} \quad (32)$$

Now the situation is similar to that in (21)–(24), and therefore

$$\begin{aligned} K\mathbf{a} &= \mathbf{P} \wedge \mathbf{Q}, \\ K\mathbf{b} &= i(P_4\mathbf{Q} - Q_4\mathbf{P}). \end{aligned} \quad (33)$$

Since

$$U^Dx = \begin{pmatrix} -\mathbf{b} \wedge \mathbf{x} - i\mathbf{a}x_4 \\ i\mathbf{a}\cdot\mathbf{x} \end{pmatrix},$$

it follows that elimination of P and Q between (32) and (33) yields a set of nonlinear equations for the solution of $\mathbf{a}$, $\mathbf{b}$, and ϕ.

The formula for θ, corresponding to (31), is given by

$$\cos\theta = x'U^2x/(xU^2x). \quad (34)$$

In the general case where $\theta \neq 0$, $\theta \neq \pi$, no success was attained with an analytical solution of (30) and (33). This holds also for attempts to solve for $\mathbf{a}$ and $\mathbf{b}$ in terms θ or ϕ.

Using the NAG Fortran routine C05NBF, numerical solutions of (30) for $\mathbf{a}$ and $\mathbf{b}$ for arbitrary θ and of (33) for $\mathbf{a}$ and $\mathbf{b}$ for arbitrary ϕ were obtained. The missing ϕ or θ follows from (31) and (34). These two sets of values of $\mathbf{a}$, $\mathbf{b}$, θ, and ϕ suggest that two arbitrary initial vectors and the resulting final vectors do not contain enough information to determine the Lorentz operator uniquely.

Additional information may take the form of the trace of Λ_1, which is given by

$$\mathrm{Tr}\,\Lambda_1(\mathbf{a},\mathbf{b},\theta,\phi) = 2(\cos\theta + \cosh\phi). \quad (35)$$

Attempts to solve the eight equations (30), (31), and (35) for the eight unknowns $\mathbf{a}$,$\mathbf{b}$,θ,ϕ have had moderate success.

The numerical method relies heavily on the accuracy of the initial estimates.

Another source of the missing information could be the introduction of a third arbitrary initial vector z and the resulting final vector z'.

V. THE SINGULAR OPERATOR

We now turn to the operator Λ_1^s discussed before.[3] It is given by

$$\Lambda_1^s(\mathbf{a},\mathbf{b}) = \exp(iU)$$
$$= I + iU - \tfrac{1}{2}U^2, \quad (36)$$

where U and U^D are given by (3) with

$$\mathbf{a}\cdot\mathbf{a} - \mathbf{b}\cdot\mathbf{b} = 0, \quad \mathbf{a}\cdot\mathbf{b} = 0, \quad (37)$$

whence

$$U^3 = 0.$$

As before we assume that in

$$x' = \Lambda_1^s(\mathbf{a},\mathbf{b})x, \quad y' = \Lambda_1^s(\mathbf{a},\mathbf{b})y, \quad (38)$$

where x and y are given arbitrary vectors, x' and y' are also given, while $\Lambda_1^s(\mathbf{a},\mathbf{b})$ is to be determined in terms of x, x', y, and y'. As before,

$$U^D(x'-x) = 0,$$

and therefore

$$k\mathbf{a} = \mathbf{p}\wedge\mathbf{q}, \quad k\mathbf{b} = i(p_4\mathbf{q} - q_4\mathbf{p}), \quad (39)$$

or

$$ikU = pq^T - qp^T,$$

where

$$p = x' - x, \quad q = y' - y.$$

The desired operator is then

$$\Lambda_1^s = I + (1/k)(pq^T - qp^T) + (1/2k^2)(pq^T - qp^T)^2. \quad (40)$$

The only unknown parameter in (40) is the constant k introduced in (39). From (37) and (39) it follows that

$$(pq)^2 - p^2q^2 = 0, \quad (41)$$

leaving k undetermined. The value

$$k = \tfrac{1}{2}(x'+x)q = \tfrac{1}{2}(xy' - x'y) \quad (42)$$

follows when we require (40) to satisfy (38).

According to (41), the initial and final vectors of Λ_1^s define a null space.

According to previous work[3] one would expect the operator (40) to be closely associated with (16) and (26). In particular (40) would be equal to a product of two suitable operators of the type (27). To demonstrate this in the absence of a presentation of the spinor operator in terms of initial and final states would require a considerable amount of algebra.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the assistance of B. M. Herbst, J. A. C. Weideman, and M. F. Maritz with the numerical work.

[1]C. B. van Wyk, Am. J. Phys. **47**, 747 (1979).
[2]D. M. Fradkin, J. Math. Phys. **23**, 2520 (1982).
[3]C. B. van Wyk, J. Math. Phys. **27**, 1306 (1986).

A brief study on the transformation of Maxwell equations in Euclidean four-space

E. Zampino
Technicare (Johnson & Johnson), 29100 Aurora Road, Solon, Ohio 44139

(Received 27 March 1985; accepted for publication 14 January 1986)

Electromagnetic-type fields in Euclidean four-space are studied by changing the sign of the time differential term in Faraday's law of induction. Although a covariant set of field equations can be derived, difficulties arise in the case of time-dependent fields.

I. INTRODUCTION

Euclidean four-space and its associated coordinate transformations may provide a feasible model to pursue studies in superluminal physics. The coordinate transformations of this particular four-space are analytical and invertible for observers moving at any constant speed that is finite. This is a result of the fact that the metric of flat Euclidean four-space is positive definite, and there is no light cone. It may be possible to develop a tachyon theory for Euclidean four-space, although there is a great deal of controversy about superluminal electromagnetic fields.[1]

Studies on electromagnetic fields subject to subluminal and superluminal Lorentz transformations are extensive.[2] The fundamental structure of electromagnetic-type fields has been extensively studied by a number of authors.[3] However, an attractive feature of flat Euclidean four-space and its associated transformations is the absence of any preferred reference frames,[4] or imaginary position coordinates.[5] In Sec. II, fundamental definitions pertaining to flat Euclidean four-space are presented. In Sec. III, the transformation equations for electromagnetic-type fields are presented, along with the covariant field equations. In Sec. IV, remarks are made on dynamic fields, which includes the solutions to wave equations for the cases of the vacuum and an arbitrary source distribution. In Sec. V, a summary of this paper is presented.

II. FUNDAMENTAL DEFINITIONS

The Riemannian space metric for a flat Euclidean four-space is expressed as

$$g_{\mu\nu} = \begin{cases} 1, & \mu = \nu, \\ 0, & \mu \neq \nu, \end{cases} \tag{2.1}$$

where $\mu = 0,1,2,3$ and $\nu = 0,1,2,3$. The metric signature is $(+,+,+,+)$. The generalized Riemannian space coordinates are defined to represent Cartesian coordinates. A particular dimension in Euclidean four-space is arbitrarily chosen to represent time:

$$(x^0,x^1,x^2,x^3) = (t,x,y,z)\,. \tag{2.2}$$

Consider two parallel frames S and S' in relative motion along the X direction, such that their origins coincide at the time $t = t' = 0$. The square of a four-space "interval" $d\sigma$, between two infinitesimally separated events, is defined by

$$(d\sigma)^2 = (dt)^2 + \sum_{\mu=1}^{3} (dx^\mu)^2 \tag{2.3}$$

in frame S, and is defined by

$$(d\sigma')^2 = (dt')^2 + \sum_{\mu=1}^{3} (dx^{\mu\prime})^2 \tag{2.4}$$

in frame S'. There exists a set of coordinate transformations (in differential form) defined as

$$dt' = \Lambda(dt + v\,dx)\,, \tag{2.5}$$

$$dx' = \Lambda(dx - v\,dt)\,, \tag{2.6}$$

$$dy' = dy\,, \tag{2.7}$$

$$dz' = dz\,, \tag{2.8}$$

$$\Lambda = (1 + v^2)^{-1/2}\,, \tag{2.9}$$

where v is the relative speed of S' with respect to S, expressed as a fraction of the speed of light. The coordinate transformations above insure that the four-space "interval" is invariant:

$$(d\sigma')^2 = (d\sigma)^2\,. \tag{2.10}$$

The relationship between the Euclidean four-space transformations and the Lorentz transformations is obtained by a simple pure imaginary mapping,

$$x \to ix\,, \tag{2.11}$$

$$x' \to ix'\,, \tag{2.12}$$

$$v \to iv\,, \tag{2.13}$$

$$i = (-1)^{1/2}\,, \tag{2.14}$$

performed on the transformation equations (2.5)–(2.9). Contravariant vectors and partial derivatives are related to their covariant forms by

$$w^\mu = g^{\mu\nu} w_\nu\,, \tag{2.15}$$

$$\partial^\mu = g^{\mu\nu}\,\partial_\nu\,, \tag{2.16}$$

therefore, in a flat Euclidean four-space endowed with the metric (2.1),

$$w^\nu = w_\nu\,, \tag{2.17}$$

$$\partial^\nu = \partial_\nu\,. \tag{2.18}$$

The components of a source density four-vector would be expressed as

$$J^\mu = (\rho, J_x, J_y, J_z)\,, \tag{2.19}$$

where ρ is the scalar source density, and J_x, J_y, J_z are the Cartesian components of a three-space source density vector. The components of an electromagnetic four-vector potential are expressed as

 0022-2488/86/051315-04$02.50

$$A^{\mu} = (-\phi, A_x, A_y, A_z) , \quad (2.20)$$

where ϕ is the scalar potential and A_x, A_y, A_z are the Cartesian components of a three-space vector potential. The minus sign in front of the scalar potential in the first component of A^{μ} is necessary to obtain a set of covariant field equations in Euclidean four-space. The electric-type and magnetic-type fields in Euclidean four-space are connected to a scalar potential and three-space vector potential. In three-vector notation,

$$\mathbf{B} = \nabla \times \mathbf{A} , \quad (2.21)$$

$$\mathbf{E} = \nabla\phi + \frac{\partial \mathbf{A}}{\partial t} , \quad (2.22)$$

$$[\mathbf{B} = (B_x, B_y, B_z),\ \mathbf{E} = (E_x, E_y, E_z),\ \mathbf{A} = (A_x, A_y, A_z)] .$$

The electromagnetic field strength tensor in Euclidean four-space will be expressed as

$$Q^{\mu\nu} = \begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{bmatrix} , \quad (2.23)$$

which is the electromagnetic field strength tensor of the Maxwell theory.[6] The tensor $Q^{\mu\nu}$ obeys the transformation law

$$Q^{\mu'\nu'} = \frac{\partial x^{\mu'}}{\partial x^{\mu}} \frac{\partial x^{\nu'}}{\partial x^{\nu}} Q^{\mu\nu} . \quad (2.24)$$

Expanding the Laplace–Beltrami operator

$$\Box^2 = \frac{1}{\tilde{g}} \frac{\partial}{\partial x^{\mu}} \left\{ \tilde{g} g^{\mu\nu} \frac{\partial}{\partial x^{\nu}} \right\} , \quad (2.25)$$

where

$$\tilde{g} = |\det(g_{\mu\nu})|^{1/2} \quad (2.26)$$

on Euclidean four-space yields

$$\Box^2 = \frac{\partial^2}{\partial t^2} + \nabla^2 , \quad (2.27)$$

where ∇^2 is the usual Laplacian operator. Equation (2.27) is form invariant with respect to tranformations in Euclidean four-space.

III. ELECTROMAGNETIC FIELD EQUATIONS

Maxwell's equations are not form invariant with respect to coordinate transformations of Euclidean four-space [(2.5)–(2.9)]. Maxwell's equations can be "fixed up" by changing the sign of the time differential term in Faraday's law of induction. This simple sign change has important consequences; it changes the meaning of the induction law of electrodynamics and the form of wave equations. (More will be said on this subject in Sec. IV.)

Consider a frame S' (as defined in Sec. II) boosted to a speed v, relative to a parallel frame S. The field equations in frame S' (expressed in three-vector notation) are assumed to be

$$\nabla' \cdot \mathbf{E}' = 4\pi\rho' , \quad (3.1)$$

$$\nabla' \cdot \mathbf{B}' = 0 , \quad (3.2)$$

$$\nabla' \times \mathbf{E}' = \frac{\partial \mathbf{B}'}{\partial t'} , \quad (3.3)$$

$$\nabla' \times \mathbf{B}' = 4\pi \mathbf{J}' + \frac{\partial \mathbf{E}'}{\partial t'} . \quad (3.4)$$

The differential operators, and components of the electric and magnetic-type fields may be transformed from the boosted frame S' to the lab frame S. The chain rule of partial differentiation can be applied utilizing the inverse coordinate transformations of Euclidean four-space. The following set of operator transformations are obtained:

$$\frac{\partial}{\partial t'} = v\Lambda \frac{\partial}{\partial x} + \Lambda \frac{\partial}{\partial t} , \quad (3.5)$$

$$\frac{\partial}{\partial x'} = \Lambda \frac{\partial}{\partial x} - v\Lambda \frac{\partial}{\partial t} , \quad (3.6)$$

$$\frac{\partial}{\partial y'} = \frac{\partial}{\partial y} , \quad (3.7)$$

$$\frac{\partial}{\partial z'} = \frac{\partial}{\partial z} . \quad (3.8)$$

Making use of (2.23), (2.24), and the Euclidean four-space coordinate transformations, field components transform as follows:

$$E'_x = E_x , \quad (3.9)$$

$$E'_y = \Lambda(E_y + vB_z) , \quad (3.10)$$

$$E'_z = \Lambda(E_z - vB_y) , \quad (3.11)$$

$$B'_x = B_x , \quad (3.12)$$

$$B'_y = \Lambda(B_y + vE_z) , \quad (3.13)$$

$$B'_z = \Lambda(B_z - vE_y) . \quad (3.14)$$

Finally, the application of operator and field component transformations in Euclidean four-space, combined with the transformation equations for the source density four-vector, will transform the field equations in the boosted frame to

$$\nabla \cdot \mathbf{E} = 4\pi\rho , \quad (3.15)$$

$$\nabla \cdot \mathbf{B} = 0 , \quad (3.16)$$

$$\nabla \times \mathbf{E} = \frac{\partial \mathbf{B}}{\partial t} , \quad (3.17)$$

$$\nabla \times \mathbf{B} = 4\pi \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t} . \quad (3.18)$$

The above result shows that the Maxwell equations containing a sign modification for Faraday's law of induction are form invariant with respect to coordinate transformations in Euclidean four-space. An important special case of the above result is that of static electromagnetic fields. In Eqs. (3.17) and (3.18) the time-dependent derivatives can be set equal to 0, which yields the Maxwell equations for static fields. The important implication of this is that the Maxwell equations for static electromagnetic fields are applicable for both Minkowski space-time and Euclidean four-space. Just as impressive is the fact that a covariant formulation of electromagnetic-type field equations is possible in Euclidean four-space. Utilizing the definitions for the source density four-vector, four-vector potential, field strength tensor, and Laplace–Beltrami operator found in Sec. II, the covariant equations are stated in four-vector notation as

$$\partial_\mu J^\mu = 0\,, \tag{3.19}$$

$$\partial_\mu A^\mu = 0\,, \tag{3.20}$$

$$\partial_\mu Q^{\mu\nu} = 4\pi J^\nu\,, \tag{3.21}$$

$$\partial^\mu Q^{\nu\kappa} + \partial^\nu Q^{\kappa\mu} + \partial^\kappa Q^{\mu\nu} = 0\,, \tag{3.22}$$

$$\square^2 A^\mu = -4\pi J^\mu\,, \tag{3.23}$$

where $\mu,\nu,\kappa = 0,1,2,3$.

Briefly stated, (3.19) is the continuity equation, and (3.20) is the Euclidean four-space analog of the Lorentz condition. The field equations (3.21) and (3.22) yield the Maxwell equations on Euclidean four-space and (3.23) yields the "wave" equations in terms of potentials. It may appear that a theory of electromagnetic-type fields in Euclidean four-space could be constructed that is "parallel" to the Maxwell theory. However, some of the simple consequences of the above field equations are problematic, and this can be demonstrated by some simple examples.

IV. REMARKS ON TIME-DEPENDENT FIELDS

Although the static electromagnetic fields in Euclidean four-space can be described by Maxwell equations, there are at least two problems in case of dynamic fields. Consider the equation,

$$\nabla\times\mathbf{E} = \frac{\partial \mathbf{B}}{\partial t}\,, \tag{4.1}$$

where the sign preceding the time derivative term is deliberately changed to achieve form invariance in Euclidean four-space. The relation (4.1) (by way of elementary arguments utilizing Stokes' theorem) leads directly to the statement

$$\mathscr{E} = \frac{d\Phi}{dt}\,, \tag{4.2}$$

where $\mathscr{E}$ is the emf and Φ is the magnetic flux. The emf and induced current in a metal loop, produced from a time rate of change in magnetic flux, would be in a direction opposite to that specified by Lenz's law. The positive sign in the time differential term, therefore, implies that the magnetic field produced by an induced current is in a direction such as to reinforce the original change in magnetic flux that produces it. This reinforcement effect will increase the total magnetic flux through the loop, which, in turn, would increase the emf. The result would be a "runaway" magnetic flux and emf. Difficulties arise with the solutions to wave equations as well. The vacuum wave equation for the electric field in Euclidean four-space is

$$\nabla^2\mathbf{E} + \frac{\partial^2\mathbf{E}}{\partial t^2} = 0\,, \tag{4.3}$$

which is derived by taking the curl on both sides of field equation (3.17) and utilizing (3.15) and (3.18) for the vacuum case. Following the approach of Reitz and Milford,[7] the complex solution for the electric field is

$$\mathbf{E}(\mathbf{x},t) = \mathbf{E}_0 \exp(\pm\omega z)\exp(-i\omega t)\,, \tag{4.4}$$

where $\omega > 0$. (A similar solution for $\mathbf{B}$ can be obtained.) Monochromatic wave solutions for the vacuum in Euclidean four-space are either exponentially growing or decaying as a function of distance along the direction of propagation. The wave equations in terms of potentials and in the presence of sources are given by (3.23). The equations are of the general form

$$\nabla^2\psi + \frac{\partial^2\psi}{\partial t^2} = 4\pi\alpha f(\mathbf{x},t)\,, \tag{4.5}$$

where $f(\mathbf{x},t)$ is a known source distribution. The constant α is either $+1$ or -1. A Green's function is assumed to exist such that

$$\left[\nabla^2 + \frac{\partial^2}{\partial t^2}\right] G(\mathbf{x}',t';\mathbf{x},t) = 4\pi\alpha\delta(\mathbf{x}-\mathbf{x}')\delta(t-t') \tag{4.6}$$

and

$$\psi(\mathbf{x},t) = \int\int dt'\,d^3\mathbf{x}'\,G(\mathbf{x}',t';\mathbf{x},t)f(\mathbf{x}',t')\,. \tag{4.7}$$

Using the straightforward approach by Butkov,[8] we have

$$G(\mathbf{x}',t';\mathbf{x},t) = \begin{cases} \dfrac{\alpha}{\pi}\dfrac{1}{|\mathbf{x}-\mathbf{x}'|^2 + (t-t')^2}\,, & t' < t\,, \\ 0\,, & t' > t\,, \end{cases} \tag{4.8}$$

and the general solution is

$$\psi(\mathbf{x},t) = \frac{\alpha}{\pi}\int d^3\mathbf{x}'\int_{-\infty}^{t} dt'\,\frac{f(\mathbf{x}',t')}{|\mathbf{x}-\mathbf{x}'|^2 + (t-t')^2}\,. \tag{4.9}$$

The time-dependent integration in the general solution shown above can be performed, and as a result, the solution can be further reduced to the form

$$\psi(\mathbf{x},t) = \alpha\int d^3\mathbf{x}'\,\frac{f(\mathbf{x}',t\pm i|\mathbf{x}-\mathbf{x}'|)}{|\mathbf{x}-\mathbf{x}'|}\,. \tag{4.10}$$

V. SUMMARY

Changing the sign of the time differential term in the Maxwell equation, which expresses Faraday's law of induction, permits a set of field equations that are form invariant with respect to coordinate transformations in Euclidean four-space. Static electromagnetic-type fields are well described by Maxwell equations in a flat Euclidean four-space. A covariant set of electromagnetic-type field equations can be found in Euclidean four-space. However, changing the sign in Faraday's law of induction changes the meaning of the induction law and the form of wave equations. Monochromatic solutions to the wave equations in the Euclidean four-space vacuum (with no boundary conditions applied) contradict what is predicted by the Maxwell theory. The solutions for the vector and scalar potentials contains a Euclidean four-space analog of advanced and retarded time that is complex valued.

ACKNOWLEDGMENTS

A special thanks is expressed to the referee for his very helpful critique. Gratitude is expressed to Dr. J. Lock for his encouragement and many helpful discussions associated to this work.

The author would also like to thank Technicare Corporation for providing financial assistance.

[1]O. P. S. Negi and B. S. Rajput, Phys. Lett. B **113**, 180 (1982).
[2]O. P. S. Negi, H. C. Chandola, K. D. Purohit, and B. S. Rajput, Phys. Lett. B **105**, 281 (1981).
[3]S. Weinberg, Phys. Rev. B **138**, 988 (1965).
[4]A. Antippa, Phys. Rev. D **11**, 724 (1975).
[5]O. P. S. Negi and B. S. Rajput, J. Math Phys. **23**, 1964 (1982).
[6]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed., Sec. 11.9.
[7]J. R. Reitz and F. J. Milford, *Foundations of Electromagnetic Theory* (Addison–Wesley, Reading, MA, 1967), 2nd ed., Sec. 15-8, p. 313.
[8]E. Butkov, *Mathematical Physics* (Addison–Wesley, Reading, MA, 1968), Sec. 14.7, p. 621.

Geometric quantization and constrained systems

Abhay Ashtekar[a)] and Matthew Stillerman
Physics Department, Syracuse University, Syracuse, New York 13210

(Received 1 August 1985; accepted for publication 5 December 1985)

The problem of obtaining the quantum theory of systems with first class constraints is discussed in the context of geometric quantization. The precise structure needed on the constraint surface of the full phase space to obtain a polarization on the reduced phase space is displayed in a form that is particularly convenient for applications. For unconstrained systems, *any* polarization on the phase space leads to a mathematically consistent quantum description, although not all of these descriptions may be viable from a physical standpoint. It is pointed out that the situation is worse in the presence of constraints: a general polarization on the full phase space need not lead to even a mathematically consistent quantum theory. Examples are given to illustrate the general constructions as well as the subtle difficulties.

I. INTRODUCTION

Let us begin with a brief summary of the geometric quantization scheme and the Hamiltonian formulation of classical systems with first class constraints. This discussion will serve as a platform on which the new results of the paper are based.[1]

A. Geometric quantization

Geometric quantization[2,3] offers a systematic procedure to isolate the new input that is required in the passage from the classical description of a physical system to its quantum description. Let Γ be the $2n$-dimensional manifold, equipped with a symplectic structure Ω_{ab}, representing the classical phase space of a given physical system.[4] To obtain its quantum description via geometric quantization, one proceeds in two steps. The first is called *prequantization*. One begins with the complex vector space of cross sections ψ of a U(1) bundle over Γ and introduces on it a Hermitian inner product using the natural (i.e., Liouville) measure on Γ induced by Ω_{ab}. The elements of the resulting Hilbert space H_p are called *prequantum states*. The remarkable thing is that, given *any* classical observable, i.e., a real-valued function f on Γ, there exists a (densely defined) symmetric operator $\mathbf{O}_f$ on H_p such that the association $f \to \mathbf{O}_f$ is one to one, linear, and maps Poisson brackets to commutators:

$$[\mathbf{O}_f, \mathbf{O}_g] = (\hbar/i)\mathbf{O}_{\{f,g\}}, \quad (1)$$

where $\{f, g\}$ is the Poisson bracket between the classical observables f and g. The explicit expression of $\mathbf{O}_f$ is given by

$$\mathbf{O}_f \circ \psi := (\hbar/i)\, X_f^b \nabla_b\, \psi + f\psi \quad (2)$$

$$\equiv (\hbar/i)\Omega^{ab}(\partial_a f)\nabla_b\, \psi + f\psi, \quad (2')$$

where $X_f^b \equiv \Omega^{ab}\,\partial_a f$ is the Hamiltonian vector field of f, and where ∇ is a U(1) connection with curvature Ω_{ab}; $2i\hbar\nabla_{[a} \times \nabla_{b]}\, \psi \equiv \Omega_{ab}\, \psi$. In the terminology more commonly used in physics, ψ can be considered as a U(1) Higgs scalar (with charge $1/\hbar$) and Ω_{ab} as a U(1) Yang–Mills field on Γ. (Recall that, since Ω_{ab} is a symplectic structure, it is curl-free.) For simplicity, let us assume that Ω_{ab} admits a global potential[5] A_a; $\Omega_{ab} = 2\,\partial_{[a} A_{b]}$. One can "trivialize" ψ to obtain a complex-valued function on Γ (which we also denote by ψ), and, given a specific choice of A_a, express the covariant ∇ as

$$\nabla_a\, \psi = (\partial_a - (i/\hbar)\, A_a)\psi. \quad (3)$$

Under a "gauge" transformation,

$$A_a \to A_a + \partial_a\, \Lambda, \quad (4)$$

the complex function ψ transforms via

$$\psi \to (\exp(i/\hbar)\Lambda)\psi. \quad (4')$$

This framework is for mathematical convenience only; there are, of course, no *physical* gauge fields in the system under consideration.

Although the above prequantum description has many appealing features, it is *not* the quantum description. Indeed, if the phase space is $2n$ dimensional, the quantum states are functions of only n variables. The prequantum states, on the other hand, correspond (on trivialization, or, on "gauge" fixing) to functions on *phase space*, i.e., of $2n$ variables. Thus, to obtain quantum states from the prequantum ones, we must impose additional conditions that get rid of the dependence of ψ on n of the $2n$ variables.

This is achieved in the second step of the geometric quantization program. Introduce, at each point of Γ, an n-dimensional subspace of the tangent space such that[6] (i) the pullback of Ω_{ab} to this subspace vanishes identically and (ii) the subspaces are integrable. Thus, the integral manifolds of these subspaces provide us with a foliation of Γ by Lagrangian submanifolds. The assignment of such a subspace to each point of Γ is called a *polarization*. We shall denote by $\mathbf{P}(\gamma)$ the subspace in the tangent space of γ selected by a polarization $\mathbf{P}$. Given a polarization $\mathbf{P}$, one can consider the U(1) Higgs scalars ψ on Γ satisfying

$$\mathscr{L}_V\, \psi := V^a\, \nabla_a\, \psi = 0, \quad (5)$$

for all vector fields V^a that lie in $\mathbf{P}(\gamma)$ at any point γ of Γ. Note that, because of the use of the "gauge"-covariant derivative, (5) is "gauge" invariant. Since $V^a|_\gamma$ is *any* vector in the n-dimensional subspace $P(\gamma)$ at γ, (5) implies that ψ is in effect a function of only n variables. Condition (i) above on $P(\gamma)$ ensures that the n variables on which ψ effectively depends form a complete set of commuting observables, while (ii) ensures that (5) admits a "sufficient number" of

[a)] Alfred P. Sloan Research Fellow.

 0022-2488/86/051319-12$02.50

solutions. The Hilbert space of *quantum* states is to consist of suitably normalizable solutions to (5). (Note, however, that, in general, ψ satisfying (5) will fail to be square integrable w.r.t. the Liouville measure on Γ. Therefore, from the viewpoint of the final quantum theory, the prequantum Hilbert space structure H_p is spurious.[6]) To obtain the quantum Hilbert space **H** we proceed as follows. Note, first, that since the pullback of Ω_{ab} to the Lagrangian submanifolds of the polarization vanishes, we can always choose a "gauge" such that the pullback of A_a to these submanifolds also vanishes,[7] i.e., such that

$$V^a A_a = 0, \tag{6}$$

for all V^a tangential to P. In this "gauge," Higgs scalars ψ satisfying the polarization condition (5) are pullbacks to Γ of complex-valued functions—which we denote again by ψ—on the space $\mathscr{C}$ of Lagrangian submanifolds of **P**. ($\mathscr{C}$ is an n-manifold and may be thought of as the "*configuration space*" selected by the polarization **P**. Note, however, that, in general, Γ cannot be identified with the cotangent bundle over $\mathscr{C}$ since the Lagrangian submanifolds of **P** need not be vector spaces.) Choose on $\mathscr{C}$ a nowhere vanishing n-form $\epsilon_{\alpha_1\cdots\alpha_n}$ and set

$$\langle \psi,\chi \rangle = \int_{\mathscr{C}} \psi^* \chi \epsilon_{\alpha_1\cdots\alpha_n} dS^{\alpha_1\cdots\alpha_n}. \tag{7}$$

The complex Hilbert space obtained by Cauchy completion of the space of Higgs scalars satisfying (5) is the space of *quantum* states. Denote it by **H**.

Next, we define a class of quantum operators. Let f be a classical observable whose Hamiltonian vector field X_f^a preserves the given polarization **P**, i.e., satisfies

$$\mathscr{L}_V X_f|_\gamma \in \mathbf{P}(\gamma), \tag{8}$$

for all vector fields V^a tangential to **P**. Then, X_f^a on Γ can be projected down unambiguously to a vector field X_f^a on $\mathscr{C}$. Define the divergence, Div X_f, of X_f^a on $\mathscr{C}$ via

$$\mathscr{L}_{X_f} \epsilon_{\alpha_1\cdots\alpha_n} = (\text{Div}\, X_f)\epsilon_{\alpha_1\cdots\alpha_n}. \tag{9}$$

Now, associated with every classical observable f satisfying (8) we define a (densely defined) quantum operator **F** (see Ref. 8):

$$\mathbf{F}\circ\psi := (\hbar/i)\, X_f^\alpha \partial_\alpha \psi + (f - X_f^a A_a)\psi + (\hbar/2i)(\text{Div}\, X_f)\psi \tag{10}$$

$$\equiv \mathbf{O}_f \circ \psi + (\hbar/2i)(\text{Div}\, X_f)\psi. \tag{10'}$$

It is straightforward to check that if ψ satisfies the polarization condition (5), so does $\mathbf{F}\circ\psi$, and that the addition of the factor $(\hbar/2i)(\text{Div}\, X_f)$ to $\mathbf{O}_f$ makes **F** symmetric w.r.t. the inner product (7). Thus, given a polarization **P**, we obtain a class of classical observables f that have unambiguous quantum analogs **F**.[9] For this class, the Poisson brackets do go over to quantum commutators. If

$$\{f,f'\} = f'', \tag{11}$$

then

$$[\mathbf{F},\mathbf{F}'] = (\hbar/i)\mathbf{F}''. \tag{11'}$$

There *are* ways of associating quantum operators with certain classical observables that fail to satisfy (8), an outstanding example being the procedure introduced by Blattner, Kostant, and Sternberg (BKS).[10] However, this association fails to preserve the relation between Poisson brackets and commutators.

Finally, we note that although we had to introduce a nowhere vanishing n-form $\epsilon_{\alpha_1\cdots\alpha_n}$ on $\mathscr{C}$ in addition to the polarization **P** on Γ, the resulting quantum theory is independent of the specific choice of the n-form. Choose another nowhere vanishing n-form $\hat{\epsilon}_{\alpha_1\cdots\alpha_n}$ on $\mathscr{C}$. Since $\mathscr{C}$ is an n-manifold, there exists a nowhere vanishing function μ such that $\hat{\epsilon}_{a\cdots a} = \mu\epsilon_{a\cdots a}$. Hence, the mapping Λ: $\psi \to \hat{\psi} := |\mu|^{-1/2}\psi$ is an isomorphism from the Hilbert space **H** obtained using $\epsilon_{\alpha_1\cdots\alpha_n}$ in (7) and the Hilbert space $\hat{\mathbf{H}}$ obtained from $\hat{\epsilon}_{\alpha_1\cdots\alpha_n}$. Furthermore, if **F** and $\hat{\mathbf{F}}$ are the quantum operators on **H** and $\hat{\mathbf{H}}$ obtained via (9), (10), and (10′) from a classical observable f, then one has

$$\Lambda\circ\mathbf{F}\circ\Lambda^{-1} = \hat{\mathbf{F}}. \tag{12}$$

Thus, the quantum theories obtained from $\epsilon_{\alpha_1\cdots\alpha_n}$ and $\hat{\epsilon}_{\alpha_1\cdots\alpha_n}$ on $\mathscr{C}$ are naturally isomorphic. (In fact one can avoid the introduction of n-forms altogether by using $\frac{1}{2}$-forms—rather than functions—on $\mathscr{C}$ as quantum states.) On the other hand, the quantum description does depend on the choice of polarization. This is not surprising; many mathematically viable quantum theories can have the same classical limit.

B. Classical systems with first class constraints

In presence of constraints, not all points of the $2n$-dimensional phase space Γ are accessible to the physical system. Let there be m constraints $(m<n)$. Then, the points representing the allowable states form a $(2n-m)$-dimensional submanifold $\bar{\Gamma}$ of Γ. $\bar{\Gamma}$ will be referred to as the *constraint surface*. The constraints are said to be *first class* if and only if $\bar{\Gamma}$ is such that, given any covector n_a (at any point $\bar{\gamma}$ of $\bar{\Gamma}$) normal to $\bar{\Gamma}$, $\Omega^{ab}n_b$ is tangential to $\bar{\Gamma}$. In this paper, we shall restrict ourselves to systems with first class constraints.[11]

By introducing a suitable chart on Γ, $\bar{\Gamma}$ can be (locally) characterized by m equations, $C_i(\gamma) = 0$, $i = 1,\dots, m$. Since the covector fields $\partial_a C_i$ are normal to $\bar{\Gamma}$, it follows that the m Hamiltonian vector fields X_i^a defined by the constraint functions, $X_i^a = \Omega^{ba}\partial_b C_i$, are tangential to $\bar{\Gamma}$. These X_i^a will be referred to as *constraint vector fields*. (More generally, a vector field X^a, tangential to $\bar{\Gamma}$, is a constraint vector field if $\Omega_{ab}X^b$ is everywhere normal to $\bar{\Gamma}$.) Since X_i^a is tangential to $\bar{\Gamma}$, it follows that $X_i^a \circ \partial_a C_j = 0$, for all $j = 1,2,\dots, m$. Now, since

$$X_i^a \partial_a C_j \equiv \Omega^{ba}(\partial_b C_i)(\partial_a C_j) \tag{13}$$

$$\equiv \{C_i, C_j\}, \tag{13'}$$

it follows that the Poisson bracket $\{C_i, C_j\}$ between any two constraint functions vanishes on the constraint surface:

$$\{C_i, C_j\} = f_{ij}{}^k C_k \cong 0, \tag{14}$$

for some functions $f_{ij}{}^k$ on Γ, where $\cong$ stands for "*equals at points of* $\bar{\Gamma}$ *to*." [Thus the first equality in (14) holds on a neighborhood in Γ of $\bar{\Gamma}$, while the second holds at points of $\bar{\Gamma}$ only.] Equation (14) says that, if the constraints are first class, the constraint functions are closed under the Poisson bracket. It is straightforward to check that the converse is also true. In fact Eq. (14) is often used to *define* first class constraints. We have proceeded in a different manner for aesthetic reasons: our definition does not require the introduction of a chart, i.e., a specific choice of m function C_i on Γ to define $\overline{\Gamma}$.

Recall that the Poisson bracket Lie algebra between functions on Γ is intimately related to the Lie algebra of Hamiltonian vector fields:

$$[X_f, X_g]^a = X^a_{\{f,g\}}, \tag{15}$$

for any functions f and g on Γ, where [,] on the left side is the Lie bracket between vector fields. Hence, Eq. (14) implies that, on $\overline{\Gamma}$, the constraint vector fields X^a_i are closed under the Lie bracket. Hence the m-flats formed by the constraint vector fields in the tangent space of any point of $\overline{\Gamma}$ are integrable. In the Dirac treatment of constrained systems, motions along constraint vector fields correspond to gauge.[12] We shall, therefore, refer to the m-manifolds spanned by the constraint vector fields as *gauge equivalence classes*. Thus, in presence of first class constraints, not only are the points outside $\overline{\Gamma}$ not accessible to the system, but even within $\overline{\Gamma}$, distinct points do not represent physically distinct states. A physical state is represented by a gauge equivalence class of points in $\overline{\Gamma}$.

Consider, therefore, the space $\hat{\Gamma}$ each point of which represents a gauge equivalence class. The space $\hat{\Gamma}$ can be naturally endowed with the structure of a $2(n-m)$-dimensional manifold. Furthermore, it naturally inherits a symplectic structure $\hat{\Omega}$. To see this, let us first examine the pull-back $\underleftarrow{\Omega}_{ab}$ to $\overline{\Gamma}$ of the symplectic structure Ω_{ab} on Γ. Does $\underleftarrow{\Omega}_{ab}$ have degenerate directions? Let V^a, tangential to $\overline{\Gamma}$, be a degenerate direction. Then, given *any* W^a tangential to $\overline{\Gamma}$,

$$0 = \underleftarrow{\Omega}_{ab}\, V^a\, W^b \cong (\Omega_{ab}\, V^a) W^b, \tag{16}$$

whence $\Omega_{ab}\, V^b$ is normal to $\overline{\Gamma}$. Using the fact that Ω_{ab} is nondegenerate, it follows that V^a is a linear combination of constraint vector fields. Thus, $\underleftarrow{\Omega}_{ab}$ does admit degenerate directions and these are *precisely* the constraint m-flats. Furthermore, given any constraint vector field $X^a_i = \Omega^{ba}\,\partial_b\, C_i$, we have

$$\mathscr{L}_{X_i}\,\Omega_{ab} = 0 \text{ on } \Gamma, \quad \text{whence } \mathscr{L}_{X_i}\,\underleftarrow{\Omega}_{ab} = 0 \text{ on } \overline{\Gamma}. \tag{17}$$

Equations (16) and (17) are necessary and sufficient for $\underleftarrow{\Omega}_{ab}$ to be a pullback to $\overline{\Gamma}$ of a symplectic structure on $\hat{\Gamma}$, which we denote by $\hat{\Omega}_{AB}$. Since each point of $\hat{\Gamma}$ represents a physically distinct state of the system, $(\hat{\Gamma},\hat{\Omega})$ may be thought of as the physical phase space. [Since $\hat{\Gamma}$ is $2(n-m)$ dimensional, the system is physically interesting only if $m<n$.]

C. Statement of the problem

Let us now apply the techniques of geometric quantization to systems with first class constraints. Let Γ be a $2n$-dimensional manifold with symplectic structure Ω_{ab} and let there be m first class constraints, $C_{\mathbf{k}} = 0$, $\mathbf{k} = 1,\ldots, m$ $(m<n)$. Denote the constraint surface by $\overline{\Gamma}$ as before. Using (Γ,Ω), we can carry out prequantization. Since each classical observable defines a prequantum operator unambiguously, it is natural to adopt the following strategy. Impose on the prequantum wave functions, the m prequantum operator constraints

$$\mathbf{O}_{\mathbf{k}} \circ \psi \equiv (\hbar/i)\, X^a_{\mathbf{k}}(\partial_a - (i/\hbar)\, A_a)\psi + C_{\mathbf{k}}\, \psi = 0. \tag{18}$$

These m equations are mutually consistent, i.e., do not imply further constraints on the prequantum wave functions ψ: Because the classical constraints are first class, Eq. (1) implies that the commutator of any two prequantum constraints is again a prequantum constraint. Thus, because constraints exist already at the classical level and because the transition from the classical to the prequantum description is completely unambiguous, it is natural to "take care" of the constraints in the first step of the geometrical quantization program, before one introduces the new structure needed for quantization. Let us now carry out the second step, i.e., introduce a polarization **P** on Γ and require that the quantum wave functions satisfy Eq. (5):

$$\mathscr{L}_V\, \psi := V^a\, \nabla_a\, \psi = 0, \tag{5}$$

for all vector fields V in the polarization **P**. To begin with, ψ is a general complex-valued function of $2n$ variables. Equation (18) imposes m conditions on ψ and Eq. (5) imposes n conditions. Hence, it would appear that the permissible quantum wave functions—which satisfy (18) as well as (5)—are functions of exactly the right number, $(n-m)$, of variables. However, for a general choice of polarization this need not be the case. Although each set of equations, (18) and (5), is "internally" consistent, in general the two sets may *not* be consistent with each other. The consistency conditions add further restrictions on permissible wave functions, which, as a result, depend on *less than* $(n-m)$ independent variables. Indeed, in specific examples, the consistency conditions can be so severe that the space of constrained quantum states is zero dimensional! Thus, in presence of constraints, one cannot just pick out any old polarization on Γ; the polarization has to be compatible with constraints in a suitable sense. Thus, we are led to ask: Given a classical system with constraints, what is the class of polarizations **P** on Γ that can lead to a meaningful quantum theory, i.e., a theory in which the physical wave functions depend precisely on $(n-m)$ independent variables? What is the class of classical observable that gets promoted unambiguously to quantum operators?

Recall that, in the classical description, one can simply get rid of constraints by working on the reduced phase space $(\hat{\Gamma},\hat{\Omega})$. Since points of $\hat{\Gamma}$ represent physical states of the system, one might attempt to apply the geometric quantization procedure directly to $(\hat{\Gamma},\hat{\Omega})$. The key step now is to introduce a polarization $\hat{\mathbf{P}}$ on $\hat{\Gamma}$. In practice, however, it is usually awkward to work directly on $\hat{\Gamma}$ since points of $\hat{\Gamma}$ represent equivalence classes of points on $\overline{\Gamma}$. One is therefore led to ask: What exactly is the structure that we must introduce on $\overline{\Gamma}$ in order to obtain a polarization $\hat{\mathbf{P}}$ on $\hat{\Gamma}$? This structure will be called a *constrained polarization* and denoted by $\overline{\mathbf{P}}$. It is important to obtain $\overline{\mathbf{P}}$ in a form that is convenient to use in practice.

Finally, we are led to ask for the relation between the *prequantum operator constraint method* [Eqs. (18) and (5)] based on Γ and the *reduced phase-space method* based on $\hat{\Gamma}$. Does a polarization $\mathbf{P}$ on Γ, which leads to a meaningful quantum description via (18) and (5), automatically induce a structure $\overline{\mathbf{P}}$ on $\overline{\Gamma}$ that is equivalent to a polarization $\hat{\mathbf{P}}$ on $\hat{\Gamma}$? If so, is there a simple relation between the quantum theories obtained from $\mathbf{P}$ on (Γ,Ω) and $\hat{\mathbf{P}}$ on $(\hat{\Gamma},\hat{\Omega})$?

The purpose of this paper is to answer these questions.

II. POLARIZATIONS ON Γ, $\overline{\Gamma}$, AND $\hat{\Gamma}$

This section is divided into four parts. In the first, we present a simple example in which the difficulties associated with the incompatibility of Eqs. (5) and (18) appear explicitly. In the second, we develop a general scheme in which these difficulties are avoided by working on the constraint surface $\overline{\Gamma}$. It turns out that the scheme is equivalent to the reduced phase space method, i.e., to the standard geometric quantization on $(\hat{\Gamma},\hat{\Omega})$. In the third part, we apply the general techniques to the example introduced in the beginning of this section and construct a viable quantum theory of the system in question. In the last part, we return to the full phase space Γ and specify conditions that a polarization $\mathbf{P}$ on Γ must satisfy in order to obtain a viable quantum description via (5) and (18). It turns out that these conditions are also sufficient to ensure that $\overline{\Gamma}$ is equipped with just the structure needed to induce a polarization on the reduced phase space $\hat{\Gamma}$. We conclude the section with a discussion of the relation between the reduced phase space method and the prequantum operator constraint method.

A. Example: Prequantum operator constraints

Consider a particle and let its configuration space $\mathscr{C}$ be the interior of the future null cone of a point 0 of Minkowski space.[13] The phase space Γ is the cotangent bundle over $\mathscr{C}$. Next, we introduce certain constraints:

$$C_1(q,p) \equiv \eta^{\alpha\beta} p_\alpha p_\beta = 0, \tag{19}$$

$$C_2(q,p) \equiv v^\alpha p_\alpha = 0, \tag{20}$$

where $\eta^{\alpha\beta}$ is the Minkowskian metric and v^α is a vector field on Minkowski space to be specified shortly. Let us compute the Poisson brackets between C_1 and C_2. We have

$$\{C_1,C_2\} = -(\mathscr{L}_v \eta^{\alpha\beta})\, p_\alpha p_\beta. \tag{21}$$

Hence, if v^α is a conformal killing field of $\eta^{\alpha\beta}$, i.e., if $\mathscr{L}_v \eta_{\alpha\beta} = 2\phi\, \eta_{\alpha\beta}$ for some function ϕ, the right side of (21) would vanish on the constraint surface; the constraints would be first class. To be specific, we shall choose v^α as follows:

$$v^\alpha(q) = (k_\beta q^\beta) q^\alpha - \tfrac{1}{2}(q_\beta q^\beta) k^\alpha, \tag{22}$$

where k^α is a constant, unit, spacelike vector field and q^α is the position vector of a point in Minkowski space with respect to the origin[13,14] 0.

Let us attempt to apply the prequantum operator constraint method to this system. We begin with U(1) Higgs scalars $\psi(q,p)$ on Γ. It will be convenient to use the gauge in which the potential A_a of Ω_{ab} is given by $A_a = p_\alpha \partial_a q^\alpha$. Then, the prequantum operator constraints (18) yield

$$\mathbf{O}_1 \circ \psi \equiv 2\frac{\hbar}{i}\eta^{\alpha\beta} p_b \frac{\partial}{\partial q^\alpha}\psi - (\eta^{\alpha\beta} p_\alpha p_\beta)\psi = 0, \tag{23}$$

$$\mathbf{O}_2 \circ \psi \equiv \frac{\hbar}{i}\left(v^\alpha \frac{\partial}{\partial q^\alpha}\psi - p_\beta \frac{\partial v^\beta}{\partial q^\alpha}\frac{\partial}{\partial p_\alpha}\psi\right) = 0. \tag{24}$$

Because of Eq. (1), it follows that $[\mathbf{O}_1,\mathbf{O}_2]$ automatically annihilates ψ if (23) and (24) hold, whence the system of prequantum constraints is closed; there are no new consistency conditions. Next, let us choose a polarization $\mathbf{P}$ on Γ. An obvious choice is the vertical polarization:

$$V \in \mathbf{P} \quad \text{iff} \quad V \equiv f_\alpha(q,p)\frac{\partial}{\partial p_\alpha}, \tag{25}$$

for *some* functions f^α on Γ. Then, the pullback of our A_a to the polarization flats $\mathbf{P}(\gamma)$ *is* zero for all γ in Γ and the polarization condition on the wave functions reduces to

$$\frac{\partial}{\partial p_\alpha}\psi(q,p) = 0 \quad \Leftrightarrow \quad \psi \equiv \psi(q). \tag{26}$$

Using (26) one can simplify (24) to obtain

$$\mathbf{O}_2 \circ \psi \equiv \frac{h}{i} v^\alpha \frac{\partial}{\partial q_\alpha}\psi = 0 \tag{24'}$$

on polarized wave functions. We now ask if (26) is consistent with (23) and (24'). There is no consistency problem between (26) and (24'). Together, the two conditions simply say that ψ is a function on the configuration space $\mathscr{C}$, which, in addition, is constant along the integral curves of v^α. However, there is clearly a problem with (23) and (26). The only $\psi \equiv \psi(q)$ which satisfies (23) everywhere on Γ is the one which vanishes identically!

Remark: One might attempt to rectify the situation by noting that the problem arises simply because $\mathbf{O}_1$ maps polarized wave functions—i.e., wave functions that depend only on q—to those that depend on both q and p. One may, therefore, first try to define an operator which "corresponds to" the classical observable $C_1(q,p)$, which, in addition, preserves the space of polarized wave functions, and use it in (23) in place of $\mathbf{O}_1$. Now, due to factor ordering problems, a general classical observable does not have an unambiguous operator analog. However, $C_1(q,p)$ is fortunately a simple function: it is quadratic in momentum with a nondegenerate metric $(\eta^{\alpha\beta})$ as coefficients. For this set of classical observables and the vertical polarization, the Blatnner–Kostant–Sternberg (BKS) procedure[10] is applicable. The unambiguous quantum operator $\mathbf{O}_1$ is simply the wave operator (since scalar curvature of $\eta^{\alpha\beta}$ is identically zero): we replace (23) by

$$\mathbf{O}_1' \circ \psi = -\hbar^2 \Box \psi = 0. \tag{23'}$$

This modified operator constraint is clearly compatible with the polarization condition (26). However, now there is a new problem: (23') is incompatible with (24')! This comes about because the BKS modification procedure destroys the simple relation [Eq. (1)] between the Poisson brackets and the commutators. More precisely, an explicit calculation yields

$$[\mathscr{L}_v, \Box] = 2\,\mathscr{L}_k,$$

where k^α is the constant vector field on Minkowski space used in the definition of the conformal killing field v^α [Eq. (22)]. Hence, if ψ satisfies the polarization condition (26), Eqs. (23') and (24') imply

$$0 = [\mathbf{O}_2', \mathbf{O}_1']\,\psi = -(\hbar^3/i)2k^\alpha\,\partial_\alpha\,\psi, \tag{27}$$

which is a new constraint on the permissible ψ. Further, (27) and (24') now imply

$$(k_\beta\, q^\beta)\, q^\alpha\,\partial_\alpha\,\psi = 0,$$

which, together with (26), implies the ψ must[15] be constant on $\mathscr{C}$! Thus, again the space of permissible wave functions is so small that the resulting quantum theory is trivial.

Note that, classically, there is nothing pathological about this system. It is just that if we use ideas from geometric quantum mechanics to promote classical constraints to the quantum level—and, in general, this seems to be the only systematic procedure available—the use of the vertical polarization leads to a trivial quantum theory. [It is straightforward to check that the use of the horizontal polarization, $V \in \mathbf{P}_{\rm Hor}$ iff $V \equiv f^\alpha(q,p)(\partial/\partial q^\alpha)$ for *some* $f^\alpha(q,p)$, also leads to a trivial theory.] The problem lies in the choice of polarization: In presence of constraints, not all polarizations are permissible.

B. Constrained polarization and the reduced phase-space method

To overcome this difficulty, it is convenient to adopt a somewhat different route than the one outlined above and then return to the prequantum operator constraint method at the end of the discussion.

Let us consider a general, $2n$-dimensional phase space Γ and a $(2n-m)$-dimensional constraint surface $\overline{\Gamma}$ therein. Classically, the physical system has access only to points on $\overline{\Gamma}$. Furthermore, off $\overline{\Gamma}$ there is considerable ambiguity in defining the gauge equivalence classes. Therefore, an alternative strategy is to work entirely on $\overline{\Gamma}$. Let us restrict the prequantum wave functions ψ to $\overline{\Gamma}$. Thus, we obtain complex-valued functions $\underline{\psi}$ of $(2n-m)$ variables. Let us impose the prequantum operator constraints (18) on these $\underline{\psi}$. Since the $\underline{\psi}$ have support only on $\overline{\Gamma}$, where $C_{\rm k}(q,p)=0$, we have

$$\mathbf{O}_k\,\underline{\psi} \equiv (\hbar/i)\, X^a_{\rm k}\,\nabla_a\,\underline{\psi} = 0. \tag{28}$$

These m equations are internally consistent, whence, we now have complex-valued functions $\underline{\psi}$ of only $(2n-2m)$ variables. The quantum wave functions, on the other hand, are to depend only on $(n-m)$ variables. Thus, we need $(n-m)$ additional, polarization-type conditions on $\underline{\psi}$. Furthermore, the new conditions have to be consistent with Eq. (28).

We therefore begin by introducing suitable structure on $\overline{\Gamma}$. A *constrained polarization* $\overline{\mathbf{P}}$ is an assignment to each point $\bar{\gamma}$ of $\overline{\Gamma}$ of an n-dimensional subspace of the tangent space $\overline{T}_{\bar{\gamma}}$ (within $\overline{\Gamma}$) of $\bar{\gamma}$ such that (i) the pullback to $\overline{\mathbf{P}}(\bar{\gamma})$ of Ω_{ab} is zero for all $\bar{\gamma}$ in $\overline{\Gamma}$; (ii) the n-flats $\overline{\mathbf{P}}(\bar{\gamma})$ are integrable; and (iii) $\overline{\mathbf{P}}(\bar{\gamma})$ contains the m gauge directions at $\bar{\gamma}$, for all $\bar{\gamma}$ in $\overline{\Gamma}$. Choose a constrained polarization $\overline{\mathbf{P}}$ and consider those prequantum wave functions $\underline{\psi}$ on $\overline{\Gamma}$ that satisfy

$$\overline{V}^a\,\nabla_a\,\underline{\psi} = 0, \tag{29}$$

for all vector fields $\overline{V}^a$ on $\overline{\Gamma}$ which are everywhere tangential to $\overline{\mathbf{P}}$. We claim that $\underline{\psi}$ on $\overline{\Gamma}$, satisfying Eq. (29), are the appropriate quantum wave functions. To see this, let us examine the three conditions satisfied by $\overline{\mathbf{P}}$. The first two ensure that the $\underline{\psi}$ satisfying (29) are functions of precisely $(2n-m)-n\equiv n-m$, commuting (under Poisson bracket) variables, while the third implies that $\underline{\psi}$ are gauge invariant, i.e., satisfy (28) automatically. The last step is to introduce an Hermitian inner product on the space of solutions to (29). Since $\overline{\mathbf{P}}(\bar{\gamma})$ are integrable n-flats tangential to $\overline{\Gamma}$, we can quotient $\overline{\Gamma}$ by the leaves of $\overline{\mathbf{P}}$. Denote the quotient by $\mathscr{C}$. Since each point of $\mathscr{C}$ represents an integral manifold of $\overline{\mathbf{P}}$, $\mathscr{C}$ is naturally endowed with the structure of a $(n-m)$-manifold. In the "gauge" given by Eq. (6), wave functions $\underline{\psi}$ satisfying Eq. (29) on $\overline{\Gamma}$ are pullbacks to $\overline{\Gamma}$ of complex-valued functions on $\mathscr{C}$, which we again denote by $\underline{\psi}$. Hence, to define an inner product on the space of solutions to Eq. (29), we first introduce, as in Sec. I A, a nowhere vanishing $(n-m)$-form $\bar{\epsilon}_{\alpha_1\cdots\alpha_{n-m}}$ on $\mathscr{C}$ and set

$$\langle \underline{\psi}, \chi \rangle := \int_{\mathscr{C}} \underline{\psi}^*\,\chi\,\bar{\epsilon}_{\alpha_1\cdots\alpha_{n-m}}\, dS^{\alpha_1\cdots\alpha_{n-m}}. \tag{30}$$

The Cauchy completion $\overline{\mathbf{H}}$ of the space of $\underline{\psi}$ [satisfying (29)] w.r.t. (30) is the Hilbert space of quantum states. As in Sec. I A, we can introduce a privileged class of quantum operators. Let f be a classical observable whose Hamiltonian vector field X^a_f is tangential to $\overline{\Gamma}$ and whose action preserves the constrained polarization $\overline{\mathbf{P}}$. Then, $X^a_f|_{\overline{\Gamma}}$ has an unambiguous projection $\overline{X}^a_f$ on $\mathscr{C}$. As before, we define divergence of $\overline{X}^a_f$ by

$$\mathscr{L}_{\overline{X}}\,\bar{\epsilon}_{\alpha_1\cdots\alpha_{n-m}} = (\mathrm{Div}\,\overline{X}_f)\,\bar{\epsilon}_{\alpha_1\cdots\alpha_{n-m}}, \tag{31}$$

and use it to introduce a quantum operator $\overline{\mathbf{F}}$ on $\overline{\mathbf{H}}$:

$$\overline{\mathbf{F}}\circ\underline{\psi} := (\hbar/i)(\overline{X}^a_f\,\partial_a + \tfrac{1}{2}\,\mathrm{Div}\,\overline{X}^a_f)\circ\underline{\psi} + (f - X^a_f A_a)\,\underline{\psi}. \tag{32}$$

By construction, each $\overline{\mathbf{F}}$ is a densely defined, symmetric operator on $\overline{\mathbf{H}}$ and the association $f\to\overline{\mathbf{F}}$ sends Poisson brackets to commutators. [See (11) and (11').] These classical observables f with unambiguous quantum analogs are the *basic* observables associated with $\overline{\mathbf{P}}$. How many such observables are there? Given a vector field $\overline{V}^a$ on $\mathscr{C}$, one can find a *basic* observable f such that $\overline{X}^a_f = \overline{V}^a$. However, f is not uniquely determined by $\overline{V}^a$. To see the ambiguity, let us set $\overline{V}^a = 0$. Then, f is such that its Hamiltonian vector field is tangential to $\overline{\mathbf{P}}$ within $\overline{\Gamma}$. Since each leaf of $\overline{\mathbf{P}}$ is a Lagrangian submanifold of Γ, it follows that f must be the pullback to $\overline{\Gamma}$ of a function on $\mathscr{C}$. Thus, there are "as many" *basic* classical observables as there are functions and vector fields on $\mathscr{C}$. Since one can think of $\mathscr{C}$ as the "physical configuration space" of the system, the class of *basic* observables is a direct generalization of the class naturally available on the cotangent bundle of the configuration space of an unconstrained system.[9] Finally, as in Sec. I A, the quantum theory depends only on the choice of a constrained polarization $\overline{\mathbf{P}}$ on $\overline{\Gamma}$; change of the $(n-m)$-form $\bar{\epsilon}_{\alpha_1\cdots\alpha_{n-m}}$ on $\mathscr{C}$ just provides an isomorphic quantum theory.

To summarize, then, to obtain a mathematically viable quantum theory in which the wave functions depend on the correct—i.e., $(n-m)$—independent variables, one has to enlarge the integrable, degenerate (w.r.t. Ω_{ab}) gauge m-flat in the tangent space of each point of $\bar{\Gamma}$ to an integrable, degenerate n-flat. The distribution of these n-flats gives us, precisely, a constrained polarization.

Although the notion of constrained polarization was motivated by considerations involving the constraint surface $\bar{\Gamma}$ alone, it turns out that there is a one to one correspondence between constrained polarizations on $\bar{\Gamma}$ and polarizations on the reduced phase-space $\hat{\Gamma}$.

Lemma 1.1: Every constrained polarization $\bar{\mathbf{P}}$ on $\bar{\Gamma}$ yields, via the natural projection from $\bar{\Gamma}$ to $\hat{\Gamma}$, a polarization $\hat{\mathbf{P}}$ on $\hat{\Gamma}$.

Proof: Denote the natural projection mapping from $\bar{\Gamma}$ to $\hat{\Gamma}$, which sends each point of $\bar{\Gamma}$ to the gauge-equivalence class it belongs to, by Π. Given any point $\bar{\gamma}$ of $\bar{\Gamma}$, and a tangent vector $\bar{V}^a$ within $\bar{\Gamma}$, its push-forward, $\underset{\rightarrow}{\Pi} \circ \bar{V}^a = :\hat{V}^A$ is a tangent vector at $\Pi \circ \bar{\gamma} = :\hat{\gamma}$ of $\hat{\Gamma}$. Here $\underset{\rightarrow}{\Pi}$ is a linear map whose kernel is precisely the m-dimensional gauge flat at $\bar{\gamma}$. Since $\bar{\mathbf{P}}(\bar{\gamma})$ contains this m-flat, the image $\underset{\rightarrow}{\Pi} \circ \bar{\mathbf{P}}(\bar{\gamma}) = :\hat{\mathbf{P}}(\hat{\gamma})$ is a $(n-m)$-dimensional subspace of the tangent space of $\hat{\Gamma}$. Furthermore, since the n-flats $\bar{\mathbf{P}}(\bar{\gamma})$ are integrable, i.e., since

$$[\bar{V},\bar{W}]^a \in \bar{\mathbf{P}},$$

for all $\bar{V},\bar{W}$ in $\bar{\mathbf{P}}$, it follows that $\underset{\rightarrow}{\Pi} \circ \bar{\mathbf{P}}(\bar{\gamma}_1) = \underset{\rightarrow}{\Pi} \circ \bar{\mathbf{P}}(\bar{\gamma}_2)$ if $\bar{\gamma}_1$ and $\bar{\gamma}_2$ are gauge related, i.e., if $\Pi \circ (\bar{\gamma}_1) = \Pi \circ (\bar{\gamma}_2)$. Thus, under $\underset{\rightarrow}{\Pi}$, the distribution of n-dimensional subspaces $\bar{\mathbf{P}}$ on $\bar{\Gamma}$ projects down unambiguously to a distribution of $(n-m)$-dimensional subspaces $\hat{\mathbf{P}}$ on $\hat{\Gamma}$.

We now wish to show that $\hat{\mathbf{P}}$ are tangential to Lagrangian submanifolds of $\hat{\Gamma}$. By definition of the symplectic structure $\hat{\Omega}_{AB}$ on $\hat{\Gamma}$, we have

$$\hat{\Omega}_{AB}\, \hat{V}^A\, \hat{W}^B|_{\hat{\gamma}} = \Omega_{ab}\, \bar{V}^a\, \bar{W}^b\,|_{\bar{\gamma}}.$$

Now, by construction, $\hat{V}$ and $\hat{W}$ belong to $\hat{\mathbf{P}}$ iff $\bar{V}$ and $\bar{W}$ belong to $\bar{\mathbf{P}}$. Since the pullback of Ω_{ab} to $\bar{\mathbf{P}}$ vanishes, so does the pullback of $\hat{\Omega}_{AB}$ to $\hat{\mathbf{P}}$. It only remains to show that the $(n-m)$-flats $\hat{\mathbf{P}}(\hat{\gamma})$ are integrable on $\hat{\Gamma}$. Let $\hat{f}$ be *any* function on $\hat{\Gamma}$ and $\bar{f}:=\underset{\leftarrow}{\Pi} \circ \hat{f}$, its pullback under Π. Then, given any $\bar{V}^a$ and $\bar{W}^a$ in $\bar{\mathbf{P}}$, with $[\bar{V},\bar{W}]^a = :\bar{X}^a$ in $\bar{\mathbf{P}}$, we have

$$[\hat{V},\hat{W}]^A\, \partial_A \hat{f} = [\bar{V},\bar{W}]^a\, \partial_a \bar{f} = \bar{X}^a\, \partial_a \bar{f}$$
$$= \hat{X}^A\, \partial_A \hat{f},$$

whence $\hat{\mathbf{P}}$ is closed under the operation of taking Lie brackets. Thus, $\hat{\mathbf{P}}$ is a polarization on $\hat{\Gamma}$. □

Next, we consider the converse.

Lemma 1.2: Every polarization $\hat{\mathbf{P}}$ on the reduced phase-space $\hat{\Gamma}$ yields a constrained polarization $\bar{\mathbf{P}}$ upon pullback to $\bar{\Gamma}$.

Proof: Fix a point $\bar{\gamma}$ of $\bar{\Gamma}$ and set $\hat{\gamma} = \Pi \circ \bar{\gamma}$ as before. Let $\bar{\mathbf{P}}(\bar{\gamma}) := \underset{\leftarrow}{\Pi} \circ (\hat{\mathbf{P}}(\hat{\gamma}))$ be the maximal subspace of the tangent space of $\bar{\gamma}$ (within $\bar{\Gamma}$) which projects down to $\hat{\mathbf{P}}(\hat{\gamma})$. Since $\hat{\mathbf{P}}(\hat{\gamma})$ is $(n-m)$ dimensional and since the kernel of the mapping $\underset{\rightarrow}{\Pi}$ is m dimensional, $\bar{\mathbf{P}}(\bar{\gamma})$ is an n-dimensional subspace of the tangent space (within $\bar{\Gamma}$) at $\bar{\gamma}$. Furthermore, since the m-dimensional gauge flat at $\bar{\gamma}$ is the kernel of $\underset{\rightarrow}{\Pi}$, $\bar{\mathbf{P}}(\bar{\gamma})$ contains this m-flat. Thus, using $\hat{\mathbf{P}}$ on $\hat{\Gamma}$, we have obtained an n-dimensional distribution $\bar{\mathbf{P}}$ on $\bar{\Gamma}$ which contains the m-dimensional gauge distribution.

We have to show that the pullback of Ω_{ab} to $\bar{\mathbf{P}}$ vanishes and that $\bar{\mathbf{P}}$ is integrable. Given any $\hat{V}^A$ and $\hat{W}^B$ in $\mathbf{P}$, we have

$$0 = \hat{\Omega}_{AB}\, \hat{V}^A\, \hat{W}^B = \Omega_{ab}\, \bar{V}^a\, \bar{W}^b,$$

where $\bar{V}$ and $\bar{W}$ are *any* vectors on $\bar{\Gamma}$ such that $\underset{\rightarrow}{\Pi} \circ \bar{V} = \hat{V}$ and $\underset{\rightarrow}{\Pi} \circ \bar{W} = \hat{W}$. Hence, it follows that $\Omega_{ab}\, \bar{V}^a\, \bar{W}^b = 0$ for any $\bar{V}$ and $\bar{W}$ in $\bar{\mathbf{P}}$, whence the pullback of Ω_{ab} to $\bar{\mathbf{P}}$ vanishes. To show integrability of the n-flats $\bar{P}(\bar{\gamma})$, we first note that for any function $\hat{f}$ on $\hat{\Gamma}$,

$$\mathscr{L}_{\bar{V}}\, \mathscr{L}_{\bar{W}}\, \underset{\leftarrow}{\Pi} \circ \hat{f} = \mathscr{L}_{\bar{V}} \circ \underset{\leftarrow}{\Pi}\, \mathscr{L}_{\hat{W}} \hat{f} = \underset{\leftarrow}{\Pi} \circ \mathscr{L}_{\hat{V}}\, \mathscr{L}_{\hat{W}} \hat{f},$$

where $\underset{\leftarrow}{\Pi}$ is, as before, the pullback operation from $\hat{\Gamma}$ to $\bar{\Gamma}$, and where $\hat{W}^A = \underset{\rightarrow}{\Pi} \circ \bar{W}^a$ and $\hat{V}^A = \underset{\rightarrow}{\Pi} \circ \bar{V}^a$. Hence, if $\hat{V}$ and $\hat{W}$ are everywhere tangential to $\hat{\mathbf{P}}$, we have

$$\underset{\rightarrow}{\Pi} \circ [\bar{V},\bar{W}] = [\hat{V},\hat{W}] \in \hat{\mathbf{P}},$$

whence $[\bar{V},\bar{W}] \in \bar{\mathbf{P}}$ for all $\bar{V}$ and $\bar{W}$ tangential to $\bar{\mathbf{P}}$. Thus, $\bar{\mathbf{P}}$ is indeed a constrained polarization. □

Thus, we have shown that there is a 1–1 correspondence between constrained polarization $\bar{\mathbf{P}}$ on $\bar{\Gamma}$ and polarization $\hat{\mathbf{P}}$ on the reduced phase space $\hat{\Gamma}$. We now show that the quantum theory obtained using a constrained polarization $\bar{\mathbf{P}}$ in the procedure outlined in this subsection [Eqs. (29)–(32)] is naturally isomorphic to the quantum theory obtained from the corresponding polarization $\hat{\mathbf{P}}$ on the unconstrained reduced phase-space in the procedure of Sec. I A [Eqs. (5)–10)].

Let us begin with the "configuration spaces" $\bar{\mathscr{C}}$ and $\hat{\mathscr{C}}$. Each point of $\bar{\mathscr{C}}$ represents an n-dimensional, constrained-polarization leaf in $\bar{\Gamma}$, everywhere tangential to $\bar{\mathbf{P}}$. A point on $\hat{\mathscr{C}}$, on the other hand, represents an $(n-m)$-dimensional leaf in $\hat{\Gamma}$, everywhere tangential to $\hat{\mathbf{P}}$. However, the relation between $\bar{\mathbf{P}}$ and $\hat{\mathbf{P}}$ is such that there is a 1–1 correspondence between the leaves of the two polarizations: A leaf of $\hat{\mathbf{P}}$ is precisely the projection of a leaf of $\bar{\mathbf{P}}$ via the natural projection map Π from $\bar{\Gamma}$ to $\hat{\Gamma}$. Hence $\bar{\mathscr{C}}$ and $\hat{\mathscr{C}}$ are naturally isomorphic. Hence, there is a natural isomorphism Λ from the quantum Hilbert space $\bar{\mathscr{H}}$ obtained from the constrained polarization $\bar{\mathbf{P}}$ to the quantum Hilbert space $\hat{\mathbf{H}}$ obtained from the polarization $\hat{\mathbf{P}}$. Next, we consider the *basic* quantum operators in the two schemes. Let $\bar{f}$ be a function on $\bar{\Gamma}$ such that (i) X_f^a is tangential to $\bar{\Gamma}$; and (ii) X_f^a preserves the constrained polarization $\mathscr{L}_{X_f}\, \bar{\mathbf{P}} \subseteq \bar{\mathbf{P}}$, where f is any extension to Γ of $\bar{f}$. (Note that, if one extension f has these properties, so do all extensions.) Then, in the constrained polarization scheme, $\bar{f}$ is a *basic* observable associated with $\bar{\mathbf{P}}$; $\bar{f}$ has an unambiguous quantum analog $\bar{\mathbf{F}}$. Now, condition (i) is equivalent to demanding

$$0 \cong X_f^a \partial_b C_i \cong \Omega_{ab} X_f^a X_i^b \cong -X_i^b \partial_b \bar{f}, \quad \text{for all } i \equiv 1,\ldots, m.$$

Where, as before $\cong$ stands for "equals, at points of $\bar{\Gamma}$, to"; and X_i^a is the Hamiltonian vector field generated by the ith constraint function C_i. Thus, (i) is equivalent to demanding that $\bar{f}$ is the pullback to $\bar{\Gamma}$ of a function $\hat{f}$ on $\hat{\Gamma}$; $\bar{f} = \underset{\leftarrow}{\Pi} \circ \hat{f}$. Condition (ii) on X_f^a is now equivalent to demanding that the Hamiltonian vector field $\hat{X}_{\hat{f}}^A$ on $(\hat{\Gamma},\hat{\Omega})$ of $\hat{f}$ should preserve the polarization $\hat{\mathbf{P}}$; $\mathscr{L}_{\hat{X}_{\hat{f}}} \hat{\mathbf{P}} \subseteq \hat{\mathbf{P}}$. But this is precisely the necessary and sufficient condition for $\hat{f}$ to be a *basic* classical observable on $\hat{\Gamma}$, associated with $\hat{\mathbf{P}}$; $\hat{f}$ has an unambiguous quantum analog $\hat{\mathbf{F}}$. A direct comparison between the explicit expressions of $\bar{\mathbf{F}}$ and $\hat{\mathbf{F}}$ yields

$$\Lambda \circ \bar{\mathbf{F}} \circ \Lambda^{-1} = \hat{\mathbf{F}}. \tag{33}$$

Thus, we have the following theorem.

Theorem 1: The quantum theory obtained from a constrained polarization $\bar{\mathbf{P}}$ on the constraint surface $\bar{\Gamma}$ is naturally isomorphic with that obtained from the corresponding polarization $\hat{\mathbf{P}}$ on the reduced phase space $\hat{\Gamma}$.

C. Example revisited: Constrained polarization

Let us now return to the example of Sec. II A. The configuration space, $\mathscr{C}$, is the four-manifold consisting of points in the interior of the null cone of an origin 0 in Minkowski space. The phase space Γ is the cotangent bundle over $\mathscr{C}$. Thus, Γ is an eight-dimensional manifold. The six-dimensional constraint surface $\bar{\Gamma}$ is specified by the two constraints

$$C_1(q,p) \equiv \eta^{\alpha\beta} p_\alpha p_\beta = 0, \tag{19}$$

$$C_2(q,p) \equiv v^\alpha p_\alpha = 0, \tag{20}$$

where v^α is the conformal killing field of $\eta^{\alpha\beta}$ given by

$$v^\alpha(q) := (k \circ q)\, q^\alpha - \tfrac{1}{2}(q \circ q) k^\alpha. \tag{22}$$

(Here $\eta^{\alpha\beta}$ is the Minkowskian metric; k^α, a constant, unit, spacelike vector field; and q^α, the position vector of a point in $\mathscr{C}$ w.r.t. the origin 0.) The two constraints are of first class.

To obtain a quantum theory via the procedure of Sec. II B, we have to introduce a constrained polarization $\bar{\mathbf{P}}$ on $\bar{\Gamma}$, i.e., enlarge the two-dimensional gauge flat in the tangent space of each point $\bar{\gamma}$ of $\bar{\Gamma}$ to a four-dimensional, Lagrangian, integrable flat. For this, it is convenient to introduce a new chart on $\mathscr{C}$. Introduce four orthonormal, constant vector fields $\mathring{e}_\beta{}^\alpha$ on $(\mathscr{C},\eta)$, $\beta = 0,1,2,3$, such that $\mathring{e}_1{}^\alpha = k^\alpha$ and consider the corresponding four conformal Killing vector fields:

$$e_\beta{}^\alpha := (\mathring{e}_\beta \circ q)\, q^\alpha - \tfrac{1}{2}(q \circ q)\mathring{e}_\beta{}^\alpha. \tag{34}$$

These vector fields leave the region $\mathscr{C}$ of Minkowski space invariant, are mutually orthogonal, and commute everywhere on $\mathscr{C}$. [e_1^α is, of course, the vector field v^α of Eq. (22).] Hence, there exists a chart x^β on $\mathscr{C}$ such that $\partial/\partial x^\beta = e_\beta$. We shall make the following specific choice: $x^\alpha = -2\, q^\alpha / q \circ q$, where, as before, $q \circ q = \eta_{\alpha\beta} q^\alpha q^\beta$. Let (x^α, p_α) be the corresponding chart on Γ. Then, Eqs. (19) and (20) defining $\bar{\Gamma}$ are equivalent to

$$C_1'(x,p) \equiv -p_0^2 + p_1^2 + p_2^2 + p_3^2 \equiv \eta^{\alpha\beta} p_\alpha p_\beta = 0, \tag{19'}$$

$$C_2'(x,p) \equiv p_1 = 0, \tag{20'}$$

and the corresponding Hamiltonian vector fields are

$$X_1 = \eta^{\alpha\beta} p_\beta \frac{\partial}{\partial x^\alpha}, \tag{35}$$

$$X_2 = \frac{\partial}{\partial x^1}. \tag{36}$$

An obvious choice for the constrained polarization is therefore[16]:

$$\bar{\mathbf{P}}|_{(x,p)} \equiv \text{linear span of} \left\{ \frac{\partial}{\partial x^1}, \eta^{\alpha\beta} p_\beta \frac{\partial}{\partial x^\alpha}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right\}\Bigg|_{(x,p)}$$

$$\equiv \text{linear span of} \left\{ \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}, \frac{\partial}{\partial x^0} \right\}\Bigg|_{(x,p)}. \tag{37}$$

It is obvious that $\bar{\mathbf{P}}(x,p)$ contains the two-dimensional gauge flat at (x,p); that the pullback of the symplectic structure $\Omega \equiv dp_\alpha \wedge dx^\alpha$ on Γ to $\bar{\mathbf{P}}$ is zero, and that the four-flats $\bar{\mathbf{P}}(x,p)$ are integrable. The effective configuration space $\bar{\mathscr{C}}$ is the quotient of $\bar{\Gamma}$ by the integral manifolds of $\bar{\mathbf{P}}$. Thus, each point $\bar{q}$ of $\bar{\mathscr{C}}$ is an equivalence class of points (x^α, p_α) satisfying (19') and (20'), subject to the equivalence relation

$$(x^\alpha, p_\alpha) \approx (x'^\alpha p_\alpha).$$

Hence, $\bar{\mathscr{C}}$ can be coordinated by three-variables p_0, p_2, p_3, subject only to the condition $-p_0^2 + p_1^2 + p_2^2 = 0$. Thus, $\bar{\mathscr{C}}$ has the structure of the light cone in a *three-dimensional* Minkowski space; as expected, $\bar{\mathscr{C}}$ is two-dimensional.[16] Quantum states are complex-valued functions on $\bar{\mathscr{C}}$ that are square-integrable w.r.t. the volume element given by an arbitrarily chosen but fixed nowhere vanishing two-form on $\bar{\mathscr{C}}$. The *basic* classical observables—i.e., the ones with unambiguous quantum analogs—consist of functions on $\bar{\Gamma}$ that are either independent of x^α or linear in x^α. Thus, using the constrained polarization framework developed in Sec. II B, we have obtained a coherent description of quantum kinematics of the system.

Remarks: (i) The constrained polarization $\bar{\mathbf{P}}$ was constructed using the structure of the conformal killing fields of Minkowski space. The vector field, $v^\alpha \equiv e_1^\alpha$, which enters the constraint (22), is an "inverted translation" or "acceleration" in the one-direction. We augmented the gauge two-flat at each point of $\bar{\Gamma}$ by Hamiltonian vector fields on $\bar{\Gamma}$ defined by the "inverted translations" e_2^α and e_3^α along the two- and three-directions. Since e_1^α, e_2^α, and e_3^α all commute with one another, the $e_\beta^\alpha p_\alpha$ form a complete set of commuting observables on Γ. The quantum wave functions are restrictions to $\bar{\Gamma}$ of (normalizable) complex-valued functions of these variables.

(ii) Quantum wave functions, the polarized solutions to the prequantum operator constraints, can be identified with distributions of the type $\delta(-(e_0 \circ p)^2 + (e_2 \circ p)^2 + (e_3 \circ p)^2)\, \psi(e_0 \circ p, e_2 \circ p, e_3 \circ p)$. Note, however, that since the vector fields e_α are *not* constant (w.r.t the derivative operator of $\eta^{\alpha\beta}$) in the Cartesian chart on $\mathscr{C}$, the wave functions appear as rather complicated functions of the coordinates q^α

and the corresponding components of momenta. Thus, the constrained polarization is neither vertical nor the usual horizontal one (i.e., the one given by $\partial/\partial q^\alpha$); in the standard terminology, the representation given by $\overline{\mathbf{P}}$ is neither the configuration representation nor the momentum one. This is to be expected because neither the vertical polarization nor the usual horizontal one is compatible with the prequantum operator constraints.

D. Prequantum operator constraint method

In Sec. II A, we showed, by means of an example, that a general polarization $\mathbf{P}$ on Γ need not be compatible with prequantum operator constraints. We shall now discuss the necessary and sufficient conditions on $\mathbf{P}$ needed to ensure compatibility, then show that these conditions suffice for $\mathbf{P}$ to induce a constrained polarization $\overline{\mathbf{P}}$ on $\overline{\Gamma}$, and conclude by comparing the quantum theories resulting from $\mathbf{P}$ and the corresponding $\overline{\mathbf{P}}$.

As before, we assume that Γ is $2n$ dimensional and that there are m first class constraints $C_i(\gamma)=0$, $i=1,2,\ldots,m$. In the passage to quantum theory, we begin by considering complex-valued, U(1)-Higgs scalars ψ on Γ satisfying the prequantum operator constraints:

$$\mathbf{O}_{\mathbf{K}}\circ\psi\equiv(\hbar/i)\,X^a_{\mathbf{K}}\,\nabla_a\,\psi+C_{\mathbf{K}}\,\psi=0. \tag{18}$$

Our next task is to introduce a polarization $\mathbf{P}$ on (Γ,Ω) and consider only those solutions ψ to (18) that also satisfy

$$\mathscr{L}_V\,\psi\equiv V^a\,\nabla_a\,\psi=0, \tag{5}$$

for all vector fields V^a tangential to $\mathbf{P}$. We know that the m conditions (18) are internally consistent and that so are the n conditions (5). Now the question is that of compatibility *between* (18) and (5). Given *any* U(1) Higgs scalar ψ, we have

$$\begin{aligned}
&\mathscr{L}_V\circ\mathbf{O}_{\mathbf{K}}\circ\psi-\mathbf{O}_{\mathbf{K}}\circ\mathscr{L}_V\,\psi\\
&\quad=V^a\nabla_a\,((\hbar/i)\,X^b_{\mathbf{K}}\,\nabla_b+C_{\mathbf{K}})\psi\\
&\qquad-((\hbar/i)\,X^b_{\mathbf{K}}\,\nabla_b+C_{\mathbf{K}})\,V^a\nabla_a\,\psi\\
&\quad=(\hbar/i)(\mathscr{L}_V\,X^b_{\mathbf{K}})\,\nabla_b\,\psi\\
&\qquad+(2\hbar/i)\,V^a X^b_{\mathbf{K}}\,\nabla_{[a}\,\nabla_{b]}\,\psi\\
&\qquad+(V^a\nabla_a\,C_{\mathbf{K}})\,\psi\\
&\quad=(\hbar/i)(\mathscr{L}_V\,X^b{}_{\mathbf{K}})\,\nabla_b\,\psi.
\end{aligned} \tag{38}$$

Hence, (18) and (5) are mutually consistent iff the right side of Eq. (38) vanishes identically (on all of Γ) if ψ itself satisfies the two sets of equations. This, in turn, is possible iff

$$\mathscr{L}_{X_{\mathbf{K}}}\,V^a\in\mathbf{P}, \tag{39}$$

$\forall\,\mathbf{K}=1,\ldots,m$ and $\forall\,V^a\in\mathbf{P}$. Thus, the consistency condition between (18) and (5) is precisely that the polarization $\mathbf{P}$ be preserved by the m constraint vector fields. Such a $\mathbf{P}$ will be referred to as a *polarization compatible with constraints*. Given such a polarization, (18) and (5) give us a consistent set of $(n+m)$ conditions on ψ. Since, to begin with, ψ effectively depends on $2n$ variables, solutions to (18) and (5) effectively depend only on $(n-m)$ variables and therefore can be thought of as the physical quantum states of the system. Note, however, that *a priori* there is no obvious Hermitian inner product on the space of these wave functions. We shall return to this point later in this subsection.

Next, we wish to show that any polarization that is compatible with constraints naturally induces a constrained polarization $\overline{\mathbf{P}}$ on $\overline{\Gamma}$. Let us begin with a general polarization $\mathbf{P}$ not necessarily compatible with constraints. Fix *any* point $\bar\gamma$ of $\overline{\Gamma}$ and consider the subspace $\overline{\mathbf{P}}(\bar\gamma)$ in the tangent space $\overline{T}_{\bar\gamma}$ of $\bar\gamma$ within $\overline{\Gamma}$ defined by

$$\overline{\mathbf{P}}(\bar\gamma):=\text{linear span of }\{\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma},G(\bar\gamma)\}, \tag{40}$$

where $G(\bar\gamma)$ is the m-dimensional gauge flat in $\overline{T}_{\bar\gamma}$ spanned by the constraint vector fields. Since

$$\Omega_{ab}\,\bar g^a\,\bar u^b=0\quad\text{and}\quad\Omega_{ab}\,V^a\,W^b=0, \tag{41}$$

for all $\bar g^a\in G(\bar\gamma)$, $\bar u^b\in\overline{T}_{\bar\gamma}$, and V^a, $W^a\in\mathbf{P}(\bar\gamma)$, it follows that the pullback of Ω_{ab} to $\overline{\mathbf{P}}(\bar\gamma)$ vanishes identically; $\overline{\mathbf{P}}(\bar\gamma)$ is isotropic w.r.t. Ω_{ab}. Next, using the fact that Ω_{ab} is nondegenerate, the constraints are first class, and that $\mathbf{P}(\bar\gamma)$ is Lagrangian, we can show that $\overline{\mathbf{P}}(\bar\gamma)$ is n-dimensional everywhere on $\overline{\Gamma}$.

Lemma 2.1:

$$\dim\overline{\mathbf{P}}(\bar\gamma)=n,\quad\text{for all}\quad\bar\gamma\text{ in }\overline{\Gamma}.$$

Proof: Let $V:=[\mathbf{P}(\bar\gamma)\cap G(\bar\gamma)]$ and let V' be the subspace of the tangent space of $\bar\gamma$ in Γ that is isotropic w.r.t. V, i.e., is such that $Y^b\in V'$ iff $\Omega_{ab}\,X^a\,Y^b=0$, $\forall\,X^a\in V$. Let the dimension of V be s. Then, since Ω_{ab} is a nondegenerate two-form, $\dim V'=2n-s$. Now, by inspection, the linear span of $\mathbf{P}(\bar\gamma)$ and $\overline{T}_{\bar\gamma}$ is a subspace of V'. Let us compute the dimension of this span:

$$\dim(\text{linear span of }\mathbf{P}(\bar\gamma)\cup\overline{T}_{\bar\gamma})=3n-m-k,$$

where $k=\dim(\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma})$. Hence, we have a bound on the dimension of V':

$$\dim V'=2n-s\geqslant 3n-m-k,$$

whence it follows that $-s\geqslant n-m-k$.

Now, $\dim\overline{\mathbf{P}}(\bar\gamma)=k+m-s$. Hence,

$$\dim\overline{\mathbf{P}}(\bar\gamma)\geqslant k+m+n-m-k\geqslant n.$$

However, we know that $\overline{\mathbf{P}}(\bar\gamma)$ is isotropic w.r.t. Ω_{ab}. Hence, it follows that $\dim\overline{\mathbf{P}}(\bar\gamma)\leqslant n$. Combining the two bounds we have

$$\dim\overline{\mathbf{P}}(\bar\gamma)=n.$$ □

Thus, given *any* polarization $\mathbf{P}$ on Γ, one can consider the intersection of the n-dimensional polarization flat and the $(2n-m)$-dimensional tangent space $\overline{T}_{\bar\gamma}$ within $\overline{\Gamma}$ of $\bar\gamma$. In general, although one has the bound $n-m\leqslant\dim\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma}\leqslant n$, the precise value of the dimension of $\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma}$ is not universal; it depends on the choice of $\mathbf{P}$. However, irrespective of this choice, the dimension of the space obtained by augmenting the intersection $\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma}$ by the gauge directions not already contained in it is *always* n. If $\mathbf{P}(\bar\gamma)$ contains no gauge directions, then $\dim\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma}$ is $n-m$, while if it contains all gauge directions, $\dim\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma}$ is n; the more the gauge directions in $\mathbf{P}(\bar\gamma)$, the less is the number of directions in $\mathbf{P}(\bar\gamma)$ transverse to $\overline{\Gamma}$.

Thus, by augmenting $\mathbf{P}(\bar\gamma)\cap\overline{T}_{\bar\gamma}$ by gauge directions, one obtains, at each point $\bar{\gamma}$ of $\bar{\Gamma}$, a Lagrangian subspace (w.r.t. Ω_{ab}) $\bar{\mathbf{P}}(\bar{\gamma})$ of $\bar{T}_{\bar{\gamma}}$. By construction, the gauge flat $G(\bar{\gamma})$ is contained in polarization $\bar{\mathbf{P}}(\bar{\gamma})$. Hence, $\bar{\mathbf{P}}$ already enjoys all the properties of a constrained polarization $\bar{\mathbf{P}}$ except integrability. It is here that we need $\mathbf{P}$ to be compatible with constraints.

Lemma 2.2: $\bar{\mathbf{P}}(\bar{\gamma})$ are integrable if $\mathbf{P}$ is compatible with constraints.

Proof: By definition of $\bar{\mathbf{P}}$, any vector field $\bar{V}^a$ that is tangential to $\bar{\mathbf{P}}$ everywhere on $\bar{\Gamma}$ is a sum of a vector field $\bar{Y}^a$ everywhere tangential to $\mathbf{P}(\bar{\gamma}) \cap \bar{T}_{\bar{\gamma}}$ and a constraint vector field $\bar{X}^a$: $\bar{V}^a = \bar{Y}^a + \bar{X}^a$. Hence,

$$[\bar{V}, \bar{V}'] = [\bar{Y}, \bar{Y}'] + [\bar{X}, \bar{X}'] + [\bar{Y}, \bar{X}'] + [\bar{X}, \bar{Y}]$$
$$= \bar{Y}'' + \bar{X}'' + [\bar{Y}, \bar{X}'] + [\bar{X}, \bar{Y}'],$$

where $\bar{Y}''$ is a vector field tangential to $\mathbf{P}(\bar{\gamma}) \cap \bar{T}_{\bar{\gamma}}$ and $\bar{X}''$ is a constraint vector field. Since $\mathbf{P}$ is constraint compatible, $[\bar{Y}, \bar{X}']$ and $[\bar{X}, \bar{Y}']$ are tangential to $\mathbf{P}$. Next, since $\bar{X}, \bar{Y}, \bar{X}', \bar{Y}'$ are all tangential to the submanifold $\bar{\Gamma}$ of Γ, $[\bar{Y}, \bar{X}']$ and $[\bar{X}, \bar{Y}']$ are also tangential to $\bar{\Gamma}$. Hence, each of the four vector fields in the final expression of $[\bar{V}, \bar{V}']$ lies in $\bar{\mathbf{P}}$, whence $\bar{\mathbf{P}}$ is integrable. □

Combining the results of Lemma 2.1 and 2.2, we have the following theorem.

Theorem 2: Let $\mathbf{P}$ be a polarization on Γ, compatible with constraints. Then $\bar{\mathbf{P}}$, defined by $\bar{\mathbf{P}}(\bar{\gamma})$: = linear span of $\{\mathbf{P}(\bar{\gamma}) \cap \bar{T}_{\bar{\gamma}}, G(\bar{\gamma})\}$ for all $\bar{\gamma}$ in $\bar{\Gamma}$, is a constrained polarization on $\bar{\Gamma}$.

Finally, let us compare the space of quantum wave functions obtained from $\mathbf{P}$ via Eqs. (18) and (5) with the one obtained from $\bar{\mathbf{P}}$ via the constrained-polarization condition, Eq. (29).

Lemma 3.1: Let $\mathbf{P}$ be a polarization on Γ, compatible with constraints. Then, the restriction $\underline{\psi}$ to $\bar{\Gamma}$ of every solution ψ to Eqs. (18) and (5) satisfies the constrained-polarization equation (29) w.r.t. $\bar{\mathbf{P}}$.

Proof: On $\bar{\Gamma}$, (18) reduces to

$$X^a{}_{\mathbf{K}} \nabla_a \psi \cong 0, \quad \mathbf{K} = 1,2,..., m,$$

and (5) implies that

$$\bar{V}^a \nabla_a \psi \cong 0,$$

for all vector fields $\bar{V}^a$ tangential to $\mathbf{P}(\bar{\gamma}) \cap \bar{T}_{\bar{\gamma}}$ for all $\bar{\gamma}$ in $\bar{\Gamma}$. Since $\bar{\mathbf{P}}(\bar{\gamma})$: = linear span $\{G(\bar{\gamma}), \mathbf{P}(\bar{\gamma}) \cap \bar{T}_{\bar{\gamma}}\}$, it follows immediately that the restriction $\underline{\psi}$ of ψ to $\bar{\Gamma}$ satisfies (29):

$$\bar{W}^a \nabla_a \underline{\psi} \cong 0,$$

for all $\bar{W}$ tangential to $\bar{\mathbf{P}}$. □

To analyze the relation between quantum theories resulting from $\mathbf{P}$ and $\bar{\mathbf{P}}$, we shall *make an additional assumption* on the polarization $\mathbf{P}$. We now assume that $G(\bar{\gamma}) \cap \mathbf{P}(\gamma)$ is spanned by the s constraint vector fields $X_1,..., X_s$ in a neighborhood of $\bar{\Gamma}$. Then, we can show that the solutions ψ to Eqs. (18) and (5) are completely determined in a neighborhood of $\bar{\Gamma}$ by their restrictions $\underline{\psi}$ to $\bar{\Gamma}$. If the neighborhood extends to all of Γ, the quantum description given by $\mathbf{P}$ is equivalent to that given by the constrained polarization $\bar{\mathbf{P}}$ induced by $\mathbf{P}$ on $\bar{\Gamma}$.

Lemma 3.2: There exists a neighborhood N of $\bar{\Gamma}$ in Γ such that the mapping Δ_N from the space of solutions ψ to (18) and (5) in N to the space of solutions $\underline{\psi}$ to (29) on $\bar{\Gamma}$ (obtained by restricting ψ to $\bar{\Gamma}$) is one to one and onto.

Proof: In a neighborhood N_1 of $\bar{\Gamma}$ within Γ, solutions ψ to (18) and (5) must satisfy $C_1 \psi = 0,..., C_s \psi = 0$. Denote by $\tilde{\Gamma}$ the $(2n - s)$-dimensional submanifold of N_1 defined by

$$C_1(\gamma) = 0,..., C_s(\gamma) = 0.$$

Since solutions to (18) and (5) in N_1 vanish outside $\tilde{\Gamma}$, we now focus on $\tilde{\Gamma}$. Given any vector field V^a in N_1, we have

$$V^a \nabla_a C_\mathrm{i} = \Omega_{ab} V^a X^b_\mathrm{i}.$$

Therefore, since the pullback of Ω_{ab} to the polarization n-flats vanishes and since $X_1,..., X_s$ are tangential to $\mathbf{P}$, we have

$$V^a \nabla_a C_1 = 0,..., V^a \nabla_a C_s = 0,$$

for all V^a in P. Thus, the n-dimensional polarization leaf through any point of Γ lies *entirely* within $\tilde{\Gamma}$.

Let us now consider leaves of $\mathbf{P}$ passing through points $\bar{\gamma}$ of $\bar{\Gamma}$. Since, by Lemma 2.1, the linear span of $\{\mathbf{P}(\bar{\gamma}) \cap \bar{T}_{\bar{\gamma}}, G(\bar{\gamma})\}$ is n dimensional and since, by assumption, $\mathbf{P}(\bar{\gamma}) \cap G(\bar{\gamma})$ is s dimensional at any $\bar{\gamma}$ in $\bar{\Gamma}$, it follows that $\mathbf{P}(\bar{\gamma}) \cap \bar{T}_{\bar{\gamma}}$ is $(n - m + s)$ dimensional. Since $\mathbf{P}(\bar{\gamma})$ is n dimensional, there exists an $(m - s)$-dimensional subspace in the tangent space (within $\tilde{\Gamma}$) of each $\bar{\gamma}$ in $\bar{\Gamma}$ that lies entirely in $\mathbf{P}(\bar{\gamma})$ and which is transverse to $\bar{\Gamma}$. Note, furthermore, that the codimension of $\bar{\Gamma}$ in $\tilde{\Gamma}$ is precisely $(m - s)$. Therefore, leaves of $\mathbf{P}$, passing through points $\bar{\gamma}$ of $\bar{\Gamma}$, provide us an open neighborhood $\tilde{N}$ of $\bar{\Gamma}$ within $\tilde{\Gamma}$. Let N be an open neighborhood of $\bar{\Gamma}$ within Γ whose intersection with $\tilde{\Gamma}$ is $\tilde{N}$.

Given a solution ψ to (18) and (5), in N, $\underline{\psi}$: $= \Delta_n \psi$, the restriction of ψ to $\bar{\Gamma}$, is a solution to (29). Let ψ be such that $\underline{\psi} = 0$. Then, since ψ is constant on each leaf of polarization [Eq. (5)], it follows that ψ must vanish in $\tilde{N}$. We have already shown that (18) and (5) imply that ψ must vanish in $\Gamma - \tilde{\Gamma}$. Hence, ψ vanishes everywhere in N. Thus, the mapping Δ_N is 1–1. Next, we show that Δ_N is onto. Given any solution $\underline{\psi}$ to (29) on $\bar{\Gamma}$, we extend it to ψ on $\tilde{N}$ by requiring that ψ be (covariantly) constant on each leaf of P. The extension is unambiguous because, in virtue of (29), $\underline{\psi}$ is covariantly constant on the intersection of $\bar{\Gamma}$ with any leaf of $\mathbf{P}$. Finally, we extend ψ to N by requiring that it vanish in the complement of $\tilde{N}$ in N. By construction, the resulting ψ satisfies (5) and the first s of equations (18). The satisfaction of the last $(m - s)$ of Eqs. (18) follows from the fact that these equations are satisfied on $\bar{\Gamma}$ in virtue of (29) and the fact that the polarization $\mathbf{P}$ is compatible with constraints. Thus, every $\underline{\psi}$ on $\bar{\Gamma}$ satisfying (29) is the image under Δ_N of *some* ψ satisfying (18) and (5) in N. That is, Δ_N is also onto. □

Combining the results of Lemmas 3.1 and 3.2, we have the following theorem.

Theorem 3: Let $\mathbf{P}$ be a polarization on Γ that is compatible with constraints and that contains precisely s gauge directions $X_1, X_2,...,X_s$ at each point in a neighborhood of $\bar{\Gamma}$ within Γ $(0 \leqslant s \leqslant m)$. Then, there exists a neighborhood N of $\bar{\Gamma}$ within Γ such that the space of solutions ψ to the prequantum operator constraint equations (18) and the $\mathbf{P}$-polarization conditions (5) in N is naturally isomorphic to the space of solutions to the $\bar{\mathbf{P}}$-constrained polarization conditions (29) on $\bar{\Gamma}$, where $\bar{\mathbf{P}}$ is obtained from $\mathbf{P}$ via Theorem 2.

Let us now illustrate the ideas underlying this theorem by means of a simple example. Consider a nonrelativistic particle. The configuration space $\mathscr{C}$ is the Euclidean three-space and the phase space Γ is the cotangent bundle over $\mathscr{C}$. Let the constraint be

$$C(\vec{x},\vec{p}) := p_x = 0.$$

Then $\overline{\Gamma}$ is a five-manifold with a global chart (x, y, z, p_y, p_z) and the constraint vector field $X_C := \partial/\partial x$ is tangential to $\overline{\Gamma}$. Let us now introduce a polarization $\mathbf{P}_1$ on Γ:

$$\mathbf{P}_1(\vec{x},\vec{p}) := \text{span of}\left\{\frac{\partial}{\partial p_x}, \frac{\partial}{\partial p_y}, \frac{\partial}{\partial p_z}\right\}.$$

Then, $X_C \equiv \partial/\partial x$ preserves $\mathbf{P}_1$; the polarization is compatible with constraints. Since X_C is everywhere transverse to $\mathbf{P}_1$, the intersection $G(\bar{\gamma}) \cap \mathbf{P}_1$ is zero dimensional, whence $\overline{\mathbf{P}}_1(\bar{\gamma})$ is obtained by augmenting $\overline{T}_{\bar{\gamma}} \cap \mathbf{P}_1(\bar{\gamma})$ by the constraint vector $X_C|_{\bar{\gamma}}$,

$$\overline{\mathbf{P}}_1 := \text{span of}\left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial p_y}, \frac{\partial}{\partial p_z}\right\}.$$

Let us now use the constrained polarization method. It is convenient to use a "gauge" adapted to $\mathbf{P}$, e.g., the one in which the symplectic potential A_a is given by $A = -x\, dp_x + p_y\, dy + p_z\, dz$. Then, the $\overline{\mathbf{P}}$-polarization conditions [Eq. (29)] imply that the permissible wave functions $\underline{\psi}$ have the form

$$\underline{\psi} \equiv \underline{\psi}(y, z).$$

Let us now compare this result with the one obtained from the prequantum operator constraint method using the polarization $\mathbf{P}$ on Γ. Now, it is convenient to use the "gauge" adapted to $\mathbf{P}$, i.e., to choose $A = p_x\, dx + p_y\, dy + p_z\, dz$. Then, the prequantum operator constraint (18) yields

$$\frac{\hbar}{i}\frac{\partial}{\partial x}\psi(\vec{x},\vec{p}) - p_x\,\psi + p_x\,\psi = 0$$
$$\Leftrightarrow \psi \equiv \Psi(y, z, p_x, p_y, p_z),$$

while the $\mathbf{P}_1$-polarization conditions (5) yield

$$\frac{\partial}{\partial p_\alpha}\psi(\vec{x},\vec{p}) = 0 \Leftrightarrow \psi \equiv \psi(\vec{x}), \quad \alpha = 1,2,3$$

The two conditions together now imply that $\psi \equiv \psi(y, z)$. Thus, the space of solutions to (18) and (5) on Γ is naturally isomorphic to the space of solutions to (29). In this example, $s = \dim G(\gamma) \cap \mathbf{P}_1(\gamma) = 0$. Hence $\widetilde{\Gamma}$ is $2n \equiv$ six dimensional. (In fact, $\Gamma = \widetilde{\Gamma}$.) The constraint surface $\overline{\Gamma}$ is five dimensional. Every $\underline{\psi} \equiv \underline{\psi}(y, z)$, a solution to (29) on $\overline{\Gamma}$, is extended to all of Γ unambiguously by (5) because one can reach any point (x, y, z, p_x, p_y, p_z) in Γ by the integral curve of the vector field $\partial/\partial p_x$ (which is everywhere tangential to $\mathbf{P}_1$) that passes through the point $(x, y, z, p_x = 0, p_y, p_z)$ of $\overline{\Gamma}$.

Let us now use another polarization $\mathbf{P}_2$:

$$\mathbf{P}_2(\vec{x},\vec{p}) := \text{span}\left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right\}.$$

In this case, $s = \dim G(\gamma) \cap \mathbf{P}_2(\gamma) = 1 = m$, the number of constraints. In general, $0 \leqslant s \leqslant m$. With $\mathbf{P}_1$, we had $s = 0$, while with $\mathbf{P}_2$ we have $s = m$; the two cases correspond to the two extreme scenarios. Again $\mathbf{P}_2$ is compatible with the constraint and induces the constrained polarization

$$\overline{\mathbf{P}}_2 := \text{span}\left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right\}$$

on $\overline{\Gamma}$; since $s = m$, $\overline{\mathbf{P}}_2$ is just the restriction of $\mathbf{P}_2$ to $\overline{\Gamma}$. In the "gauge" adapted to $\overline{\mathbf{P}}$, $A = -x^\alpha\, dp_\alpha$, the solutions $\underline{\psi}$ to the constrained polarization equation (29) are

$$\underline{\psi} = \underline{\psi}(p_y, p_z).$$

Let us now apply the prequantum operator constraint method. Equation (5) now implies (using $A = -x^\alpha\, dp_\alpha$)

$$\frac{h}{i}\frac{\partial}{\partial x}\psi(\vec{x},\vec{p}) + p_x\,\psi(\vec{x},\vec{p}) = 0,$$

while the $\mathbf{P}_2$-polarization conditions (29) yield

$$\frac{\partial}{\partial x^\alpha}\psi(\vec{x},\vec{p}) = 0, \quad \alpha = 1,2,3.$$

Thus, the solutions to the two sets of equations have support *only* on the $p_x = 0$ surfaces, i.e., only on $\overline{\Gamma}$, and there they are independent of x^α. That is, the solutions are functions of p_y and p_z only; $\psi \equiv \psi(p_y, p_z)$, whence the spaces of solutions to (18) and (5) on Γ and of (29) on $\overline{\Gamma}$ are again naturally isomorphic. [Note that since $s = m$ now, $\widetilde{\Gamma}$ is $(2n - m)$ dimensional; in fact $\widetilde{\Gamma} = \overline{\Gamma}$.]

In general, $0 \leqslant s \leqslant m$ and one obtains a "combination" of the two scenarios given above.

Remarks: (i) In the prequantum operator constraint method, one needs not only the constraint surface $\overline{\Gamma}$, but also constraint functions C_i (up to linear combinations) on all of Γ. One would therefore have expected the resulting quantum theory to depend sensitively on the choice of C_i. However, from the proof of Lemma 3.2, it follows that, to a large extent, this is not the case. More precisely, we have the following. Choose two sets of constraint functions C_i and C_i' on Γ all of which vanish precisely on $\overline{\Gamma}$. Choose a polarization $\mathbf{P}$ that is compatible with both sets of constraints and that has the property that $\mathbf{P}(\gamma) \cap G(\gamma) = \mathbf{P}(\gamma) \cap G'(\gamma)$ [with $\dim P(\gamma) \cap G(\gamma) = s$] everywhere in a neighborhood of $\overline{\Gamma}$, where $G(\gamma)$ and $G'(\gamma)$ are, respectively, the m-flats spanned by the constraint vector fields X_i and X_i' at γ. Then, in the neighborhood N (which is determined only by the polarization $\mathbf{P}$), $\psi(\gamma)$ satisfies (18) w.r.t. C_i and (5) iff it satisfies (18) w.r.t. C_i' and (5), since ψ is completely determined in N by $\underline{\psi}$ and the polarization $\mathbf{P}$.

(ii) How large is the neighborhood N? Although, in simple examples, such as the one considered above, one can choose N to be all of Γ, it *is* possible to concoct cases in which it is a *proper* subset of Γ. In a general case, the maximal choice of $\widetilde{N}$ is the set of those points of $\widetilde{\Gamma}$ the polarization leaf through which intersects $\overline{\Gamma}$. Thus, the polarization leaves passing through the complement of this $\widetilde{N}$ never intersect $\overline{\Gamma}$. Consequently, if the complement has the same dimension, $(2n - s)$, as $\widetilde{\Gamma}$, there exist nontrivial wave functions ψ on Γ in the prequantum operator constraint method whose restriction $\underline{\psi}$ to $\overline{\Gamma}$ vanishes identically. Such wave functions have no analog whatsoever in the constrained polarization or the reduced phase-space approach. Consequently, in this case, the map Δ_Γ from the space of solutions ψ to (18) and (5) on Γ to the space of solutions $\underline{\psi}$ to (29) on $\overline{\Gamma}$ fails to be one to one.

(iii) Let us suppose that **P** is such that the polarization leaf through *any* point $\bar{\gamma}$ of $\bar{\Gamma}$ intersects $\bar{\Gamma}$ in precisely one $(n - m + s)$-dimensional submanifold. Then, Δ_Γ is an isomorphism. We can use this fact to pull back the inner product on $\bar{\mathbf{H}}$ to the space of solutions to (18) and (5) and to introduce the *basic* quantum operators on the resulting Hilbert space. Thus, in this case, using **P**, the program of the prequantum operator constraints method can be completed naturally and yields a quantum theory that is equivalent to that obtained from the constrained polarization $\bar{\mathbf{P}}$, induced on $\bar{\Gamma}$ by **P**. If there exist leaves of **P** that fail to intersect $\bar{\Gamma}$ or that intersect $\bar{\Gamma}$ more than once, Δ_Γ fails to be an isomorphism. In this case, a general procedure to complete the prequantum operator constraint program does not appear to be available. In specific examples, using the fact that Δ_Γ fails to be an isomorphism, one can easily complete the program in such a way that the resulting quantum theory is *inequivalent* to the one obtained via $\bar{\mathbf{P}}$. However, the arbitrariness involved in the completion of the program makes it difficult to draw any meaningful conclusion from this inequivalence. It is possible to analyze in detail the type of global pathologies due to which Δ_Γ can fail to be an isomorphism. However, such a detailed discussion is beyond the scope of the present paper.

III. DISCUSSION

Let us summarize the main results. Let Γ be a $2n$-dimensional manifold representing the phase space of a classical system and let $\bar{\Gamma}$ be a $(2n - m)$-dimensional constraint surface in Γ. We assume that the constraints are first class. Had there been no constraints, we could have introduced *any* polarization Γ and obtained a mathematically viable quantum theory for the system. In the presence of constraints, this procedure needs modification. In this paper, we have introduced three procedures—based on choices of appropriately defined polarization[6] on Γ, $\bar{\Gamma}$, and the reduced phase space $\hat{\Gamma}$—to incorporate constraints at the quantum level and compared the strategies involved.

The first procedure involves the introduction of a *constrained polarization* $\bar{\mathbf{P}}$ on $\bar{\Gamma}$. To obtain a $\bar{\mathbf{P}}$, one has to extend the m-dimensional gauge flat $G(\bar{\gamma})$ spanned by the constraint vector fields, to a n-dimensional, integrable, totally degenerate (w.r.t. Ω_{ab}) flat in the tangent space $\bar{T}_{\bar{\gamma}}$ within $\bar{\Gamma}$ of every point $\bar{\gamma}$ in $\bar{\Gamma}$. A mathematically viable quantum theory then results by demanding that the quantum states be U(1) Higgs scalars on $\bar{\Gamma}$, which are covariantly constant on leaves of $\bar{\mathbf{P}}$. The entire procedure is motivated by general considerations involving $\bar{\Gamma}$ alone. Yet, it turned out that there is a 1–1 correspondence between the constrained polarization $\bar{\mathbf{P}}$ on $\bar{\Gamma}$ and polarization $\hat{\mathbf{P}}$ on the reduced phase space $\hat{\Gamma}$. Furthermore, the quantum theory resulting from $\bar{\mathbf{P}}$ turns out to be naturally isomorphic with that resulting from the corresponding $\hat{\mathbf{P}}$ (Theorem 1). In practice, unless the constraints happen to be exceptionally simple, it is much more convenient to work on $\bar{\Gamma}$ than on $\hat{\Gamma}$ because each point of $\hat{\Gamma}$ is an equivalence class of points of $\bar{\Gamma}$. The notion of constrained polarization is therefore very useful in practice. This is illustrated by the example discussed in Sec. II C, where the introduction of a constrained polarization $\bar{\mathbf{P}}$ was relatively straightforward; it would have been much more complicated to first pass to the reduced phase space $\hat{\Gamma}$ and then find a polarization $\hat{\mathbf{P}}$ on it because one of the constraints is quadratic in momenta. Finally, we have the possibility of introducing a polarization **P** on the full phase space Γ and incorporating the constraints as prequantum operator conditions on permissible wave functions. However, we saw (in Sec. II A) that a general polarization **P** need *not* be compatible with the prequantum operator constraints. When incompatibility arises, the Hilbert space of quantum states is simply not "large enough" for the quantum description to be viable. The necessary and sufficient conditions for compatibility is that **P** be preserved under the action of the Hamiltonian vector fields generated by constraints. Such a polarization is said to be *compatible with constraints. A priori*, these considerations appear to be independent of the structure available on $\bar{\Gamma}$. Yet, it turned out that every constraint-compatible polarization **P** naturally induces a constrained polarization $\bar{\mathbf{P}}$ on $\bar{\Gamma}$ (Theorem 2). Furthermore, there is a natural 1–1 correspondence between the $\bar{\mathbf{P}}$-polarized wave functions $\underline{\psi}$ on $\bar{\Gamma}$ and the **P**-polarized wave functions ψ satisfying the prequantum operator constraints in a neighborhood of $\bar{\Gamma}$ in Γ (Theorem 3). If the maximal such neighborhood fills all of Γ—it may not, because of global problems associated with leaves of **P**—one can complete the program of obtaining a quantum theory using **P** and the final theory is again naturally isomorphic with the one resulting from the constrained polarization $\bar{\mathbf{P}}$ induced on $\bar{\Gamma}$ by **P**. If the neighborhood fails to be all of Γ, however, there is, apparently, no natural way of introducing a Hilbert space structure on the space of quantum states obtained from **P**.

In practice, Γ is often the cotangent bundle over a configuration space $\mathscr{C}$. In this case, if the constraints are either independent of or linear in momenta—as for example, in Yang–Mills theory—the reduced phase space $\hat{\Gamma}$ itself inherits a natural cotangent bundle structure, and one can just use the vertical polarization on $\hat{\Gamma}$ in the passage to quantum theory. However, in many interesting cases—general relativity being an outstanding example—the constraints have a more complicated dependence on momenta and this simple procedure does not work. A natural strategy then is to look for a constrained polarization $\bar{\mathbf{P}}$ on $\bar{\Gamma}$. If one can find a $\bar{\mathbf{P}}$, one would have a mathematically viable quantum theory that then can be tested against experiments. The example in Secs. II A and II C illustrates these ideas. Since the constraints in this example are somewhat similar to those of general relativity—one is linear in momenta, and the other, quadratic—one might hope that general relativity also admits a constrained polarization $\bar{\mathbf{P}}$. Although the discovery of such a polarization will not, by itself, provide us with a complete theory of quantum gravity—the problem of introducing a Hermitian inner product on the space of polarized wave functions is highly nontrivial for systems with infinite degrees of freedom—it would represent a major step towards the goal.

We should emphasize, however, that the techniques introduced in this paper do not by any means exhaust the way in which one can obtain a quantum description of systems with first class constraints, nor is there any *a priori* guarantee that the physically correct description can always be obtained via these techniques. For instance, in specific examples, one can use the following alternative strategy. One may simply introduce a polarization **P** on all of Γ and *then* impose constraints as suitable operator equations—which, in general, would *not* be the prequantum operator constraints—on the space of polarized wave functions. By very construction, this procedure would avoid the problem of compatibility of polarization conditions with constraint equations; quantum constraints, being operators on the space of polarized wave functions, would automatically leave this space invariant. In specific examples, one *can* carry out this procedure and show that the resulting quantum theory would *not* be equivalent to that obtained from *any* constrained polarization whatsoever.[17] Thus, the quantum theories resulting from the techniques introduced in this paper do not exhaust all possibilities. The interest in these techniques lies, rather, in the fact that they constitute a general framework; they provide avenues that can *always* be followed to obtain a complete and consistent quantum theory. The alternative strategy given above, on the other hand, does not share this "universality"; it requires special input—or "tricks"—at a number of intermediate steps which may or may not be available for the system under consideration. For instance, *a priori*, no factor ordering procedure is available to pass from classical constraints to quantum—rather than prequantum—operator equations; there is no guarantee that the quantum operator constraints would be closed under the commutator bracket[18]; and, we do not have a prescription to introduce a Hermitian inner product on the space of solutions to the operator constraints. In the constrained-polarization method, on the other hand, one has to provide a new input just once; in the choice of $\overline{\mathbf{P}}$. Furthermore, it is the type of input that one must provide in the passage to quantum theory even in absence of constraints. Once the input is given, the framework automatically leads one to a mathematically satisfactory and complete quantum kinematics.

ACKNOWLEDGMENTS

We wish to thank Kumar Narain for discussions.

This work was supported in part by the National Science Foundation Contract No. PHY83-10041 with Syracuse University and by the Alfred P. Sloan Foundation.

[1]Several of these results were reported by one of us (AA) at the special trimester on mathematical physics at the Stefan Banach Institute of the Polish Academy of Sciences, Warsaw, in October 1983.

[2]See, e.g., J. Śniatycki, *Geometrical Quantization and Quantum Mechanics* (Springer, New York, 1980).

[3]See, e.g., N. M. J. Woodhouse, *Geometric Quantization* (Clarendon, Oxford, 1980).

[4]For simplicity, we shall always begin with C^∞ manifolds and fields. Various operations such as Cauchy completions, solving operator constraint equations, etc. may subsequently lead us to nonsmooth fields.

[5]In fact, in Eqs. (3), (4), and (4') we are assuming that global potentials A_a exist and are all homologous. If this fails, one has to cover Γ by patches in each of which a potential A_a exists and impose appropriate overlap conditions on the resulting prequantum wave functions. See, e.g., M. Stillerman, Ph.D. thesis, Syracuse University, 1985.

[6]For simplicity, we restrict ourselves to *real* polarizations. Note, however, that not all symplectic manifolds admit a real polarization and the use of complex polarization then becomes essential. An example is $\Gamma = S^2$. Also, certain complex polarizations—the Kähler ones—have the nice property that the prequantum Hilbert space H_p itself provides the correct quantum Hermitian inner product. For details, see, e.g., M. Stillerman, Ref. 5.

[7]For simplicity, we have assumed here that the first homology of polarization leaves is trivial. If this fails, we have to introduce "patches" (see Ref. 5) on $\mathscr{C}$.

[8]It is easy to verify that $\mathscr{L}_V(f - X_f^a A_a) = 0$ for all v tangential to the polarization leaves. Therefore $f - X_f^a A_a$ can be projected down from Γ to $\mathscr{C}$ unambiguously. (Div X_f^a) is, by definition, a function on $\mathscr{C}$. Hence F is well-defined.

[9]If (Γ,Ω) is the cotangent bundle over an n-manifold $\mathscr{C}$, this class of observables consists precisely of linear combination of configuration observables (i.e., pullbacks to Γ of functions on $\mathscr{C}$) and momentum observables (i.e., functions on Γ of the type $v^a p_a$, where v^a is a vector field on $\mathscr{C}$) and the quantum description constructed here reduces to the usual one, given, e.g., in A. Ashtekar, Commun. Math. Phys. 77, 59 (1980).

[10]See, e.g., Ref. 2, Chaps. 5 and 7; Ref. 3, Chap. 5.

[11]The theory of first class constraints is due to P. G. Bergmann and P. A. M. Dirac. See, e.g., P. A. M. Dirac, *Lectures on Quantum Mechanics, Belfer Graduate Monograph Series*, Number 2 (Yeshiva U. P., New York, 1964), and references therein. The treatment given here is taken from A. Ashtekar and R. Geroch, Rep. Prog. Phys. **37**, 1211 (1974), Appendix 3.

[12]This notion of gauge is, of course, quite different from that related to the freedom in the choice of the symplectic potential, encountered in Sec. I A. To distinguish between the two, we consistently use quotation marks ("gauge") while discussing symplectic potentials.

[13]We use the interior of the future null cone of a point 0—rather than all of Minkowski space—as the configuration space $\mathscr{C}$ to avoid having to excise points to obtain a smooth constraint surface $\overline{\Gamma}$ later on.

[14]Note that v^a leaves the configuration space $\mathscr{C}$—i.e., the interior of the future light cone of 0—invariant. Had we chosen a Killing field or a pure dilation for v^a, the inconsistency that we point out in this subsection would not have occurred.

[15]Because of (23') and (27), ψ is a solution of the wave equation in (an open region of) a three-dimensional Minkowski space. Because $q^a \partial_a \psi = 0$, ψ is, in addition, constant along radial directions in this space. Hence ψ is completely determined by its restriction $\underline{\psi}$ to the two-dimensional hyperboloid of unit, future-directed timelike vectors, and $\underline{\psi}$ satisfies the two-dimensional Laplace equation on the hyperboloid. Hence the only ψ which remains bounded at infinity is the constant one. Even if we were to drop the condition at infinity, the Hilbert space is too restrictive: it does not contain arbitrary functions of $n - m \equiv 2$ variables.

[16]We have to excise the points $p_a = 0$ at which $\overline{\Gamma}$ is nonsmooth. Therefore $\mathscr{C}$ is the null cone $-p_0^2 + p_2^2 + p_3^2 = 0$ without its origin; it consists of two copies of $S^1 \times R$.

[17]See, e.g., A. Ashtekar and G. T. Horowitz, Phys. Rev. D **26**, 3342 (1982).

[18]For example, the consistency of quantum operator constraints in the configuration representation is a major problem in quantum gravity. It is known that "obvious" factor orderings run into inconsistencies; J. L. Anderson, in *Eastern Theoretical Physics Conference* (Gordon and Breach, New York, 1967).

Time-dependent invariants and quantum mechanics in two dimensions

A. K. Dhara and S. V. Lawande
Theoretical Physics Division, Bhabha Atomic Research Centre, Trombay, Bombay-400 085, Maharashtra, India

(Received 29 August 1984; accepted for publication 30 December 1985)

A class of time-dependent problems in two space dimensions possessing time-dependent invariants, bilinear in momenta, is considered. Explicit expression of the potentials and the corresponding invariants are derived. Quantum mechanics is introduced in these time-dependent problems directly through a Feynman propagator defined as a path integral involving the classical action. The propagators are shown to admit expansions in terms of the eigenfunctions of the corresponding invariant operators. Equivalence of the present theory to that of Lewis and Riesenfeld [H. R. Lewis, Jr. and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969)] is discussed.

I. INTRODUCTION

The problem of obtaining exact invariants of motion for certain time-dependent systems has received a great deal of attention. Most of the systems considered so far have been one dimensional for which invariants that are either linear or quadratic in momentum are reported. These invariants have invoked interest partly because of their relation to certain pairs of nonlinear equations,[1–3] and partly because of their utility in solving a class of time-dependent quantum mechanical problems.[4–9] Apart from their intrinsic mathematical interest, the invariants have invoked much attention because of their use in discussing several physical problems;[4,10–12] for example, the fact that an invariant satisfied Liouville equation has been exploited recently to construct exact time-dependent solutions of the Vlasov–Poisson and Vlasov–Maxwell equations.[10,11]

In the present paper, we consider the class of time-dependent potentials in two dimensions possessing the most general type of invariant that is bilinear in momenta. For time-independent problems in two dimensions the existence of an invariant, other than the energy integral, containing the bilinear term has been discussed by several authors.[13–15] However, for time-dependent systems the algebra is more involved; the derivation is essentially based on an extension of the direct method due to Lewis and Leach.[16]

One of the motivations in looking for such invariants is towards solving the time-dependent quantum mechanical problems. Lewis and Riesenfeld[4] first exploited the invariant operators in this direction. They showed that the general solution of the time-dependent Schrödinger equation can be expressed in terms of the eigenfunctions of the invariant operator of the corresponding problem.

A more direct way of introducing quantum mechanics for a problem specified by classical Lagrangian is through the Feynman propagator. Dhara and Lawande[8,9] have shown recently that the existence of an invariant greatly simplifies the derivation of the propagator. It turns out that the Feynman propagator for the time-dependent problem in one dimension is related to the propagator for an associated time-independent problem. We show, in this paper, that this result is also valid for two-dimensional problems.

A brief outline of the present paper is as follows. In Sec. II, we outline the derivations of the forms of the potential that admit the invariants and the corresponding expressions for the invariants. Applications to quantum mechanics via the Feynman propagator are discussed in Sec. III. Furthermore, we show that the propagator admits an expansion in terms of the eigenfunctions of the invariant operator, and discuss the equivalence of the Lagrangian approach to that of Lewis and Riesenfeld.[4] Finally, some concluding remarks are added in Sec. IV.

II. DERIVATION OF THE INVARIANTS

We consider a particle moving in a two-dimensional space under a time-dependent potential $V(\mathbf{q},t)$, $\mathbf{q} = (q_1, q_2)$. The Hamiltonian is

$$H(\mathbf{q}, \mathbf{p},t) = (\mathbf{p}^2/2) + V(\mathbf{q}, t), \quad \mathbf{p} = (p_1, p_2). \tag{2.1}$$

We assume that the Hamiltonian (2.1) admits most general time-dependent invariants, bilinear in momenta, of the form

$$\begin{aligned} I(\mathbf{q}, \mathbf{p},t) = {} & p_1^2 f_1(\mathbf{q}, t) + p_1 p_2 f_2(\mathbf{q}, t) \\ & + p_2^2 f_3(\mathbf{q}, t) + p_1 f_4(\mathbf{q},t) \\ & + p_2 f_5(\mathbf{q}, t) + f_0(\mathbf{q}, t). \end{aligned} \tag{2.2}$$

The defining equation of the invariant is

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + \sum_{i=1}^{2} \left(\frac{\partial I}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial I}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0. \tag{2.3}$$

Equation (2.3) imposes restriction on the potentials. Our aim is to find out the explicit form of the potential as well as that of the invariant. Equating different powers of the p's we get the following set of first-order partial differential equations (PDE):

$$\frac{\partial f_1}{\partial q_1} = 0, \tag{2.4}$$

$$\frac{\partial f_3}{\partial q_2} = 0, \tag{2.5}$$

$$\frac{\partial f_2}{\partial q_1} + \frac{\partial f_1}{\partial q_2} = 0, \tag{2.6}$$

$$\frac{\partial f_2}{\partial q_2} + \frac{\partial f_3}{\partial q_1} = 0, \tag{2.7}$$

$$\frac{\partial f_2}{\partial t} + \frac{\partial f_5}{\partial q_1} + \frac{\partial f_4}{\partial q_2} = 0, \tag{2.8}$$

 0022-2488/86/051331-09$02.50

$$\frac{\partial f_1}{\partial t}+\frac{\partial f_4}{\partial q_1}=0, \qquad (2.9)$$

$$\frac{\partial f_3}{\partial t}+\frac{\partial f_5}{\partial q_2}=0, \qquad (2.10)$$

$$\frac{\partial f_4}{\partial t}+\frac{\partial f_0}{\partial q_1}-2f_1\frac{\partial V}{\partial q_1}-f_2\frac{\partial V}{\partial q_2}=0, \qquad (2.11)$$

$$\frac{\partial f_5}{\partial t}+\frac{\partial f_0}{\partial q_2}-f_2\frac{\partial V}{\partial q_1}-2f_3\frac{\partial V}{\partial q_2}=0, \qquad (2.12)$$

$$\frac{\partial f_0}{\partial t}-f_4\frac{\partial V}{\partial q_1}-f_5\frac{\partial V}{\partial q_2}=0. \qquad (2.13)$$

The solution of these PDE determine the forms of the potential and the invariant. After solving the Eqs. (2.4)–(2.13) consistently, we arrive at the following expressions of the potentials and the corresponding invariants.[17]

A. Invariants having a bilinear form in momenta

There are two invariants in this case. The potential giving rise to these as well as the form of the invariants are listed below.

Case (A):

$$V(\mathbf{q},t) = -\sum_{i=1}^{2}\ddot{x}_i(q_i-x_i) + \frac{1}{r^2}\Gamma\left(\frac{q_1-x_1}{q_2-x_2}\right)+G(r^2,t), \qquad (2.14)$$

$$I(\mathbf{q},\mathbf{p},t)=\tfrac{1}{2}[(p_1-\dot{x}_1)(q_2-x_2)-(p_2-\dot{x}_2)(q_1-x_1)]^2 + \Gamma((q_1-x_1)/(q_2-x_2)), \qquad (2.15)$$

where

$$r^2=\sum_{i=1}^{2}(q_i-x_i)^2, \qquad (2.16)$$

and $x_1(t), x_2(t), G$, and Γ are the arbitrary functions of their arguments.

Note that although the potential (2.14) contains an arbitrary function G, the corresponding invariant (2.15) does not contain any term involving G. The term involving Γ in potential looks like the usual centrifugal term modified by a factor depending on the angle between the shifted coordinates. The first term in (2.14) expresses the linear perturbation. The quadratic term in the invariant (2.15) appears as time-dependent angular momentum in shifted coordinates. We may, however, remark that the appearance of the time-dependence in the potential (2.14) is more relaxed through arbitrary function G in view of the time-dependence in the corresponding invariant (2.15). An interesting special case, for which $\Gamma\equiv 0$ and $G=\omega^2(t)r^2$, is the linearly perturbed time-dependent isotropic oscillator. The invariant (2.15) shows that time-dependent angular momentum is indeed the conserved quantity for it.

Case (B):

$$V=\frac{f(u)+g(v)}{u^2-v^2}+\ddot{\gamma}Q_1, \qquad (2.17)$$

$$I=\tfrac{1}{2}[Q_2P_1-Q_1P_2]^2+(c^2/2)P_1P_2 + \left(v^2-\frac{c^2}{2}\right)\frac{f(u)}{u^2-v^2}+\left(u^2-\frac{c^2}{2}\right)\frac{g(v)}{u^2-v^2}, \qquad (2.18)$$

where

$$2u^2=(R^2+c^2)+[(R^2+c^2)^2-4c^2\tilde{Q}_1^2]^{1/2}, \qquad (2.19)$$

$$2v^2=(R^2+c^2)-[(R^2+c^2)^2-4c^2\tilde{Q}_1^2]^{1/2}, \qquad (2.20)$$

$$R^2=\tilde{Q}_1^2+\tilde{Q}_2^2, \qquad (2.21)$$

$$\tilde{Q}_1=(1/\sqrt{2})(Q_1+Q_2), \qquad (2.22)$$

$$\tilde{Q}_2=(1/\sqrt{2})(-Q_1+Q_2), \qquad (2.23)$$

$$Q_1=q_1+\gamma(t), \qquad (2.24)$$

$$Q_2=q_2+\alpha_0, \qquad (2.25)$$

$$P_1=\dot{Q}_1, \qquad (2.26)$$

$$P_2=\dot{Q}_2, \qquad (2.27)$$

and f, g, and $\gamma(t)$ are arbitrary functions of their respective arguments, and α_0 and c are some arbitrary constants.

Another class of potential can be obtained by replacing Q_1 by Q_2 and $\gamma(t)$ by $\alpha(t)$ in Eq. (2.17). In this case, γ will be an arbitrary constant γ_0 and the invariant will be same as Eq. (2.18).

We note that the potential (2.17) contains elliptic coordinates. So these type of potentials are useful where the problem has an elliptic symmetry and in these coordinates the problem is separable.[18] For the time-independent case, it goes over to the known result exactly.[14,15]

B. Invariants having a quadratic form in the momenta

For $f_2(\mathbf{q},t)\equiv 0$, we obtain the invariants that are quadratic in momenta. The potentials and the corresponding invariants are listed below.

Case (C):

$$V(\mathbf{q},t)=\sum_{k1}^{2}\left\{-\frac{\ddot{\rho}_k q_k^2}{2\rho_k}+(\ddot{\rho}_k x_k-\rho_k\ddot{x}_k)\frac{q_k}{\rho_k} + \rho_k^{-2}G_k\left(\frac{q_k-x_k}{\rho_k}\right)\right\}, \qquad (2.28)$$

$$I(\mathbf{q},\mathbf{p},t)=\sum_{k=1}^{2}\left\{\frac{1}{2}[\rho_k(p_k-\dot{x}_k)-\dot{\rho}_k(q_k-x_k)]^2 + G_k\left(\frac{q_k-x_k}{\rho_k}\right)\right\}, \qquad (2.29)$$

where $\rho_1(t)$, $\rho_2(t)$, $x_1(t)$, $x_2(t)$, G_1, and G_2 are arbitrary functions of their respective arguments. It is interesting to note here that the invariant (2.29) is a generalization of the one-dimensional invariant obtained by Lewis and Leach.[16] As an example we consider a two-dimensional anisotropic harmonic oscillator with time-dependent frequency acted on by the external force. The potential then has the form

$$V(\mathbf{q},t)=\tfrac{1}{2}[\Omega_1^2(t)q_1^2+\Omega_2^2(t)q_2^2] + F_1(t)q_1+F_2(t)q_2. \qquad (2.30)$$

Comparing with the form (2.28), we have $G_k=0$ $(k=1,2)$ and two auxiliary equations for each k:

$$\ddot{\rho}_k + \Omega_k^2(t)\rho_k = 0,$$
$$\ddot{x}_k + \Omega_k^2(t)x_k = F_k(t), \qquad (2.31)$$

for determining the functions ρ_k and x_k. The equations (2.28) and (2.29), however, imply that the more general external forces derivable from $\rho_k^{-2}G_k(q_k - x_k)/\rho_k$ may be added to basic harmonic oscillator potential (2.30). These forces depend on the ρ_k and x_k and require tailoring according to the auxiliary equations (2.31).

Case (D):

$$V(\mathbf{q}, t) = -\frac{1}{2}\left(\frac{\ddot{\rho}}{\rho} + \frac{\gamma^2}{\rho^4}\right)(q_1^2 + q_2^2) + \frac{\dot{\gamma}}{\rho^2} q_1 q_2 + \left[(\ddot{\rho}x_1 - \rho\ddot{x}_1) + \gamma\frac{d}{dt}\left(\frac{x_2}{\rho}\right)\right]\frac{q_1}{\rho} + \left[(\ddot{\rho}x_2 - \rho\ddot{x}_2) + \gamma\frac{d}{dt}\left(\frac{x_1}{\rho}\right)\right]\frac{q_2}{\rho}. \qquad (2.32)$$

The corresponding invariant has the form

$$I(\mathbf{q}, \mathbf{p}, t) = \tfrac{1}{2}\{[\rho(p_1 - \dot{x}_1) - \dot{\rho}(q_1 - x_1)]^2 - [\rho(p_2 - \dot{x}_2) - \dot{\rho}(q_2 - x_2)]^2\} + \gamma(q_2p_1 - q_1p_2) + (\gamma^2/2\rho^2)(-q_1^2 + q_2^2) + (\gamma/\rho)\{(\rho\dot{x}_2 - \dot{\rho}x_2)q_1 - (\rho\dot{x}_1 - \dot{\rho}x_1)q_2\}, \qquad (2.33)$$

where $\rho(t), \gamma(t), x_1(t)$, and $x_2(t)$ are arbitrary functions of time. Equation (2.32) can be looked upon as the potential of two linearly perturbed interacting time-dependent harmonic oscillators. Note, however, the presence of $\gamma(t)$ in the quadratic term in the potential implies that the frequency of two isotropic oscillators is modulated by the interacting strength. A particularly simple case of the potential (2.32) is obtained if $\gamma = \Omega_0 = \text{const}$. The resulting potential corresponds to that of the two-dimensional isotropic oscillator acted on by an external time-dependent force. For this case, the new invariant (2.33) we obtained is functionally independent from the one obtained from (2.29) by setting $G_k \equiv 0$ and $\Omega_1(t) = \Omega_2(t)$.

Case (E):

$$V(\mathbf{q}, t) = \frac{1}{2}\left(-\frac{\ddot{\rho}}{\rho} + \frac{\gamma^2}{\rho^4}\right)(q_1^2 + q_2^2) - \left\{(\rho\ddot{x}_1 - \ddot{\rho}x_1) - \gamma\frac{d}{dt}\left(\frac{x_2}{\rho}\right)\right\}\frac{q_1}{\rho} - \left\{(\rho\ddot{x}_2 - \ddot{\rho}x_2) + \gamma\frac{d}{dt}\left(\frac{x_1}{\rho}\right)\right\}\frac{q_2}{\rho} + \rho^{-2}G(C_1, C_2). \qquad (2.34)$$

The corresponding invariant has the expression

$$I(\mathbf{q}, \mathbf{p}, t) = \tfrac{1}{2}\{[\rho(p_1 - \dot{x}_1) - \dot{\rho}(q_1 - x_1)]^2 + [\rho(p_2 - \dot{x}_2) - \dot{\rho}(q_2 - x_2)]^2\} + \gamma(q_2p_1 - q_1p_2) + (\gamma^2/2\rho^2)(q_1^2 + q_2^2) + (\gamma/\rho)[(\rho\dot{x}_2 - \dot{\rho}x_2)q_1 - (\rho\dot{x}_1 - \dot{\rho}x_1)q_2] + G(C_1, C_2), \qquad (2.35)$$

$$C_1 = ((q_1/\rho)\cos\gamma\tau + (q_2/\rho)\sin\gamma\tau) - \int^t \left\{\cos\gamma\tau'\frac{d}{dt'}\left(\frac{x_1}{\rho}\right) + \sin\gamma\tau'\frac{d}{dt'}\left(\frac{x_2}{\rho}\right)\right\}dt', \qquad (2.36)$$

$$C_2 = (-(q_1/\rho)\sin\gamma\tau + (q_2/\rho)\cos\gamma\tau) + \int^t \left\{\sin\gamma\tau'\frac{d}{dt'}\left(\frac{x_1}{\rho}\right) - \cos\gamma\tau'\frac{d}{dt'}\left(\frac{x_2}{\rho}\right)\right\}dt',$$

$$\tau = \int^t \frac{dt'}{\rho^2(t')}, \qquad (2.37)$$

while $\rho(t), x_1(t)$, and $x_2(t)$ are arbitrary functions of time and γ is some constant. The potential (2.34) corresponds to the case of a two-dimensional isotropic oscillator with a time-dependent frequency perturbed linearly by a time-dependent force and an additional arbitrary force derivable from the potential $\rho^{-2}G(C_1, C_2)$. Note that according to (2.36) this additional force generally depends (in a complicated manner) on the constant γ and the time-dependent functions ρ, x_1, and x_2. A particular case is obtained when $\gamma = 0$ whereby (2.34) takes the simple form

$$V(\mathbf{q}, t) = -(\ddot{\rho}/2\rho)(q_1^2 + q_2^2) - (\rho\ddot{x}_1 - \ddot{\rho}x_1)q/\rho - (\rho\ddot{x}_2 - \ddot{\rho}x_2)\frac{q_2}{\rho} + \rho^{-2}G\left(\frac{q_1 - x_1}{\rho}, \frac{q_2 - x_2}{\rho}\right). \qquad (2.38)$$

III. APPLICATIONS TO QUANTUM MECHANICS

A. Feynman propagator

In Sec. II, we presented the class of time-dependent potentials in two dimensions admitting invariants that are bilinear in momenta. We now discuss the application of such invariants towards solving the corresponding quantum mechanical problems. Given the classical Lagrangian, quantum mechanics can be introduced through the Feynman propagator. This approach has the advantage that the quantum superposition principle is already contained in the propagator. Further, it is expected that the existence of invariants may simplify the derivation of the propagator.

We recall here that the Feynman propagator $K(\mathbf{q}'', t''; \mathbf{q}', t')$ is the quantum mechanical amplitude for finding a particle at position $\mathbf{q}''$ at time t'' if the particle was at $\mathbf{q}'$ at an earlier time t'. The conventional definition of the propagator is the path integral[19]

$$K(\mathbf{q}'', t''; \mathbf{q}', t') = \int \exp\left[\frac{i}{\hbar}\int_{t'}^{t''} L\, dt\right]\mathscr{D}\mathbf{q}(t), \qquad (3.1)$$

where L is the classical Lagrangian

$$L(\mathbf{q}, \dot{\mathbf{q}}, t) = (\dot{\mathbf{q}}^2/2) - V(\mathbf{q}, t), \qquad (3.2)$$

while $\mathscr{D}\mathbf{q}(t)$ is the usual Feynman path differential measure implying that integrations are to be performed over all possible particle paths starting at $\mathbf{q}(t') = \mathbf{q}'$ and terminating at $\mathbf{q}(t'') = \mathbf{q}''$.

We now try to evaluate (3.1) further for a two-dimensional problem, where $\mathbf{q} \equiv (q_1, q_2)$ in each of the cases where an invariant exists. For convenience, we take the case (E) first and subsequently others in the following order.

Case (E): The admissible form of the time-dependent potential is given by Eq. (2.34). Thus the Lagrangian has the form

$$L = \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2) + \frac{1}{2}\left(\frac{\ddot{\rho}}{\rho} - \frac{\gamma^2}{\rho^4}\right)(q_1^2 + q_2^2) + \left\{(\rho\ddot{x}_1 - \ddot{\rho}x_1) - \gamma\frac{d}{dt}\left(\frac{x_2}{\rho}\right)\right\}\frac{q_1}{\rho} + \left\{(\rho\ddot{x}_2 - \ddot{\rho}x_2) + \gamma\frac{d}{dt}\left(\frac{x_1}{\rho}\right)\right\}\frac{q_2}{\rho} - \rho^{-2}G(C_1, C_2), \tag{3.3}$$

where C_1 and C_2 are as in Eqs. (2.35) and (2.36). The forms of C_1 and C_2 suggest a rotational transformation to new coordinates

$$\tilde{q}_1 = q_1 \cos\gamma\tau + q_2 \sin\gamma\tau,$$
$$\tilde{q}_2 = -q_1 \sin\gamma\tau + q_2 \cos\gamma\tau, \tag{3.4}$$

where τ is as in (2.37). With this transformation the Lagrangian takes the following form:

$$\tilde{L} = \frac{1}{2}(\dot{\tilde{q}}_1^2 + \dot{\tilde{q}}_2^2) - \frac{\gamma}{\rho^2}(\tilde{q}_2\dot{\tilde{q}}_1 - q_1\dot{\tilde{q}}_2) + \frac{1}{2}\frac{\ddot{\rho}}{\rho}(\tilde{q}_1^2 + \tilde{q}_2^2) + \left\{(\rho\ddot{X}_1 - \ddot{\rho}X_1) - 2\gamma\frac{d}{dt}\left(\frac{X_2}{\rho}\right)\right\}\frac{\tilde{q}_1}{\rho} + \left\{(\rho\ddot{X}_2 - \ddot{\rho}X_2) + 2\gamma\frac{d}{dt}\left(\frac{X_1}{\rho}\right)\right\}\frac{\tilde{q}_2}{\rho} + \rho^{-2}G\left(\frac{\tilde{q}_1 - X_1}{\rho}, \frac{\tilde{q}_2 - X_2}{\rho}\right), \tag{3.5}$$

where

$$X_1 = \rho\int^t \left\{\cos\gamma\tau'\frac{d}{dt'}\left(\frac{x_1}{\rho}\right) + \sin\gamma\tau'\frac{d}{dt'}\left(\frac{x_2}{\rho}\right)\right\}dt',$$
$$X_2 = \rho\int^t \left\{-\sin\gamma\tau'\frac{d}{dt'}\left(\frac{x_1}{\rho}\right) + \cos\gamma\tau'\frac{d}{dt'}\left(\frac{x_2}{\rho}\right)\right\}dt'. \tag{3.6}$$

Since the transformation (3.4) implies that the path differential measure transforms as $\mathscr{D}\mathbf{q} = \mathscr{D}\tilde{\mathbf{q}}$, we may write the propagator as

$$K(\mathbf{q}'', t''; \mathbf{q}', t') = \int \exp\left\{\frac{i}{\hbar}\int_{t'}^{t''}\tilde{L}\,dt\right\}\mathscr{D}\tilde{\mathbf{q}}(t). \tag{3.7}$$

As the next step, we note that it is possible to write

$$\tilde{L} = L^{(0)} + \frac{d\chi}{dt}, \tag{3.8}$$

where the new Lagrangian $L^{(0)}$ and the function χ are given by

$$L^{(0)} = \frac{1}{2}\rho^2\left\{\left[\frac{d}{dt}\left(\frac{\tilde{q}_1 - X_1}{\rho}\right)\right]^2 + \left[\frac{d}{dt}\left(\frac{\tilde{q}_2 - X_2}{\rho}\right)\right]^2\right\} - \gamma\left\{\frac{(\tilde{q}_2 - X_2)}{\rho}\frac{d}{dt}\left(\frac{\tilde{q}_1 - X_1}{\rho}\right) - \frac{(\tilde{q}_1 - X_1)}{\rho}\frac{d}{dt}\left(\frac{\tilde{q}_2 - X_2}{\rho}\right)\right\} - \rho^{-2}G\left(\frac{\tilde{q}_1 - X_1}{\rho}, \frac{\tilde{q}_2 - X_2}{\rho}\right), \tag{3.9}$$

$$\chi = W(\tilde{q}_1, \tilde{q}_2, t) - g(t), \tag{3.10}$$

where

$$W(\tilde{q}_1, \tilde{q}_2, t) = \frac{\dot{\rho}}{2\rho}(\tilde{q}_1^2 + \tilde{q}_2^2) + \rho\left[\tilde{q}_1\frac{d}{dt}\left(\frac{X_1}{\rho}\right) + \tilde{q}_2\frac{d}{dt}\left(\frac{X_2}{\rho}\right)\right] + \frac{\gamma}{\rho^2}(X_1\tilde{q}_2 - X_2\tilde{q}_1), \tag{3.11}$$

$$g(t) = \frac{1}{2}\int^t \rho^2\left\{\left[\frac{d}{dt'}\left(\frac{X_1}{\rho}\right)\right]^2 + \left[\frac{d}{dt'}\left(\frac{X_2}{\rho}\right)\right]^2\right\} - \gamma\int^t (\dot{X}_1X_2 - X_1\dot{X}_2)\frac{dt'}{\rho^2}. \tag{3.12}$$

The path integral (3.7) now assumes the form

$$K(\mathbf{q}'', t''; \mathbf{q}', t') = \exp\left\{\frac{i}{\hbar}[\chi(t'') - \chi(t')]\right\} \times K^{(0)}(\tilde{\mathbf{q}}'', t''; \tilde{\mathbf{q}}', t'), \tag{3.13}$$

where

$$K^{(0)}(\tilde{\mathbf{q}}'', t''; \tilde{\mathbf{q}}', t') = \int \exp\left\{\frac{i}{\hbar}\int_{t'}^{t''} L^{(0)}\,dt\right\}\mathscr{D}\tilde{\mathbf{q}}(t). \tag{3.14}$$

Further reduction of (3.13) is made by introducing a transformation to new time τ related to the old time t by (2.37), and letting $Q_k = (\tilde{q}_k - X_k)/\rho$. Such a transformation induces a change in the path differential measure $\mathscr{D}\tilde{q}_k(t)$ given by[7–9]

$$\mathscr{D}\tilde{q}_k(t) = (\rho'\rho'')^{-1/2}\mathscr{D}Q_k(\tau), \quad (3.15)$$

where $\rho' = \rho(t')$ and $\rho'' = \rho(t'')$. The required propagator is then given by

$$K(\mathbf{q}'', t''; \mathbf{q}', t') = \left(\frac{1}{\rho'\rho''}\right)\exp\left\{\frac{i}{\hbar}[\chi(t'') - \chi(t')]\right\} \times \bar{K}^{(0)}(\mathbf{Q}'', \tau''; \mathbf{Q}', \tau'), \quad (3.16)$$

Here the new propagator

$$\bar{K}^{(0)}(\mathbf{Q}'', \tau''; \mathbf{Q}', \tau') = \int \exp\left\{\frac{i}{\hbar}\int_{\tau'}^{\tau''} \mathscr{L}^{(0)}\, d\tau\right\}\mathscr{D}\mathbf{Q}(\tau) \quad (3.17)$$

corresponds to the new Lagrangian $\mathscr{L}^{(0)}(\mathbf{Q}, d\mathbf{Q}/d\tau)$ given by

$$\mathscr{L}^{(0)} = \frac{1}{2}\left[\left(\frac{dQ_1}{d\tau}\right)^2 + \left(\frac{dQ_2}{d\tau}\right)^2\right] - \gamma\left(Q_2\frac{dQ_1}{d\tau} - Q_1\frac{dQ_2}{d\tau}\right) - G(Q_1, Q_2). \quad (3.18)$$

Clearly, the Lagrangian $\mathscr{L}^{(0)}(\mathbf{Q}, d\mathbf{Q}/d\tau)$ corresponds to a time-independent problem. The expression (3.18) is obtained from (3.9) with the help of new space-time $(\mathbf{Q}, \tau)$. It is thus seen that the original propagator is related to that for an associated time-independent problem. Further insight into the result obtained above may be obtained if we note that the classical Hamiltonian $H^{(0)}$ corresponding to $\mathscr{L}^{(0)}$ has the form

$$H^{(0)} = \tfrac{1}{2}(P_1^2 + P_2^2) + \gamma(Q_2P_1 - Q_1P_2) + (\gamma^2/2)(Q_1^2 + Q_2^2) + G(Q_1, Q_2), \quad (3.19)$$

where the canonically conjugate momenta p_1 and p_2 are defined as usual by

$$p_1 = \frac{\partial\mathscr{L}^{(0)}}{\partial(dQ_1/d\tau)} = \frac{dQ_1}{d\tau} - \gamma Q_2,$$
$$p_2 = \frac{\partial\mathscr{L}^{(0)}}{\partial(dQ_2/d\tau)} = \frac{dQ_2}{d\tau} + \gamma Q_1. \quad (3.20)$$

On the other hand, when the coordinate transformation (3.4) and the corresponding momentum transformation

$$\tilde{p}_1 = p_1\cos\gamma\tau + p_2\sin\gamma\tau,$$
$$\tilde{p}_2 = -p_1\sin\gamma\tau + p_2\cos\gamma\tau, \quad (3.21)$$

are applied to the invariant (2.35) one obtains the invariant $\tilde{I}$ expressed as a function of the new variables:

$$\tilde{I} = \tfrac{1}{2}[\rho(\tilde{p}_1 - \dot{X}_1) - \dot{\rho}(\tilde{q}_1 - X_1)]^2 + \tfrac{1}{2}[\rho(\tilde{p}_2 - \dot{X}_2) - \dot{\rho}(\tilde{q}_2 - X_2)]^2 + \gamma(\tilde{q}_2\tilde{p}_1 - \tilde{q}_1\tilde{p}_2) + (\gamma^2/2\rho^2)(\tilde{q}_1^2 + \tilde{q}_2^2) + \gamma\rho\left\{\tilde{q}_1\frac{d}{dt}\left(\frac{X_2}{\rho}\right) - \tilde{q}_2\frac{d}{dt}\left(\frac{X_1}{\rho}\right)\right\} + G\left(\frac{\tilde{q}_1 - X_1}{\rho}, \frac{\tilde{q}_2 - X_2}{\rho}\right). \quad (3.22)$$

Further when expressed in terms of the new variables Q_k, P_k, and τ, the invariant $\tilde{I}(\tilde{\mathbf{q}}, \tilde{\mathbf{p}}, t)$ reduces to $I^{(0)}(\mathbf{Q}, \mathbf{P})$, which is identical to $H^{(0)}$. Since $\mathbf{Q}$ and $\mathbf{P}$ are canonically conjugate variables, the quantum Hamiltonian $\hat{H}^{(0)}$ and the invariant operator $\hat{I}^{(0)}$ are obtained by writing $P_k = -i\hbar(\partial/\partial Q_k)$ in (3.19):

$$\hat{H}^{(0)} = -\frac{\hbar^2}{2}\left(\frac{\partial^2}{\partial Q_1^2} + \frac{\partial^2}{\partial Q_2^2}\right) - i\hbar\gamma\left(Q_2\frac{\partial}{\partial Q_1} - Q_1\frac{\partial}{\partial Q_2}\right) + \frac{\gamma^2}{2}(Q_1^2 + Q_2^2 + G(Q_1, Q_2) \equiv \hat{I}^{(0)}. \quad (3.23)$$

The propagator $\bar{K}^{(0)}(\mathbf{Q}'', \tau''; \mathbf{Q}', \tau')$ then represents the Green's function of the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial\tau}(\mathbf{Q}, \tau) = \hat{H}^{(0)}\psi(\mathbf{Q}, \tau). \quad (3.24)$$

Thus, if the associated stationary problem

$$\hat{H}^{(0)}\phi_n(\mathbf{Q}) = \lambda_n\phi_n(\mathbf{Q}) \quad (3.25)$$

admits a complete set of normalized eigenfunctions $\phi_n(\mathbf{Q})$ corresponding to eigenvalues λ_n, the propagator has an expansion

$$\bar{K}^{(0)}(\mathbf{Q}'', \tau''; \mathbf{Q}', \tau') = \sum_n \exp\left[-\frac{i\lambda_n}{\hbar}(\tau'' - \tau')\right]\phi_n^*(\mathbf{Q}')\phi_n(\mathbf{Q}''), \quad (3.26)$$

where the summation over the index n stands for summation over discrete eigenvalues and an integration over continuous ones. The complete propagator (3.16) has then the following expansion in terms of the eigenfunctions of the invariant operator

$$K(\mathbf{q}'', t''; \mathbf{q}', t') = \left(\frac{1}{\rho'\rho''}\right)\exp\left\{\frac{i}{\hbar}[\chi(t'') - \chi(t')]\right\} \times \sum_n \exp\left[-\frac{i\lambda_n}{\hbar}\int_{t'}^{t''}\frac{dt}{\rho^2}\right]\phi_n^*(\mathbf{Q}')\phi_n(\mathbf{Q}''). \quad (3.27)$$

Case (D): In this case the admissible potential $V(\mathbf{q}, t)$ is given in expressions (2.32). We write the Lagrangian as

$$L = \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2) + \frac{1}{2}\left(\frac{\ddot{\rho}}{\rho} + \frac{\gamma^2}{\rho^4}\right)(q_1^2 + q_2^2) + \left\{(\rho\ddot{x}_1 - \ddot{\rho}x_1) - \gamma\frac{d}{dt}\left(\frac{x_2}{\rho}\right)\right\}\frac{q_1}{\rho} + \left\{(\rho\ddot{x}_2 - \ddot{\rho}x_2) - \gamma\frac{d}{dt}\left(\frac{x_1}{\rho}\right)\right\}\frac{q_2}{\rho} - \frac{\dot{\gamma}}{\rho^2}q_1q_2. \quad (3.28)$$

We first make an orthogonal transformation to a new coordinate $(\tilde{q}_1, \tilde{q}_2)$ as

$$\tilde{q}_1 = (1/\sqrt{2})(q_1 + q_2), \quad \tilde{q}_2 = (1/\sqrt{2})(-q_1 + q_2). \quad (3.29)$$

In terms of the new coordinates the Lagrangian becomes separable in $\tilde{q}_1$ and $\tilde{q}_2$. We can write

$$\tilde{L} = \tilde{L}_1(\tilde{q}_1, \dot{\tilde{q}}_1, t) + \tilde{L}_2(\tilde{q}_2, \dot{\tilde{q}}_2, t), \quad (3.30)$$

where

$$\tilde{L}_1 = \frac{\dot{\tilde{q}}_1^2}{2} + \frac{\ddot{\rho}_1 \tilde{q}_1^2}{2\rho_1} + (\rho_1 \ddot{X}_1 - \ddot{\rho}_1 X_1)\frac{\tilde{q}_1}{\rho_1},$$
$$\tilde{L}_2 = \frac{\dot{\tilde{q}}_2^2}{2} + \frac{\ddot{\rho}_2 \tilde{q}_2^2}{2\rho_2} + (\rho_2 \ddot{X}_2 - \ddot{\rho}_2 X_2)\frac{\tilde{q}_2}{\rho_2}, \quad (3.31)$$

$$\rho_1(t) = \rho(t)\exp\left(-\int^t \gamma \rho^{-2}\, dt'\right),$$
$$\rho_2(t) = \rho(t)\exp\left(\int^t \gamma \rho^{-2}\, dt'\right), \quad (3.32)$$

$$X_k = \tilde{x}_k - \rho_k \int^t \gamma \frac{\tilde{x}_k}{\rho_k}\rho^{-2}\, dt' \quad (k = 1, 2), \quad (3.33)$$

$$\tilde{x}_1 = (1/\sqrt{2})(x_1 + x_2), \quad (3.34)$$

$$\tilde{x}_2 = (1/\sqrt{2})(-x_1 + x_2). \quad (3.35)$$

Since the transformation (3.29) implies that path differential measure transforms as $\mathscr{D}\mathbf{q} = \mathscr{D}\tilde{\mathbf{q}}$, we may write the propagator (3.7) as

$$K(\mathbf{q}'', t''; \mathbf{q}', t') = \prod_{k=1}^{2} \int \exp\left(\frac{i}{\hbar}\int_{t'}^{t''} \tilde{L}_k\, dt\right)\mathscr{D}\tilde{q}_k$$
$$= \prod_{k=1}^{2} \tilde{K}_k(\tilde{q}_k'', t''; \tilde{q}_k', t'). \quad (3.36)$$

In order to obtain the propagators $\tilde{K}_1$ and $\tilde{K}_2$, we perform steps similar to those described in (3.8) and onwards for case (E). We will quote the results. The final propagator for this case is

$$K(\mathbf{q}'', t''; \mathbf{q}', t')$$
$$= \prod_{k=1}^{2} (\rho_k' \rho_k'')^{-1/2} \exp\left\{\frac{i}{\hbar}[\chi_k(t'') - \chi_k(t')]\right\}$$
$$\times \bar{K}_k^{(0)}(Q_k'', \tau_k''; Q_k', \tau_k'), \quad (3.37)$$

$$\bar{K}_k^{(0)}(Q_k'', \tau_k''; Q_k', \tau_k')$$
$$= [2\pi i\hbar(\tau_k'' - \tau_k')]^{-1/2}$$
$$\times \exp\left\{\frac{i}{2\hbar}\frac{[Q_k(\tau_k'') - Q_k(\tau_k')]^2}{(\tau_k'' - \tau_k')}\right\}, \quad (3.38)$$

where

$$\tau_k = \int^t \rho_k^{-2}(s)\, ds, \quad (3.39)$$

$$Q_k = (\tilde{q}_k - X_k)/\rho_k, \quad (3.40)$$

$$\chi_k = \frac{\dot{\rho}_k \tilde{q}_k^2}{2\rho_k} + \frac{W_k \tilde{q}_k}{\rho_k} - g_k, \quad (3.41)$$

$$W_k = \rho_k \dot{X}_k - \dot{\rho}_k X_k = \rho_k^2 \frac{d}{dt}\left(\frac{X_k}{\rho_k}\right), \quad (3.42)$$

$$g_k = \frac{1}{2}\int^t \rho_k^{-2} W_k^2\, dt'. \quad (3.43)$$

Hence, the propagator for the original time-dependent problem of case (D) is shown to be related to a product of two free-particle propagators in the new space-time (Q_k, τ_k), $k = 1, 2$.

We now apply the transformation (3.29) to the invariant I of Eq. (2.33). We obtained thereby, the expression

$$\tilde{I} = -[\rho_1(\tilde{p}_1 - \dot{X}_1) - \dot{\rho}_1(\tilde{q}_1 - X_1)]$$
$$\times [\rho_2(\tilde{p}_2 - \dot{X}_2) - \dot{\rho}_2(\tilde{q}_2 - X_2)], \quad (3.44)$$

where the canonically conjugate momenta $\tilde{p}_k = \dot{\tilde{q}}_k$. After a further transformation to the new variables Q_k and the corresponding canonical momenta $P_k = dQ_k/d\tau$, the invariant reduces to another form $I^{(0)}$ given by

$$I^{(0)} = -P_1 P_2. \quad (3.45)$$

On the other hand, the Hamiltonian $H^{(0)}$ that follows from the new Lagrangian $\mathscr{L}^{(0)}$ corresponding to the reduced free particle propagator (3.37) reads as

$$H^{(0)} = \tfrac{1}{2}(P_1^2 + P_2^2). \quad (3.46)$$

Using the quantization rule $P_k = -i\hbar(\partial/\partial Q_k)$, we may write the corresponding quantum Hamiltonian operator and the quantum invariant as

$$\hat{H}^{(0)} = -\frac{\hbar^2}{2}\left(\frac{\partial^2}{\partial Q_1^2} + \frac{\partial^2}{\partial Q_2^2}\right), \quad (3.47)$$

$$\hat{I}^{(0)} = -\hbar^2 \frac{\partial^2}{\partial Q_1\, \partial Q_2}. \quad (3.48)$$

Since the operators $\hat{I}^{(0)}$ and $\hat{H}^{(0)}$ commute, they have simultaneous eigenfunctions, which are in fact that product of eigenfunctions corresponding to the momentum operators $\hat{P}_1$ and $\hat{P}_2$. The propagator $\bar{K}_k^{(0)}(Q_k'', \tau_k''; Q_k', \tau_k')$ is just the Green's function of the free-particle Shrödinger equation

$$i\hbar \frac{\partial \psi^{(k)}}{\partial \tau}{}^{(k)}(Q_k, \tau_k) = -\frac{\hbar^2}{2}\frac{\hbar^2}{\partial Q_k^2}\psi^{(k)}(Q_k, \tau_k), \quad (3.49)$$

and may be expanded in terms of the momentum eigenfunctions. In summary we see that the propagator admits an expansion in terms of the eigenfunctions of the invariant operator.

Case (C): In this case, initial Lagrangian corresponding to the potential (2.28) can be written as

$$L(\mathbf{q}, \dot{\mathbf{q}}, t) = \sum_{k=1}^{2} L_k(q_k, \dot{q}_k, t), \quad (3.50)$$

with

$$L_k(q_k, \dot{q}_k, t) = \frac{\dot{q}_k^2}{2} + \frac{\ddot{\rho}_k q_k^2}{2\rho_k}$$
$$+ (\rho_k \ddot{x}_k - \ddot{\rho}_k x_k)\frac{q_k}{\rho_k} + \rho_k^{-2} G_k\left(\frac{q_k - x_k}{\rho_k}\right). \quad (3.51)$$

The Lagrangian (3.51) is of the same form as in (3.31) and (3.32) except for the additional term involving G_k $(k = 1, 2)$. We may, however, note that no rotation of coordinates is needed in this case. Following steps similar to those described in (3.8) and onwards, final expression of the propagator can be recast in the form (3.37) except that the reduced propagator has the form

$$\bar{K}_k^{(0)}(Q_k'', \tau_k''; Q_k', \tau_k')$$
$$= \int \exp\left\{\frac{i}{\hbar}\int_{\tau_k'}^{\tau_k''} \mathscr{L}_k^{(0)}\, d\tau_k\right\}\mathscr{D}Q_k(\tau_k), \quad (3.52)$$

with

$$\mathscr{L}_k^{(0)} = \frac{1}{2}\left(\frac{dQ_k}{d\tau}\right)^2 - G_k(Q_k). \quad (3.53)$$

The functions τ_k, Q_k, χ_k, W_k, and g_k are now given by (3.39)–(3.43), where $\tilde{q}_k$ and X_k are replaced by q_k and x_k, respectively. The phase factor appearing in the expression of the propagator in this particular case matches with that of Ray and Hartley,[5] obtained by employing the Lewis–Riesenfeld theory.[4]

The classical Hamiltonian corresponding to the Lagrangian $\mathscr{L}_k^{(0)}$ in (3.53) is

$$H_k^{(0)} = P_k^2/2 + G(Q_k), \quad (3.54)$$

where P_k is the canonical momentum conjugate to the new variable Q_k. On the other hand, the invariant (2.29), when expressed in terms of new variables, is identical to the Hamiltonian (3.54). Hence, the propagator or in particular the reduced propagator (3.52) admits an expansion in terms of the eigenfunctions of the invariant operator corresponding to (2.29).

Case (A): In this case, no rotation of coordinates need be performed over the Lagrangian corresponding to the potential (2.14). Thus the propagator will assume the form (3.13) where $\tilde{\mathbf{q}}$ should be replaced by $\mathbf{q}$. The function χ is given by

$$\chi = \sum_{i=1}^{2}\left[\dot{x}_i(q_i - x_i) + \frac{1}{2}\int^t \dot{x}_i^2\, dt\right], \quad (3.55)$$

while the reduced propagator (3.14) in terms of new variables

$$Q_i = q_i - x_i, \quad i = 1, 2, \quad (3.56)$$

$$r^2 = Q_1^2 + Q_2^2, \quad (3.57)$$

$$\tan\theta = Q_2/Q_1, \quad (3.58)$$

will take the following form

$$K^{(0)}(\mathbf{r}'', t''; \mathbf{r}', t') = \int \exp\left(\frac{i}{\hbar}\int_{t'}^{t''} L^{(0)}\, dt\right)\mathscr{D}\mathbf{r}. \quad (3.59)$$

We note that the path differential measure in the reduced propagator transforms as $\mathscr{D}\mathbf{q} = \mathscr{D}\mathbf{r}$ in terms of the new variables (r, θ). The Lagrangian $L^{(0)}$ has the form

$$L^{(0)} = \frac{\dot{r}^2}{2} + \frac{r^2\dot{\theta}^2}{2} - \frac{\Gamma_0(\theta)}{r^2} - G(r^2, t), \quad (3.60)$$

with $\Gamma(z) = \Gamma_0(\cot^{-1} z)$. We may remark here that the reduced Lagrangian (3.60) is also explicitly time-dependent. Thus the propagator for the original time-dependent problem cannot be reduced to the propagator for a time-independent problem like other cases. Nevertheless, the reduced propagator (3.59) can be further simplified with the knowledge of the invariant (2.15). The Hamiltonian for the Lagrangian (3.60) is

$$H^{(0)} = \frac{p_r^2}{2} + G(r^2, t) + \frac{1}{r^2}\left[\frac{p_\theta^2}{2} + \Gamma_0(\theta)\right], \quad (3.61)$$

where

$$p_r = \dot{r}, \quad p_\theta = r^2\dot{\theta}. \quad (3.62)$$

Expression of the invariant (2.15) in terms of the polar coordinates (3.57) and (3.58)

$$I = p_\theta^2/2 + \Gamma_0(\theta). \quad (3.63)$$

It is now clear that in the representation where quantum mechanical operator corresponding to (3.63) is diagonal the Feynman propagator can be expanded in terms of the eigenfunctions of the invariant operator. If the corresponding eigenstates are denoted by $Y_l(\theta)$ then the propagator assumes the expansion

$$K^{(0)}(\mathbf{r}'', t''; \mathbf{r}', t') = \sum_l K_l(r'', t''; r', t')\, Y_l(\theta'')\, Y_l^*(\theta'). \quad (3.64)$$

Expression (3.64) shows the importance of the invariant operator. Indeed, the existence of the invariant operator simplifies the evaluation of the propagator. We need to evaluate the propagator for the radial part of the problem only.

Case (B): In this case, a similar procedure to that used in case A can be adopted. Explicit time dependence of the potential can be filtered out through the function χ in (3.13), defined as

$$\chi = -\left[\dot{\gamma}Q_1 - \int^t \frac{\dot{\gamma}^2}{2}\, dt\right]. \quad (3.65)$$

The remaining Lagrangian $L^{(0)}$ in (3.14) does not contain time explicitly. It is given by

$$L^{(0)} = \tfrac{1}{2}(\dot{Q}_1^2 + \dot{Q}_2^2) - (f(u) + g(v))/(u^2 - v^2). \quad (3.66)$$

With the help of transformations (2.22) and (2.23) along with the transformations

$$\dot{\tilde{Q}}_1 = (1/\sqrt{2})(\dot{Q}_1 + \dot{Q}_2),$$
$$\dot{\tilde{Q}}_2 = (1/\sqrt{2})(-\dot{Q}_1 + \dot{Q}_2), \quad (3.67)$$

the reduced Lagrangian $L^{(0)}$ in (3.66) can be written as

$$L^{(0)} = \tfrac{1}{2}(\dot{\tilde{Q}}_1^2 + \dot{\tilde{Q}}_2^2) - F(\tilde{Q}_1, \tilde{Q}_2), \quad (3.68)$$

where

$$F(\tilde{Q}_1, \tilde{Q}_2) = (f(u) + g(v))/(u^2 - v^2). \quad (3.69)$$

Since the transformations (2.22)–(2.25) change the path differential measure in (3.14) as $\mathscr{D}\mathbf{q} = \mathscr{D}\tilde{\mathbf{Q}}$, the propagator for this case finally takes the following form:

$$K(\mathbf{q}'', t''; \mathbf{q}', t') = \exp\left\{\frac{i}{\hbar}[\chi(t'') - \chi(t')]\right\}K^{(0)}(\tilde{\mathbf{Q}}'', t''; \tilde{\mathbf{Q}}', t'), \quad (3.70)$$

where $K^{(0)}$ is the reduced propagator for the time-independent Lagrangian (3.68). Next, to bring out the role played by the invariant operator, we express (2.18) in terms of the new coordinates $\tilde{Q}_1$ and $\tilde{Q}_2$. It has the form

$$I = \tfrac{1}{2}(\tilde{Q}_2\tilde{P}_1 - \tilde{Q}_1\tilde{P}_2)^2 + (c^2/4)(\tilde{P}_1^2 - \tilde{P}_2^2) + G(\tilde{Q}_1, \tilde{Q}_2), \quad (3.71)$$

where

$$G(\tilde{Q}_1, \tilde{Q}_2) = \left(v^2 - \frac{c^2}{2}\right)\frac{f(u)}{u^2 - v^2} + \left(u^2 - \frac{c^2}{2}\right)\frac{g(v)}{u^2 - v^2}, \quad (3.72)$$

and $\tilde{P}_1$ and $\tilde{P}_2$ are canonical momenta conjugate to the coordinates $\tilde{Q}_1$ and $\tilde{Q}_2$, respectively. It can be shown that invariant operator corresponding to (3.71) and Hamiltonian operator corresponding to the Lagrangian (3.68) commute. Hence they have the same eigenstates. So the propagator (3.70) admits an expansion in terms of the eigenstates of the invariant operator.

B. Comparison with the Lewis–Riesenfeld theory

In Sec. III A, we have shown that for time-dependent problems admitting an invariant quadratic in momenta, the Feynman propagator is related to the propagator of an associated time-independent problem. Moreover, the propagator admits an expansion in terms of the eigenfunctions of the invariant operator. It may be of some interest to compare the present approach based on the classical Lagrangian with that of the Lewis–Riesenfeld theory.[4] These authors showed that for a quantal system characterized by a time-dependent Hamiltonian $\hat{H}(t)$ and a Hermitian invariant $\hat{I}(t)$ the general solution of the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t}(\mathbf{q}, t) = \hat{H}(t)\Psi(\mathbf{q}, t) \tag{3.73}$$

can be written as

$$\Psi(\mathbf{q}, t) = \sum_n C_n e^{i\alpha_n(t)} \psi_n(\mathbf{q}, t). \tag{3.74}$$

Here $\psi_n(\mathbf{q}, t)$ are the normalized eigenfunctions of the invariant operator

$$\hat{I}\psi_n(\mathbf{q}, t) = \lambda_n \psi_n(\mathbf{q}, t), \tag{3.75}$$

where the eigenvalues λ_n are independent of time. The expansion coefficients C_n are constants, while the time-dependent phases $\alpha_n(t)$ are to be obtained from the equation

$$\hbar \frac{d\alpha_n(t)}{dt} = \langle \psi_n | i\hbar \frac{\partial}{\partial t} - \hat{H} | \psi_n \rangle. \tag{3.76}$$

Now consider, for example, the quantal Hamiltonian H corresponding to the classical Lagrangian (3.3). If we perform a unitary transformation

$$\psi_n'(\mathbf{q}, t) = U_R \psi_n(\mathbf{q}, t), \tag{3.77}$$

where U_R is the rotation operator

$$U_R = \exp\left\{\frac{i\gamma\tau}{\hbar}(\hat{q}_2\hat{p}_1 - \hat{q}_1\hat{p}_2)\right\}, \tag{3.78}$$

Eq. (3.75) transforms to

$$\hat{I}'\psi_n'(\mathbf{q}, t) = \lambda_n \psi_n'(\mathbf{q}, t), \tag{3.79}$$

where the transformed invariant operator

$$\hat{I}' = U_R \hat{I} U_R^\dagger. \tag{3.80}$$

Next apply another unitary transformation

$$\psi_n''(\mathbf{q}, t) = U\psi_n'(\mathbf{q}, t), \tag{3.81}$$

with the operator

$$U = \exp\{-(i/\hbar)W(\tilde{q}_1, \tilde{q}_2, t)\}, \tag{3.82}$$

where $W(\tilde{q}_1, \tilde{q}_2, t)$ is as in Eq. (3.11). The operator $\hat{I}'$ now changes to $\hat{I}''$

$$\hat{I}'' = U\hat{I}'U^\dagger, \tag{3.83}$$

and Eq. (3.79) becomes

$$\hat{I}''\psi_n''(\mathbf{q}, t) = \lambda_n \psi_n''(\mathbf{q}, t). \tag{3.84}$$

This transformed invariant $\hat{I}''$ when expressed in terms of the new variables $Q_k = (\tilde{q}_k - X_k)/\rho$ is the same as $\hat{I}^{(0)}$ defined in (3.23). It therefore follows that the corresponding normalized eigenfunctions

$$\phi_n(\mathbf{Q}) = \rho\psi_n''(\mathbf{q}, t). \tag{3.85}$$

Further, after applying successively the unitary transformations U_R and U, Eq. (3.76) determining the phases reduces to

$$\hbar \frac{d\alpha_n(t)}{dt} = -\langle \phi_n | \frac{\hat{I}''}{\rho^2} + \dot{g}(t) | \phi_n \rangle, \tag{3.86}$$

where $g(t)$ is as defined in (3.12). Finally using (3.84), and integrating over time t, we obtain

$$\alpha_n(t) = -\frac{\lambda_n}{\hbar}\int^t \frac{dt'}{\rho^2} - \frac{g(t)}{\hbar}. \tag{3.87}$$

Note that these phase factors appear naturally in our expansion (3.27) for the propagator K.

Similar considerations apply to other invariants also. We thus see that our formulation is equivalent to the Lewis–Riesenfeld theory. However, in the Feynman propagator approach, the steps which are essentially equivalent to those above are carried out classically on the Lagrangian leading to a transformation of the path differential measure. Moreover, the quantum-mechanical superposition principle is evident in the reduced propagator $\bar{K}^{(0)}$.

IV. CONCLUSIONS

In this paper we have considered a new family of integrable time-dependent dynamical systems for one particle in two space dimensions possessing invariants that have a general bilinear form in the momenta. We have derived the explicit expressions for the admissible potentials and also the corresponding invariants. This is achieved through the derivation and resolution of a set of partial differential equations to be satisfied by the coefficients of the bilinear form. We have shown that the existence of the invariants considerably simplifies the derivation of the Feynman propagator. It has been shown, without carrying out an explicit path integration, that the propagator in all cases other than case A, is related to the propagator for an associated time-independent problem. Further, it is shown that each propagator for these problems admits an expansion in terms of the eigenfunctions of the corresponding invariant operator. This establishes the equivalence of the present theory with that of Lewis and Riesenfeld.

[1]V. P. Ermakov, Univ. Izv. Kiev **20**, 1 (1880).
[2]J. L. Reid and J. R. Ray, J. Math, Phys. **21**, 1583 (1980).
[3]J. R. Ray and J. L. Reid, J. Math. Phys. **22**, 91 (1981).
[4]H. R. Lewis, Jr. and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[5]J. R. Ray and J. G. Hartley, Phys. Lett. A **88**, 125 (1982).
[6]J. G. Hartley and J. R. Ray, Phys. Rev. A **24**, 2873 (1981); **25**, 2388 (1982).
[7]S. V. Lawande and A. K. Dhara, Phys. Lett. A **99**, 353 (1983).
[8]A. K. Dhara and S. V. Lawande, Phys. Rev. A **30**, 560 (1984).
[9]A. K. Dhara and S. V. Lawande, J. Phys. A **17**, 2423 (1984).

[10]H. R. Lewis and K. R. Symon, Phys. Fluids **27**, 192 (1984).
[11]B. Abraham-Shrauner, Phys. Fluids **27**, 197 (1984).
[12]R. K. Colegrave and M. S. Abdalla, J. Phys. A **14**, 2269 (1981); **15**, 1549 (1982); **16**, 3805 (1983).
[13]G. Darboux, Arch. Neerl. Sci. Exactes Nat. **VI**, 371 (1901).
[14]E. T. Whittaker, *Analytical Dynamics of Particles* (Cambridge U.P., Cambridge, 1959), 4th ed., pp. 334 and 335.
[15]B. Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **24**, 2282 (1983).
[16]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982).
[17]A. K. Dhara, Ph.D. thesis, University of Bombay, 1985.
[18]P. M. Morse and H. Feshback, *Methods of Theoretical Physics (Part I)* (McGraw–Hill, New York, 1953), 5th ed., p. 657.
[19]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965).

Dynamic-group approach to the $x^2+\lambda x^2/(1+gx^2)$ potential

V. Fack, H. De Meyer,[a] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B9000 Gent, Belgium

(Received 23 July 1985; accepted for publication 21 January 1986)

A modified operator method based upon the SO (2,1) dynamic group is applied to the one-dimensional anharmonic oscillator potential $V(x)=x^2+\lambda x^2/(1+gx^2)$. A tilting transformation is carried out in order to improve the rate of convergence of the algebraic perturbation series. Very accurate results are obtained for the energy eigenvalues and for the wave functions especially in the case of small g-values.

I. INTRODUCTION

The Schrödinger equation

$$\left[\frac{d^2}{dx^2}+E-V(x)\right]\psi(x)=0, \qquad (1.1)$$

with an interaction of the type

$$V(x)=x^2+\lambda x^2/(1+gx^2), \qquad (1.2)$$

has been studied by many authors.[1–10] This type of potential is related to certain laser theory models[11] and also to a zero-dimensional field theory with a nonlinear Lagrangian.[12] Among the methods which have been followed in the investigation of the eigenvalues and eigenfunctions of the differential equation (1.1), let us mention the variational Rayleigh–Ritz formalism,[1] perturbation algorithms,[2–5] schemes based upon Padé approximants,[6] and direct numerical integration techniques.[7] Recently, exact solutions to the Schrödinger equation (1.1) with $V(x)$ given by (1.2) have been constructed.[8–10]

In the present paper we shall treat the problem in the context of algebraic perturbation theory. Therefore, we present an approach based on the Lie algebra of SO (2,1), well-known to be the dynamic group for a number of systems.[13] In fact, this so-called dynamic-group method has previously been used to treat the Yukawa potential[14,15] and certain screened Coulomb potentials.[16–18] We also introduce a tilting transformation that relates between the physical states and perturbation series of group states that are the basis of the relevant unitary irreducible representations of SO (2,1). In configuration space this transformation amounts to a scale transformation.[19] Here we use the adjustable scale parameter to accelerate the convergence of the perturbation expansions.[20] We finally obtain very accurate results for the lowest eigenvalues for a class of typical λ- and g-values.

II. ALGEBRAIC FORMULATIONS

The SO (2,1) Lie algebra consists of the generators K_1, K_2, and K_3 satisfying the commutation relations

$$\begin{aligned}[][K_1,K_2]&=-iK_3,\\ [K_2,K_3]&=iK_1,\\ [K_3,K_1]&=iK_2,\end{aligned} \qquad (2.1)$$

and the Casimir invariant

$$Q=K_3^2-K_1^2-K_2^2, \qquad (2.2)$$

of which the eigenvalues are denoted as $k(k-1)$. We shall utilize only the $\mathscr{D}^+(k)$ representation such that the orthonormal group states $|n,k\rangle$ diagonalize K_3 as

$$K_3|n,k\rangle=(n+k)|n,k\rangle \quad (n=0,1,\ldots,\ \ k>0). \qquad (2.3)$$

Introducing further the step operators $K_\pm=K_1\pm iK_2$, it is straightforward to demonstrate that

$$\begin{aligned}K_+|n,k\rangle&=[(n+1)\ (n+2K)]^{1/2}|n+1,k\rangle,\\ K_-|n,k\rangle&=[n\ (n+2k-1)]^{1/2}|n-1,k\rangle.\end{aligned} \qquad (2.4)$$

A realization of this SO (2,1) algebra relevant for one-dimensional oscillator-type Hamiltonians is[21]:

$$\begin{aligned}K_1&=\tfrac{1}{2}(\tfrac{1}{4}p^2-x^2),\\ K_2&=\tfrac{1}{2}(xp-i/2),\\ K_3&=\tfrac{1}{2}(\tfrac{1}{4}p^2+x^2),\end{aligned} \qquad (2.5)$$

where $p=-id/dx$. Moreover, for the one-dimensional case we have $Q=-\frac{3}{16}$ so that k can take the values $\frac{1}{4}$ or $\frac{3}{4}$, the first value covering the states with even parity, and the second covering the odd parity states.

On account of (2.3) and (2.5), the orthonormal group states $|n,k\rangle$ can be expressed in configuration space in terms of the usual harmonic oscillator eigenfunctions $u_n(x)=[2^n n!\sqrt{\pi}]^{-1/2}H_n(x)\exp(-x^2/2)$ as follows:

$$|n,k\rangle=\begin{cases}(-1)^n 2^{1/4}u_{2n}(\sqrt{2}x), & \text{if } k=\frac{1}{4},\\ (-1)^n 2^{1/4}u_{2n+1}(\sqrt{2}x), & \text{if } k=\frac{3}{4}.\end{cases} \qquad (2.6)$$

The extra phase factor in the expressions (2.6) stems from the relations (2.4), whereas the factor $2^{1/4}$ is necessary to preserve the normalization to unity.

The Schrödinger equation (1.1) with $V(x)$ given by (1.2) can be reexpressed with the help of (2.5) as

$$\Omega(E)|\psi\rangle=0, \qquad (2.7)$$

with

$$\Omega(E)=[1+g(K_3-K_1)]\,[4(K_3+K_1)+(K_3-K_1)-E]+\lambda(K_3-K_1). \qquad (2.8)$$

Note that we have first multiplied (1.1) on the left with $1+gx^2$. Now we introduce a supplementary parameter on performing a tilting transformation, i.e.,

$$|\bar{\psi}\rangle=e^{-i\theta K_2}|\psi\rangle, \qquad (2.9)$$

$$\bar{\Omega}(E,\theta)=e^{-i\theta K_2}\Omega(E)e^{i\theta K_2}. \qquad (2.10)$$

Hence, the eigenvalue problem (2.7) turns into an equivalent one, namely

[a] Senior Research Associate of the National Fund for Scientific Research (Belgium).

 0022-2488/86/051340-04$02.50

$$\bar{\Omega}(E,\theta) = |\bar{\psi}\rangle = 0. \tag{2.11}$$

Through the use of the Baker–Hausdorff–Campbell formula, it is readily shown that

$$e^{-i\theta K_2}(K_3 \pm K_1)e^{i\theta K_2} = e^{\pm\theta}(K_3 \pm K_1). \tag{2.12}$$

Setting

$$e^{-\theta} = \omega, \tag{2.13}$$

it follows from (2.8), (2.10), and (2.12) that

$$\begin{aligned}\bar{\Omega}(E,\theta) = {} & (4/\omega)\,(K_3 + K_1) + \omega(1 + \lambda - gE)\,(K_3 - K_1) \\ & + 4g(K_3 - K_1)\,(K_3 + K_1) \\ & + \omega^2 g(K_3 - K_1)^2 - E. \end{aligned} \tag{2.14}$$

From this expression it is easy to calculate the matrix elements of the operator $\bar{\Omega}(E,\theta)$ in the group state basis. By the aid of (2.3) and (2.4) we find

$$\begin{aligned}&\langle n,k\,|\bar{\Omega}(E,\theta)|n,k\rangle \\ &\quad = (n+k)[(1+\lambda)\omega^2 + 4]/\omega - [1 + \omega g(n+k)]E \\ &\qquad + g(n+k)^2(3\omega^2+4)/2 + gk(1-k)\,(\omega^2-4)/2,\end{aligned} \tag{2.15}$$

$$\begin{aligned}&\langle n+1,k\,|\bar{\Omega}(E,\theta)|n,k\rangle \\ &\quad = [4/\omega + 4g - \omega(1+\lambda-gE) - \omega^2 g(2n+2k+1)] \\ &\qquad \times[(n+1)\,(n+2k)]^{1/2}/2,\end{aligned} \tag{2.16}$$

$$\begin{aligned}&\langle n-1,k\,|\bar{\Omega}(E,\theta)|n,k\rangle \\ &\quad = [4/\omega - 4g - \omega(1+\lambda-gE) - \omega^2 g(2n+2k-1)] \\ &\qquad \times[n(n+2k-1)]^{1/2}/2,\end{aligned} \tag{2.17}$$

$$\begin{aligned}&\langle n+2,k\,|\bar{\Omega}(E,\theta)|n,k\rangle \\ &\quad = g(\omega^2-4)[(n+2)\,(n+1) \\ &\qquad \times(n+2k)\,(n+2k+1)]^{1/2}/4,\end{aligned} \tag{2.18}$$

$$\begin{aligned}&\langle n-2,k\,|\bar{\Omega}(E,\theta)|n,k\rangle \\ &\quad = g(\omega^2-4)[n(n-1) \\ &\qquad \times(n+2k-1)\,(n+2k-2)]^{1/2}/4,\end{aligned} \tag{2.19}$$

all other matrix elements being zero. From the formulas (2.15)–(2.19), we are going to build up a perturbation expansion for solving the eigenvalue problem (2.11).

III. PERTURBATION EXPANSION

A first approximation to the eigenvalues of the $\bar{\Omega}(E,\theta)$ operator can be obtained by considering only its diagonal terms (2.15). Hence, we calculate $E^{(0)}_{n,k}(\theta)$ on setting

$$\langle n,k\,|\bar{\Omega}(E,\theta)|n,k\rangle = 0, \tag{3.1}$$

and by solving this equation with respect to E. Although the exact $\bar{\Omega}(E,\theta)$ eigenvalues should be independent of θ (or equivalently of ω), $E^{(0)}_{n,k}(\theta)$ is clearly θ-dependent. A possible choice for θ that yields attractive results has been proposed by Feranchuk and Komarov,[22] i.e.,

$$\left[\frac{\partial E^{(0)}_{n,k}(\theta)}{\partial\theta}\right](\theta = \theta_{n,k}) = 0, \tag{3.2}$$

in which case θ will depend on the state considered.

Here, however, we shall correct the zeroth-order levels by treating the nondiagonal terms (2.16)–(2.19) as perturbation terms. Let us therefore notice that all matrix elements of $\bar{\Omega}(E,\theta)$ are linear with respect to E. Hence, we can write

$$\langle n\pm i,k\,|\bar{\Omega}(E,\theta)|n,k\rangle = a_{n\pm i,n} + b_{n\pm i,n}E \tag{3.3}$$
$$(n=0,1,2,\ldots,\quad i=0,1,2,),$$

whereby the coefficients a and b are easily obtained from (2.15)–(2.19). Since the $\bar{\Omega}(E,\theta)$ operator leaves the k-value invariant we can develop its nth eigenstate $|\bar{\psi}_{n,k}\rangle$, belonging to a particular k-value, in terms of the zeroth-order eigenfunctions $|i,k\rangle$, which are the SO (2,1) group state of the representation $\mathscr{D}^+(k)$; i.e.,

$$|\bar{\psi}_{n,k}\rangle = \sum_{i=0}^{\infty} c_i\,|i,k\rangle. \tag{3.4}$$

Substituting (3.4) into (2.11) we thus obtain, with the help of (3.3),

$$\sum_{j=0}^{\infty}\sum_{i=j-2}^{j+2}(a_{ji} + b_{ji}E_{n,k})c_i\,|j,k\rangle = 0. \tag{3.5}$$

Since the states $|j,k\rangle$ ($j=0,1,2,\ldots$) constitute a basis, each coefficient of $|j,k\rangle$ in (3.5) should vanish separately. For $j=n$ we solve the resulting secular equation with respect to $E_{n,k}$, which yields

$$E_{n,k} = -\frac{a_{nn}c_n + \Sigma_{i\neq n}a_{ni}c_i}{b_{nn}c_n + \Sigma_{i\neq n}b_{ni}c_i}, \tag{3.6}$$

whereas for $j\neq n$ we solve the secular equation with respect to c_j, yielding

$$c_j = -\frac{\Sigma_{i\neq j}a_{ji}c_i + \Sigma_{i\neq j}b_{ji}c_iE_{n,k}}{a_{jj} + b_{jj}E_{n,k}}\quad (j\neq n). \tag{3.7}$$

It is clear that (3.6) and (3.7) have the appropriate form to establish an iteration algorithm for the calculation of $E_{n,k}$. Indeed, denoting by $E^{(l)}_{n,k}$ and $c^{(l)}_j$ the lth order approximations of $E_{n,k}$ and c_j, respectively, we can prescribe the following Gauss–Seidel iterative scheme[23]:

$$\begin{aligned}E^{(l)}_{n,k} &= -\frac{a_{nn} + \Sigma_{i\neq n}a_{ni}c^{(l)}_i}{b_{nn} + \Sigma_{i\neq n}b_{ni}c^{(l)}_i}\quad (l>0),\\ c^{(l)}_j &= -\Big(\sum_{i<j}a_{ji}c^{(l)}_i + \sum_{i<j}b_{ji}c^{(l)}_iE^{(l-1)}_{n,k}\\ &\qquad + \sum_{i>j}a_{ji}c^{(l-1)}_i + \sum_{i>j}b_{ji}c^{(l-1)}_iE^{(l-1)}_{n,k}\Big)\\ &\qquad \times(a_{jj} + b_{jj}E^{(l-1)}_{n,k})^{-1}\quad (j\neq n)\quad (l>1),\\ c^{(l)}_n &= 1\quad (l>1),\end{aligned} \tag{3.8}$$

whereby the initial c-values are given by $c^{(0)}_i = \delta_{in}$.

For any θ-value for which the scheme (3.8) is convergent, we must necessarily find that $\lim_{l\to\infty} E^{(l)}_{n,k} = E_{n,k}$. Nevertheless, we can expect that the rate of convergence is effectively θ-dependent. This fact has been confirmed by our numerical treatment, which is discussed in Sec. IV.

Whenever we have obtained in this way an acceptable approximation of $E_{n,k}$, the corresponding eigenvector $|\psi_{n,k}\rangle$ is, on account of (2.9) and (3.4), determined by

$$|\psi_{n,k}\rangle = e^{i\theta K_2}\sum c_i\,|i,k\rangle. \tag{3.9}$$

Since

TABLE I. Energy eigenvalues for the $\lambda x^2/(1+gx^2)$ potential. The correspondent ω-value and the number of iterations required are indicated between squared brackets.

$g \setminus \lambda$	0.1	1	10	100
	1.043 713 044	1.380 531 800 938	3.250 261 220 414	9.976 180 087 723
	[1.650 10]	[1.281 10]	[0.585 10]	[0.205 10]
	3.120 081 864 016	4.079 883 011 687	9.619 066 412 190	29.781 191 110 777
	[1.673 10]	[1.283 10]	[0.600 10]	[0.201 10]
0.1	5.181 094 785 885	6.667 919 100 023	15.729 336 336 800	49.292 690 504 627
	[1.565 10]	[1.350 10]	[0.620 10]	[0.205 10]
	7.231 009 980 656	9.166 567 472 792	21.591 005 510 857	68.513 062 234 511
	[1.550 15]	[1.282 15]	[0.633 10]	[0.205 10]
	1.024 109 594 849	1.232 350 723 405	2.782 330 515 932	9.359 418 026 324
	[0.580 30]	[0.578 25]	[0.578 25]	[0.193 15]
	3.051 490 192 075	3.507 388 348 905	7.417 505 896 275	26.705 965 628 381
1	[0.651 30]	[0.626 30]	[0.505 30]	[0.235 20]
	5.058 963 280 762	5.589 778 933 736	10.701 025 575 410	41.441 099 751 485
	[0.822 40]	[0.762 35]	[0.554 30]	[0.235 20]
	7.064 886 134 163	7.648 201 241 723	13.388 323 494 238	53.839 093 264 554
	[1.128 60]	[1.089 50]	[0.779 35]	[0.261 25]
	1.005 942 881	1.059 296 881	1.580 022 327	5.793 942 3002
	[0.223 100]	[0.216 100]	[0.218 100]	[0.171 100]
10	3.008 810 926	3.088 090 846	3.879 036 830	11.572 196 776
	[0.251 120]	[0.251 120]	[0.258 120]	[0.226 120]

$$K_2 = -\frac{1}{2}i\left(x\frac{d}{dx}+\frac{1}{2}\right)$$

and

$$\exp\left(\theta x\frac{d}{dx}\right)f(x) = f(x\exp(\theta))$$

for an arbitrary function $f(x)$ (see Ref. 16), we immediately derive from (2.6) that in configuration space the eigenfunction $|\psi_{n,k}\rangle$ is approximated by

$$|\psi_{n,k}\rangle = \begin{cases} A\sum_i c_i u_{2i}\left(\sqrt{\frac{2}{\omega}}x\right), & \text{if } k=\frac{1}{4}, \\ A\sum_i c_i u_{2i+1}\left(\sqrt{\frac{2}{\omega}}x\right), & \text{if } K=\frac{3}{4}. \end{cases} \qquad (3.10)$$

In both cases A is a normalization factor resulting from the condition $\langle\psi_{n,k}|\psi_{n,k}\rangle = 1$.

IV. NUMERICAL RESULTS AND DISCUSSION

We have applied the iterative scheme expounded in Sec. III for the calculation of the energy values of the ground state and some of the first excited states for certain typical λ- and g-values. In every case we have derived an optimal value of the tilting parameter $\omega = \exp(-\theta)$ for which the rate of convergence is as high as possible. The number of iterations needed in order to achieve the required accuracy is essentially g-dependent. All these results are listed in Table I.

We notice from Table I that for the smallest g-value we already find the energy values with an accuracy of 12 significant figures within a ten-step process. To our knowledge, such a high degree of precision has, for the potential considered, never been obtained before by any other method. Certainly, for $g = 0.1$, this potential is nearly a pure harmonic oscillator potential of the form $(1+\lambda)x^2$. Hence, the eigenvalues should be close to $(2n+1)\,(1+\lambda)^{1/2}$ and the corresponding eigenvectors should not differ very much from the pure harmonic oscillator states $u_n((1+\lambda)^{1/2}x)$. This gives us a theoretical foundation for the fact that the optimal ω-value is found in the neighborhood of $(2/(1+\lambda))^{1/2}$. A final verification is obtained by inspection of the c-coefficients in the expansion (3.9) or (3.10). In Table II we list these coefficients for $\lambda = g = 0.1$ and for the ground state level $n = 0$, $k = \frac{1}{4}$. As expected, the most significant contribution comes from c_0, the coefficient of $u_0((2/\omega)^{1/2}x)$.

For $g \geqslant 1$, the argument can be inverted. Indeed, since then the potential differs more significantly from the pure oscillator one, we can expect that a good approximation to the exact eigenfunctions in the form of an expansion in appropriate harmonic oscillator eigenvectors should contain a rather large number of terms. Indeed, for $g = 1$ we find that on the average about 30 iterations are needed, whereas for $g = 10$ this number has already increased up to 100 in order to achieve an accuracy of ten significant figures. We therefore can conclude that the dynamic group approach combined with an iteration scheme produces extremely accurate results for g-values that are not too large. With increasing g, the method remains very useful, although the numerical efforts become comparable to the ones required by other techniques.

As a next comment we wish to draw attention to the fact that our method applies equally well to the case of negative λ-values. In order to verify the accuracy of our approximate

TABLE II. The c-coefficients occurring in (3.10) obtained with the tilting parameter $\omega = 1.65$ for the ground state $(n = 0, k = \frac{1}{4})$ of the $\lambda x^2/(1+gx^2)$ potential with $\lambda = g = 0.1$.

c_0	1.0	$E+00$
c_1	$-$ 5.445 65	$E-02$
c_2	4.419 31	$E-03$
c_3	$-$ 3.354 93	$E-04$
c_4	3.290 04	$E-05$
c_5	2.097 23	$E-06$
c_6	3.532 66	$E-07$
c_7	8.756 31	$E-09$

TABLE III. Eigenvalues for $\lambda = -0.42$ and $g = 0.1$ calculated with $\omega = 2$. The number of iterations required is indicated between squared brackets.

E_0	0.800 000 000 000	[2]
E_1	2.455 698 585 119	[11]
E_2	4.197 895 893 444	[13]
E_3	5.991 398 837 190	[14]
E_4	7.820 097 654 268	[16]
E_5	9.674 537 312 906	[18]

results we considered typical λ- and g-values for which the exact solutions and eigenvalues of the Schrödinger equation are known.[8] Again, the correspondence was found to be remarkably good. As an example, we give in Table III the calculated eigenvalues for $g = 0.1$, $\lambda = -4g - 2g^2 = -0.42$, a case for which it is known[8] that the lowest eigenvalue is exactly produced by the formula $E_0 = 1 - 2g$.

Finally, it should be remarked that the dynamic-group technique and the iterative scheme lend themselves very well and almost immediately to such a generalization that also three-dimensional problems can be treated.

[1]A. K. Mitra, J. Math. Phys. **19**, 2018 (1978).
[2]R. Kaushal, J. Phys. A **12**, L253 (1979).
[3]N. Bessis and G. Bessis, J. Math. Phys. **21**, 2780 (1980).
[4]J. Killingbeck, Comput. Phys. Commun. **18**, 211 (1979).
[5]S. Galicia and J. Killingbeck, Phys. Lett. A **71**, 17 (1979).
[6]C. Lai and H. Lin, J. Phys. A **15**, 1495 (1982).
[7]V. Fack and G. Vanden Berghe, J. Phys. A **18**, 3355 (1985).
[8]G. Flessas, Phys. Lett. A **83**, 121 (1981); J. Phys. A **15**, L97 (1982).
[9]V. Varma, J. Phys. A **14**, L489 (1981).
[10]R. Whitehead, A. Watt, G. Flessas, and M. Nagarajan, J. Phys. A **15**, 1217 (1982).
[11]H. Haken, "Laser Theory," in *Encyclopedia of Physics*, Vol. XXV/2c (Van Nostrand, Princeton, NJ, 1970).
[12]H. Risken and H. Vollmer, Z. Phys. **201**, 323 (1967).
[13]B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[14]M. Bednar, Ann. Phys. (NY) **75**, 305 (1973).
[15]A. Bechler, Ann. Phys. (NY) **108**, 49 (1977).
[16]C. C. Gerry and J. Laub, Phys. Rev. A **30**, 1229 (1984).
[17]B. Roy and R. Choudhury, Z. Naturforsch. A **40**, 453 (1985).
[18]C. C. Gerry and S. Silverman, Phys. Lett. A **95**, 481 (1983).
[19]A. O. Barut and W. Rasmussen, J. Phys. B **6**, 1695 (1973).
[20]F. M. Fernández, A. M. Mesón, and E. A. Castro, Phys. Lett. A **104**, 401 (1984).
[21]J. Cizek and J. Paldus, Int. J. Quantum Chem. **12**, 875 (1977).
[22]I. D. Feranchuk and L. I. Komarov, Phys. Lett. A **88**, 211 (1982).
[23]F. M. Fernández, A. M. Mesón, and E. A. Castro, J. Phys. A **18**, 1389 (1985).

Electromagnetic binding of a minimally interacting, relativistic spin-0 and spin-½ constituent: Zero four-momentum solutions

G. Bruce Mainland
Department of Physics, The Ohio State University, Columbus, Ohio 43210

(Received 3 September 1985; accepted for publication 3 December 1985)

Bound states of a composite system consisting of a charged spin-0 and a charged spin-½ constituent interacting via minimal electrodynamics are investigated using the Bethe–Salpeter equation in the ladder approximation (single photon exchange). Although the interaction involves derivative couplings, in the limit that the four-momentum is zero, it is possible to solve the integral equation by performing a Wick rotation and employing a method due to Fock. The eigenvalue spectrum of the coupling constant is found to be discrete.

I. INTRODUCTION

Since the determination of the value of the muon mass,[1] it has been known that the electron–muon mass ratio is numerically approximately two-thirds of the electromagnetic fine structure constant. If this is not a coincidence and charged leptons are composite, electromagnetism must play an important role in binding the constituents. However, no structure of leptons has been observed experimentally, so any constituents must be bound very tightly. The importance and strength of magnetic interactions in forming bound states at small distances was pointed out by Barut and his collaborators[2] and then used to construct composite models of leptons.[3] In that work it was assumed that stable particles are fundamental and that the unstable particles are constructed from them. However, the strong similarity of the electron, muon, and tau argues against the electron being fundamental while the other two are composite.

Taking the approach that all charged leptons are composite, electromagnetic models in two space dimensions were shown to have two desirable features.[4]

(1) All low-lying bound states have zero orbital angular momentum. If such states are comprised of a spin-0 and a spin-½ constituent, then they would all have spin-½ as do the observed leptons.

(2) The energy gaps between successively higher levels can increase, in qualitative agreement with the charged lepton mass spectrum.

We are thus motivated to consider a spin-0 and a spin-½ constituent interacting relativistically via minimal electrodynamics. To describe this system we use the Bethe–Salpeter equation[5] in the ladder approximation. For mathematical simplicity, here only single photon exchange is considered, and the masses of the constituents are taken to be equal.

We have been unable to separate the equation when the four-momentum of the bound state is timelike. By taking the four-momentum to be zero, the equation, which can be viewed as an eigenvalue equation for the coupling constant, is separated and solved.

The zero four-momentum solutions found here are interesting mathematically because they are obtained using a method due to Fock[6] and elaborated on by Lévy,[7] which is immediately applicable only to systems for which the interaction depends solely on the magnitude of the separation of the constituents. Minimal electromagnetic interaction of scalars, of course, involves derivative couplings that are of a more complicated form.

It is speculative to assume that the charged leptons are composite and even more speculative to assume that the neutrinos are composite. Nevertheless we remark that if a neutrino is both massless and composite, solving the bound-state problem yields a constraint between the mass ratio(s) of the constituents and the coupling constant(s). For two equal-mass scalars interacting via a massive scalar, it has been shown[8] that an eigenvalue of the Bethe–Salpeter equation in the ladder approximation for zero four-momentum is also an eigenvalue of the corresponding Bethe–Salpeter equation for a lightlike bound state. If the eigenvalues for zero four-momentum and lightlike solutions are also the same for this model, then a calculation similar to the one performed here could be meaningful physically if neutrinos are composite. Such a calculation would have to include the seagull term and allow for unequal masses of the constituents.

Zero four-momentum solutions have been obtained[9] for two equal-mass spinors interacting via a massless scalar and a continuum of solutions for the coupling constant is found. On the other hand, discrete eigenvalues are found[10,11] for two scalars interacting via a massless scalar. A model intermediate between the above two was considered by Sugano and Munakata,[12] who found zero four-momentum solutions for equal-mass spin-0 and spin-½ constituents interacting via a massless scalar. They determined that the eigenvalue spectrum of the coupling constant is discrete.

II. BETHE–SALPETER EQUATION IN THE LADDER APPROXIMATION

We consider a spin-0 field $\phi(x)$, which describes a particle with charge Q and mass m interacting via minimal electrodynamics with a spin-½ field $\psi(x)$, which describes a particle with charge q and the same mass m. The (renormalizable) Lagrangian is

$$L = :[(i\partial^{\mu} - QA^{\mu})\phi][(-i\partial_{\mu} - QA_{\mu})\phi^{\dagger}] - m^2\phi^{\dagger}\phi + \bar{\psi}\gamma_{\mu}(i\partial^{\mu} - qA^{\mu})\psi - m\bar{\psi}\psi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}:, \quad (2.1)$$

where $F_{\mu\nu} = \partial_{\nu}A_{\mu} - \partial_{\mu}A_{\nu}$.

The two-particle, Bethe–Salpeter wave function is defined to be

$$\chi_K(x_1,x_2) = \langle 0|T\psi(x_1)\phi(x_2)|K\rangle . \quad (2.2)$$

 0022-2488/86/051344-07$02.50

In Eq. (2.2) the symbol T represents time ordering and the letter K labels the four-momentum of the bound state. The center-of-mass coordinates X^{μ} are defined by

$$X^{\mu}=\xi x_1^{\mu}+(1-\xi)x_2^{\mu}, \tag{2.3}$$

and the relative coordinates x^{μ} by

$$x^{\mu}=x_1^{\mu}-x_2^{\mu}. \tag{2.4}$$

If we take the constant $\xi=m/(m+m)$, we get the usual nonrelativistic definition of center-of-mass coordinates but there is no reason to make a specific choice. The dependence of $\chi_K(x_1,x_2)$ on the center-of-mass coordinates factors with the result that $\chi_K(x_1,x_2)$ can be written as

$$\chi_K(x_1,x_2)=(2\pi)^{-3/2}e^{-iX^{\mu}K_{\mu}}\chi_K(x), \tag{2.5}$$

where $\chi_K(x)$ is given by

$$\chi_K(x)=(2\pi)^{3/2}\langle 0|T\psi[(1-\xi)x]\phi(-\xi x)|K\rangle. \tag{2.6}$$

Following standard procedures,[5] in the ladder approximation $\chi_K(x_1,x_2)$ satisfies[13]

$$(i\gamma^{\mu}\partial_{1\mu}-m)(\partial_2^{\mu}\partial_{2\mu}+m^2)\chi_K(x_1,x_2)$$
$$=qQ[\gamma^{\mu}D_{\mu\nu}(x_1-x_2)\partial_2^{\nu}+\partial_2^{\nu}\gamma^{\mu}D_{\mu\nu}(x_1-x_2)]\chi_K(x_1,x_2). \tag{2.7}$$

In the above equation $D_{\mu\nu}(x)$ is the photon propagator,

$$D_{\mu\nu}(x)=-\int\frac{d^4q}{(2\pi)^4}\frac{e^{-iq^{\rho}x_{\rho}}g_{\mu\nu}}{q^{\delta}q_{\delta}+i\epsilon}. \tag{2.8}$$

To obtain a convenient form for (2.7), we rewrite it in terms of center-of-mass coordinates X^{μ} and relative coordinates x^{μ} and use (2.5). Setting the four-momentum $K^{\mu}=0$,

$$(i\gamma^{\mu}\partial_{\mu}-m)(\partial^{\mu}\partial_{\mu}+m^2)\chi_0(x)$$
$$=qQ\int\frac{d^4q}{(2\pi)^4}\frac{e^{-iq^{\rho}x_{\rho}}}{q^{\delta}q_{\delta}+i\epsilon}[2\gamma_{\nu}\partial^{\nu}-iq^{\mu}\gamma_{\mu}]\chi_0(x). \tag{2.9}$$

In (2.9) we have replaced $\chi_K(x)$ by $\chi_0(x)$ to indicate that the four-momentum K^{μ} has been set to zero. Fourier transforming and defining

$$\chi_0(p)=\frac{1}{(2\pi)^2}\int d^4x\, e^{ip^{\rho}x_{\rho}}\chi_0(x), \tag{2.10}$$

(2.9) becomes

$$(\gamma^{\mu}p_{\mu}-m)(-p^{\mu}p_{\mu}+m^2)\chi_0(p)$$
$$=\frac{-iqQ}{(2\pi)^4}\int\frac{d^4q}{(p-q)^2+i\epsilon}\gamma^{\nu}(q_{\nu}+p_{\nu})\chi_0(q). \tag{2.11}$$

To put (2.11) in final form before solving it, we use the analytic properties of the wave function and continue the equation into Euclidean space *à la* Wick[10] with the result

$$D\tilde{\chi}_0(p)\equiv(\tilde{\gamma}\cdot p+m)(p\cdot p+m^2)\tilde{\chi}_0(p)$$
$$+\frac{qQ}{(2\pi)^4}\int\frac{d^4q}{(p-q)\cdot(p-q)}\tilde{\gamma}\cdot(p+q)\tilde{\chi}_0(q)$$
$$=0, \tag{2.12}$$

where $\tilde{\chi}_0(p)\equiv\chi_0(ip^0,\mathbf{p})$, the Euclidean scalar product $p\cdot p\equiv p^0p^0+\mathbf{p}\cdot\mathbf{p}$, and $\tilde{\gamma}\cdot p\equiv\tilde{\gamma}^0p^0+\tilde{\gamma}^ip^i$. The matrices $\tilde{\gamma}^{\mu}$ are given by $\tilde{\gamma}^0\equiv-i\gamma_0$, $\tilde{\gamma}^i=\gamma^i$, where the matrices γ^{μ} are those given in Ref. 13.

III. SEPARATION OF THE BETHE–SALPETER EQUATION IN EUCLIDEAN SPACE

To separate the equation we note that it is rotationally invariant and therefore the operator D defined in (2.12) commutes with the Euclidean angular momentum operators

$$J^{\mu\nu}=-ip^{\mu}\frac{\partial}{\partial p^{\nu}}+ip^{\nu}\frac{\partial}{\partial p^{\mu}}+\frac{i}{4}[\tilde{\gamma}^{\mu},\tilde{\gamma}^{\nu}]. \tag{3.1}$$

(In Euclidean space we make no distinction between covariant and contravariant vectors.) It is straightforward to show that the following three operators commute with D and with each other:

$$\tfrac{1}{2}J^{\mu\nu}J^{\mu\nu},\quad K\equiv\gamma^0(\tilde{\sigma}^iL^i+1),\ J^{12}, \tag{3.2}$$

where

$$\tilde{\sigma}^i=\begin{pmatrix}\sigma^i & 0\\ 0 & \sigma^i\end{pmatrix}$$

and

$$L^i=-i\epsilon^{ijk}p^j\frac{\partial}{\partial p^k}.$$

The σ^i are 2×2 Pauli spin matrices, Greek letters range from 0 to 3, and Roman letters range from 1 to 3. The eigenfunctions $\tilde{\chi}_0$ thus can be chosen to be simultaneous eigenfunctions of the three operators listed in (3.2).

Simultaneous eigenfunctions of K and J^{12} are well known[14] and are constructed from $\phi^{(\pm)}(\theta_3,\phi)$ given below. Introducing the polar coordinates

$$\begin{aligned}p^0&=|p|\cos\theta_2,\\ p^3&=|p|\sin\theta_2\cos\theta_3,\\ p^1&=|p|\sin\theta_2\sin\theta_3\cos\phi,\\ p^2&=|p|\sin\theta_2\sin\theta_3\sin\phi,\end{aligned} \tag{3.3}$$

we have

$$\phi_{j,m}^{(+)}(\theta_3,\phi)=\begin{bmatrix}\sqrt{\dfrac{j+m}{2j}} & Y_{j-1/2}^{m-1/2}(\theta_3,\phi)\\ \sqrt{\dfrac{j-m}{2j}} & Y_{j-1/2}^{m+1/2}(\theta_3,\phi)\end{bmatrix},$$
$$j=l+\tfrac{1}{2},\quad l=0,1,2,..., \tag{3.4a}$$

and

$$\phi_{j,m}^{(-)}(\theta_3,\phi)=\begin{bmatrix}\sqrt{\dfrac{j+1-m}{2(j+1)}} & Y_{j+1/2}^{m-1/2}(\theta_3,\phi)\\ -\sqrt{\dfrac{j+1+m}{2(j+1)}} & Y_{j+1/2}^{m+1/2}(\theta_3,\phi)\end{bmatrix},$$
$$j=l-\tfrac{1}{2},\quad l=1,2,... . \tag{3.4b}$$

The spherical harmonics $Y_l^m(\theta_3,\phi)$ are written with the convention $(Y_l^m)^*=(-1)^mY_l^{-m}$ and $\phi_{j,m}^{(\pm)}(\theta_3,\phi)$ satisfy

$$(L^3+\tfrac{1}{2}\sigma^3)\phi_{j,m}^{(\pm)}(\theta_3,\phi)=m\phi_{j,m}^{(\pm)}(\theta_3,\phi)\,, \quad (3.5a)$$

$$\sigma^i L^i \phi_{j,m}^{(\pm)}(\theta_3,\phi) = -[1\mp(j+\tfrac{1}{2})]\phi_{j,m}^{(\pm)}(\theta_3,\phi)\,. \quad (3.5b)$$

The two-component eigenfunctions $\phi_{j,m}^{(\pm)}(\theta_3,\phi)$ can be transformed into each other using the relationship

$$\phi_{j,m}^{(\pm)}(\theta_3,\phi)=\frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{|\mathbf{p}|}\phi_{j,m}^{(\mp)}(\theta_3,\phi)\,. \quad (3.6)$$

Using (3.5) the function

$$f^{(\pm)}(|p|)\begin{bmatrix}\xi^{(\pm)}(\theta_2)\phi_{j,m}^{(\pm)}(\theta_3,\phi)\\ \eta^{(\mp)}(\theta_2)\phi_{j,m}^{(\mp)}(\theta_3,\phi)\end{bmatrix} \quad (3.7)$$

is seen to be an eigenstate of both K and J^{12} with respective eigenvalues $\pm(j+\tfrac{1}{2})$ and m.

To determine $\xi^{(\pm)}(\theta_2)$ and $\eta^{(\pm)}(\theta_2)$ such that (3.7) is also an eigenstate of $\tfrac{1}{2}J^{\mu\nu}J^{\mu\nu}$, it is convenient to introduce the operator

$$A\equiv\gamma^i\gamma^0L^{i0}+\tilde{\sigma}^iL^i+\tfrac{3}{2}\,, \quad (3.8)$$

where

$$L^{i0}=-ip^i\frac{\partial}{\partial p^0}+ip^0\frac{\partial}{\partial p^i}\,. \quad (3.9)$$

Since the operator A satisfies

$$\tfrac{1}{2}J^{\mu\nu}J^{\mu\nu}=A^2-\tfrac{9}{4}\,,$$

and the eigenvalues of $\tfrac{1}{2}J^{\mu\nu}J^{\mu\nu}$ are[15] $k_1^2+2k_1+\tfrac{1}{4}$, where $k_1=\tfrac{1}{2},\tfrac{3}{2},\tfrac{5}{2},\ldots$, the eigenvalues a of the operator A are

$$a=\pm(k_1+1)=\pm\tfrac{3}{2},\pm\tfrac{5}{2},\ldots\,. \quad (3.10)$$

Thus $\tilde{\chi}_0$ is a simultaneous eigenstate of the operators (3.2) if it is written in the form

$$\tilde{\chi}_0=f^{(\pm)}(|p|)\psi_1^{(\pm)}(\theta_2,\theta_3,\phi)+g^{(\pm)}(|p|)\psi_2^{(\pm)}(\theta_2,\theta_3,\phi)\,, \quad (3.11)$$

where

$$\psi_1^{(\pm)}(\theta_2,\theta_3,\phi)\equiv\begin{bmatrix}\xi^{(\pm)}_{k_1+1}(\theta_2)\phi_{j,m}^{(\pm)}(\theta_3,\phi)\\ \eta^{(\mp)}_{k_1+1}(\theta_2)\phi_{j,m}^{(\mp)}(\theta_3,\phi)\end{bmatrix}, \quad (3.12a)$$

and

$$\psi_2^{(\pm)}(\theta_2,\theta_3,\phi)\equiv\begin{bmatrix}\xi^{(\pm)}_{-(k_1+1)}(\theta_2)\phi_{j,m}^{(\pm)}(\theta_3,\phi)\\ \eta^{(\pm)}_{-(k_1+1)}(\theta_2)\phi_{j,m}^{(\mp)}(\theta_3,\phi)\end{bmatrix}. \quad (3.12b)$$

In (3.11) either the top signs or the bottom signs must be taken and $\psi_1^{(\pm)}$ and $\psi_2^{(\pm)}$, respectively, satisfy the equations

$$A\psi_1^{(\pm)}=(k_1+1)\psi_1^{(\pm)}\,, \quad (3.13a)$$

$$A\psi_2^{(\pm)}=-(k_1+1)\psi_2^{(\pm)}\,. \quad (3.13b)$$

Substituting (3.12) into (3.13) and solving the resulting differential equations yield the following solutions:

$$\psi_1^{(+)}=\begin{bmatrix}(k_1+j+1)P^{(2)}_{k_1-1/2,j-1/2}(\theta_2)\phi_{j,m}^{(+)}(\theta_3,\phi)\\ iP^{(2)}_{k_1-1/2,j+1/2}(\theta_2)\phi_{j,m}^{(-)}(\theta_3,\phi)\end{bmatrix}, \quad (3.14a)$$

$$\psi_2^{(+)}=\begin{bmatrix}i(k_1-j+1)\,P^{(2)}_{k_1+1/2,j-1/2}(\theta_2)\phi_{j,m}^{(+)}(\theta_3,\phi)\\ P^{(2)}_{k_1+1/2,j+1/2}(\theta_2)\phi_{j,m}^{(-)}(\theta_3,\phi)\end{bmatrix}, \quad (3.14b)$$

$$\psi_1^{(-)}=\gamma_5\psi_1^{(+)}\,, \quad (3.14c)$$

$$\psi_2^{(-)}=\gamma_5\psi_2^{(+)}\,. \quad (3.14d)$$

The eigenstates (3.14a) and (3.14b) are essentially those used in Ref. 12 to solve the corresponding problem in scalar electrodynamics. Note that $\psi_1^{(\pm)}$ and $\psi_2^{(\pm)}$ are four-component column "vectors," each component of which is a hyperspherical harmonic in a four-dimensional space.

Using (A17), (A18), (A21), and (A22), $\psi_1^{(\pm)}$ and $\psi_2^{(\pm)}$ are found to obey the useful relationships

$$\tilde{\gamma}\cdot p\psi_1^{(+)}=-|p|\psi_2^{(+)}\,, \quad (3.15a)$$

$$\tilde{\gamma}\cdot p\psi_2^{(+)}=|p|\psi_1^{(+)}\,. \quad (3.15b)$$

IV. SOLUTION OF THE BETHE–SALPETER EQUATION

To solve the Bethe–Salpeter equation, we use the method of Fock[6] and project four-dimensional momentum space onto the surface of a five-dimensional hypersphere with the transformation

$$|p|=m\tan(\theta_1/2)\,. \quad (4.1)$$

The factor 2 is included in the above formula because the range of θ_1 on a hypersphere must be $0<\theta_1<\pi$ [see (A2)], so, as θ_1 varies over this range, $|p|$ varies from 0 to ∞ as required. Defining the four-vector q in analogy with (3.3) and (4.1), except that the angles are denoted by primes, we get

$$d^4q=[m^4/16\cos^8(\theta_1'/2)]\times\sin^3\theta_1'\,\sin^2\theta_2'\,\sin\theta_3'd\theta_1'\,d\theta_2'\,d\theta_3'\,d\phi'\,. \quad (4.2)$$

The components of the unit vector $\hat{u}$ in five dimensions are

$$\hat{u}=(\cos\theta_1,\ \sin\theta_1\cos\theta_2,\ \sin\theta_1\sin\theta_2\cos\theta_3,\ \sin\theta_1\sin\theta_2\sin\theta_3\cos\phi,\ \sin\theta_1\sin\theta_2\sin\theta_3\sin\phi)\,. \quad (4.3)$$

Using a corresponding expression for the unit vector $\hat{v}$ in terms of primed angles, it is straightforward to show that

$$(p-q)\cdot(p-q)=\frac{m^2}{\cos^2(\theta_1/2)\cos^2(\theta_1'/2)}\times\tfrac{1}{2}[1-\cos\Theta]\,. \quad (4.4)$$

In (4.4), Θ is the angle between the unit vectors $\hat{u}$ and $\hat{v}$.

With (4.2) and (4.4), the Bethe–Salpeter equation (2.12) can be written in the form

$$(\tilde{\gamma}\cdot p+m)(p\cdot p+m^2)\tilde{\chi}_0(p)=-\frac{qQ}{(2\pi)^4}\int\frac{m^4}{16\cos^8(\theta_1'/2)}\times\frac{\cos^2(\theta_1/2)\cos^2(\theta_1'/2)}{(m^2/2)(1-\cos\Theta)}\times\tilde{\gamma}\cdot(q+p)\tilde{\chi}_0(q)\,d\Omega_{(4)}'\,. \quad (4.5)$$

Taking $\tilde{\chi}_0$ to be of the form (3.11) (and initially considering the case where the top signs are taken), (4.5) becomes

$$[m^2/\cos^2(\theta_1/2)][-m\tan(\theta_1/2)f^{(+)}(\theta_1)\psi_2^{(+)} + m\tan(\theta_1/2)g^{(+)}(\theta_1)\psi_1^{(+)} + mf^{(+)}(\theta_1)\psi_1^{(+)} + mg^{(+)}(\theta_1)\psi_2^{(+)}]$$
$$= -\frac{qQ}{(2\pi)^4}\frac{m^2}{8}\cos^2\frac{\theta_1}{2}\int\frac{d\Omega'_{(4)}}{1-\cos\Theta}\frac{\cos^2(\theta'_1/2)}{\cos^8(\theta'_1/2)} \times\tilde{\gamma}\cdot(q+p)[f^{(+)}(\theta'_1)\psi_1^{(+)} + g^{(+)}(\theta'_1)(\psi_2^{(+)}]. \quad (4.6)$$

If $f^{(+)}(\theta_1)$ and $g^{(+)}(\theta_1)$ are assumed to be of the form

$$\frac{f^{(+)}(\theta_1)}{\cos^8(\theta_1/2)} = \sum_{n=0}^{\infty}(2n+2\nu+3)f_n P^{(3)}_{n+\nu,k_1-1/2}(\theta_1), \quad (4.7a)$$

$$\frac{g^{(+)}(\theta_1)}{\cos^8(\theta_1/2)} = \sum_{n=0}^{\infty}(2n+2\nu+3)g_n P^{(3)}_{n+\nu,k_1+1/2}(\theta_1), \quad (4.7b)$$

where f_n and g_n are constants, (4.6) can be written

$$m^3[\tfrac{1}{2}(1+\cos\theta_1)]^2\tfrac{1}{2}\sin\theta_1 \times\Big\{-\sum_{n=0}^{\infty}(2n+2\nu+3)f_n P^{(3)}_{n+\nu,k_1-1/2}(\theta_1)\psi_2^{(+)} + \sum_{n=0}^{\infty}(2n+2\nu+3)g_n P^{(3)}_{n+\nu,k_1+1/2}(\theta_1)\psi_1^{(+)}\Big\}$$
$$+ m^3[\tfrac{1}{2}(1+\cos\theta_1)]^3 \times\Big\{\sum_{n=0}^{\infty}(2n+2\nu+3)f_n P^{(3)}_{n+\nu,k_1-1/2}(\theta_1)\psi_1^{(+)} + \sum_{n=0}^{\infty}(2n+2\nu+3)g_n P^{(3)}_{n+\nu,k_1+1/2}(\theta_1)\psi_2^{(+)}\Big\}$$
$$= -[qQ/(2\pi)^4](m^2/8)(I_1+I_2). \quad (4.8)$$

In (4.8),

$$I_1 = \cos^2\frac{\theta_1}{2}\int\frac{d\Omega'_{(4)}}{1-\cos\Theta}\cos^2\frac{\theta'_1}{2}\tilde{\gamma}\cdot q \times\Big[\sum_{n=0}^{\infty}(2n+2\nu+3)f_n P^{(3)}_{n+\nu,k_1-1/2}(\theta'_1)\psi_1^{(+)} + \sum_{n=0}^{\infty}(2n+2\nu+3)g_n P^{(3)}_{n+\nu,k_1+1/2}(\theta'_1)\psi_2^{(+)}\Big], \quad (4.9)$$

and

$$I_2 = \cos^2\frac{\theta_1}{2}\tilde{\gamma}\cdot p\int\frac{d\Omega'_{(4)}}{1-\cos\Theta}\cos^2\frac{\theta'_1}{2} \times\Big[\sum_{n=0}^{\infty}(2n+2\nu+3)f_n P^{(3)}_{n+\nu,k_1-1/2}(\theta'_1)\psi_1^{(+)} + \sum_{n=0}^{\infty}(2n+2\nu+3)g_n P^{(3)}_{n+\nu,k_1+1/2}(\theta'_1)\psi_2^{(+)}\Big]. \quad (4.10)$$

Using the identities in the Appendix, the left-hand side of (4.8) can be seen to be an infinite sum of hyperspherical harmonics of the form

$$\sum_{n'} P^{(3)}_{n',k_1-1/2}(\theta_1)\psi_1^{(+)} \text{ and } \sum_{n'} P^{(3)}_{n',k_1+1/2}(\theta_1)\psi_2^{(+)}.$$

When $\sin\theta_1$ multiplies a term of the form $P^{(3)}_{n',k-1/2}(\theta_1)$ it can raise the second index by an integer [see (A24)], and when $\sin\theta_1$ multiplies a term of the form $P^{(3)}_{n+\nu,k+1/2}(\theta_1)$ it can lower the second index by an integer [see (A25)]. Multiplication by $\cos\theta_1$ does not affect the second index [see (A23)]. Thus we reach the crucial conclusion that when $\tilde{\gamma}\cdot p$, or any integer power of $\cos\theta_1$, multiplies a hyperspherical harmonic of the form being used here, it is possible to rewrite the product as a sum of hyperspherical harmonics. Rather than write out the lengthy expression for the nth term of the series for the left-hand side (lhs) of (4.8), we only write the first few, which are sufficient to determine the eigenvalue spectrum of the coupling constant.

lhs of (4.8)

$$= \frac{m^3}{8}\frac{(\nu+k_1+\frac{3}{2})(\nu+k_1+k_1+\frac{1}{2})}{(2\nu+1)(2\nu-1)} \times\Big[f_0+g_0\Big(\nu+k_1+\frac{5}{2}\Big)\Big][P^{(3)}_{\nu-3,k_1+1/2}\psi_2^{(+)} + (\nu+k_1-\tfrac{1}{2})P^{(3)}_{\nu-3,k_1-1/2}\psi_1^{(+)}]$$
$$+\frac{m^3}{8}\frac{\nu+k_1+\frac{3}{2}}{2\nu+1}\Big\{2f_0+3\Big(\nu+k_1+\frac{5}{2}\Big)g_0 + \frac{(\nu+k_1+\frac{5}{2})}{2\nu+3}\Big[f_1+\Big(\nu+k_1+\frac{7}{2}\Big)g_1\Big]\Big\} \times P^{(3)}_{\nu-2,k_1+1/2}\psi_2^{(+)}$$
$$+\frac{m^3}{8}\frac{(\nu+k_1+\frac{3}{2})(\nu+k_1+\frac{1}{2})}{2\nu+1} \times\Big\{2g_0\Big(\nu+k_1+\frac{5}{2}\Big)+3f_0 + \frac{(\nu+k_1+\frac{5}{2})}{2\nu+3}\Big[f_1+\Big(\nu+k_1+\frac{7}{2}\Big)g_1\Big]\Big\} \times P^{(3)}_{\nu-2,k_1-1/2}\psi_1^{(+)}+\cdots. \quad (4.11)$$

Turning our attention to the integral I_1 on the right-hand side of the Bethe–Salpeter equation (4.8), we note from the preceding discussion that after multiplication of the infinite series of hyperspherical harmonics by $\cos^2(\theta'_1/2) = (1+\cos\theta'_1)/2$ and $\tilde{\gamma}\cdot q$, the series can be rewritten in terms of hyperspherical harmonics, which then can be integrated using Hecke's theorem[16] [see (A9) and (A10)]. Using (3.15), (4.1), and the identity

$$\sin(\theta'_1/2)\cos(\theta'_1/2) = (\sin\theta'_1)/2,$$

the expression (4.9) for I_1 becomes

$$I_1 = \frac{m}{2}\cos^2\frac{\theta_1}{2}\int\frac{d\Omega'_{(4)}}{1-\cos\Theta}\Big[\sum_{n=0}^{\infty}-(2n+2\nu+3)f_n\sin\theta'_1 P^{(3)}_{n+\nu,k_1-1/2}(\theta'_1)\psi_2^{(+)} + \sum_{n=0}^{\infty}(2n+2\nu+3)g_n\sin\theta'_1 P^{(3)}_{n+\nu,k_1+1/2}(\theta'_1)\psi_1^{(+)}\Big]. \quad (4.12)$$

From (A24) and (A25),

$$I_1 = \frac{m}{2} \cos^2 \frac{\theta_1}{2} \int \frac{d\Omega'_{(4)}}{1 - \cos \Theta}$$
$$\times \Big\{ \sum_{n=0}^{\infty} -f_n [P^{(3)}_{n+\nu+1,k_1+1/2}(\theta'_1) - P^{(3)}_{n+\nu-1,k_1+1/2}(\theta'_1)] \psi_2^{(+)}$$
$$+ \sum_{n=0}^{\infty} g_n \Big[\Big(\nu + k_1 + n + \frac{5}{2}\Big)\Big(\nu + k_1 + n + \frac{3}{2}\Big) \times P^{(3)}_{n+\nu-1,k_1-1/2}(\theta'_1)$$
$$- \Big(\nu - k_1 + n + \frac{3}{2}\Big)\Big(\nu - k_1 + n + \frac{1}{2}\Big) \times P^{(3)}_{n+\nu+1,k_1-1/2}(\theta'_1) \Big] \psi_1^{(+)} \Big\}. \quad (4.13)$$

All the terms in the curly brackets are hyperspherical harmonics with the result that the integral can be performed using Hecke's theorem.[16] From (A9),

$$I_1 = \frac{m}{2} \cos^2 \frac{\theta_1}{2} \sum_{n=0}^{\infty} \{ f_n [-\Lambda_{n+\nu+1} P^{(3)}_{n+\nu+1,k_1+1/2} + \Lambda_{n+\nu-1} P^{(3)}_{n+\nu-1,k_1+1/2}] \psi_2^{(+)}$$
$$+ g_n [\Lambda_{n+\nu-1} (\nu + k_1 + n + \tfrac{5}{2}) \times (\nu + k_1 + n + \tfrac{3}{2}) P^{(3)}_{n+\nu-1,k_1-1/2}$$
$$- \Lambda_{n+\nu+1} (\nu - k_1 + n + \tfrac{3}{2}) \times (\nu - k_1 + n + \tfrac{1}{2}) P^{(3)}_{n+\nu+1,k_1-1/2}] \psi_1^{(+)} \}, \quad (4.14)$$

where, from (A10),

$$\Lambda_\mu = \frac{4\pi^2}{(\mu+2)(\mu+1)} \int_{-1}^{1} (1+x) C_\mu^{3/2}(x)\, dx . \quad (4.15)$$

The integral in (4.15) is easily evaluated using the procedure discussed in the Appendix and is found to have a value of 2. Thus

$$\Lambda_\mu = 8\pi^2/(\mu+2)(\mu+1) . \quad (4.16)$$

The integral I_2 as given by (4.10) is evaluated in a similar manner. Some simplification occurs when the integrals I_1 and I_2 are added, but the nth term is still lengthy. Since the first two terms in the infinite series for $I_1 + I_2$ determine the spectrum of the coupling constant, only those are given:

$$I_1 + I_2 = \frac{m}{2} \frac{(\nu + k_1 + \frac{3}{2})\Lambda_{\nu-1}}{2\nu+1} \times [f_0 P^{(3)}_{\nu-2,k_1+1/2} \psi_2^{(+)} + g_0 (\nu + k_1 + \tfrac{5}{2}) \times (\nu + k_1 + \tfrac{1}{2}) P^{(3)}_{\nu-2,k_1-1/2} \psi_1^{(+)}] + \cdots . \quad (4.17)$$

The spectrum of the integer ν and the coupling constant $qQ/4\pi$ are found by requiring coefficients of $P^{(3)}_{n,k_1+1/2}\psi_2^{(+)}$ and $P^{(3)}_{n,k_1-1/2}\psi_1^{(+)}$ vanish in (4.8). Using (4.11) and (4.17), and requiring that the coefficients of $P^{(3)}_{\nu-3,k_1+1/2}\psi_2^{(+)}$ and $P^{(3)}_{\nu-3,k_1-1/2}\psi_1^{(+)}$ vanish, yields the condition

$$f_0 = -g_0 (\nu + k_1 + \tfrac{5}{2}) . \quad (4.18)$$

But since f_0 and g_0 are assumed to be nonzero and the identities used are all valid only if $P^{(3)}_{p,r}$ is nonzero, we also must require $P^{(3)}_{\nu-3,k_1\pm 1/2} \neq 0$. From (A7) and (A8) these two conditions are satisfied, provided $k_1 + 1/2 < \nu - 3$ or

$$\nu \geq k_1 + \tfrac{7}{2} = 4,5,6,\ldots . \quad (4.19)$$

Requiring that the coefficients of $P^{(3)}_{\nu-2,k_1\pm 1/2}$ vanish yields the condition

$$qQ/4\pi = \pi\nu(\nu+1) , \quad (4.20)$$

plus an equation involving the expansion coefficients f_0, g_0, f_1, and g_1. Requiring that the coefficients of $P^{(3)}_{\nu-2+n,k\pm 1/2}$, $n > 0$, vanish yields only equations between the f_i and g_i. The result [(4.19) and (4.20)] is also obtained using the parity conjugate solutions $\psi_1^{(-)}$ and $\psi_2^{(-)}$ given in (3.14c) and (3.14d).

For the zero four-momentum bound states, it is very surprising that the product of the charges of the constituents is positive, implying the constituents have the same sign charge. Apparently this effect is caused by the compling constant $qQ/4\pi$ being much greater than unity. As a consequence the effect is not necessarily physically meaningful.

A calculation that includes the seagull term and considers constituents of unequal mass is in progress.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor David M. Scott and Professor Shimshon Zimering for helpful conversations and Mr. Russell P. Schussler for performing a computer calculation.

APPENDIX: SUMMARY OF MATHEMATICAL RESULTS FOR HYPERSPHERICAL HARMONICS

The mathematical formulas for hyperspherical harmonics required for solving the Bethe–Salpeter equation are given here. Only the recursion relations and formulas for differentiation of $P^{(s)}_{p,r}$ are original, but it is convenient to have the relevant formulas summarized in one place, especially since Ref. 17 was published in 1926 and is difficult to locate.

We consider a $(q+2)$-dimensional space in which polar coordinates are defined by the relations[17]

$$\begin{aligned} x_1 &= r \cos \theta_1 , \\ x_2 &= r \sin \theta_1 \cos \theta_2 , \\ &\vdots \\ x_q &= r \sin \theta_1 \sin \theta_2 \cdots \cos \theta_q , \\ x_{q+1} &= r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_q \cos \phi , \\ x_{q+2} &= r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_q \sin \phi , \end{aligned} \quad (A1)$$

where

$$0 \leq \theta_i \leq \pi , \quad 0 \leq \phi \leq 2\pi . \quad (A2)$$

The Jacobian J of the transformation from Cartesian coordinates x_i to polar coordinates is[17]

$$J\, d\theta_1\, d\theta_2 \cdots d\theta_q\, d\phi = r^{q+1} d\Omega_{(q+1)} , \quad (A3)$$

where

$$d\Omega_{(q+1)} = \sin^q\theta_1 \sin^{q-1}\theta_2 \cdots \sin\theta_q \, d\theta_1 \, d\theta_2 \cdots d\theta_q \, d\phi . \tag{A4}$$

Hyperspherical harmonics $Y_\mu(\theta_1,...,\theta_q,\phi)$ (see Ref. 17) satisfy the equation

$$0 = \sum_{i=1}^{q+2} \frac{\partial^2}{\partial x_i^2} r^\mu Y_\mu , \tag{A5}$$

which, in terms of polar coordinates, becomes

$$0 = \mu(\mu+q)Y_\mu + \frac{1}{\sin^2\theta_1 \cdots \sin^2\theta_q} \frac{\partial^2 Y_\mu}{\partial\phi^2} + \sum_{j=1}^{q} \frac{1}{\sin^2\theta_1 \cdots \sin^2\theta_{j-1}} \frac{1}{\sin^{q+1-j}\theta_j} \times \frac{\partial}{\partial\theta_j}\left(\sin^{q+1-j}\theta_j \frac{\partial Y_\mu}{\partial\theta_j}\right). \tag{A6}$$

The solution to (A6) is[17]

$$Y_\mu = P^{(q)}_{\mu,p_1}(\cos\theta_1) P^{(q-1)}_{p_1,p_2}(\cos\theta_2) \cdots P^{(1)}_{p_{q-1},p_q}(\cos\theta_q) e^{\pm i p_q \phi} , \tag{A7a}$$

where the integers $p_1, p_2,..., p_q$ satisfy the conditions

$$\mu \geqslant p_1 \geqslant p_2 \geqslant \cdots \geqslant p_{q-1} \geqslant p_q \geqslant 0. \tag{A7b}$$

The functions $P^{(s)}_{p,r}$ are given by[17]

$$P^{(s)}_{p,r}(z) = (1-z^2)^{r/2} \frac{d^r}{dz^r} C_p^{s/2}(z) , \quad r \text{ integer} \leqslant p , \tag{A8}$$

where C_p^s is a Gegenbauer polynomial.

The theorem of Hecke[16] that is used to solve the Bethe–Salpeter equation is as follows[7]: In a $(q+2)$-dimensional space let the unit vectors $\hat{u}$ and $\hat{v}$ be expressed, respectively, in terms of the polar coordinates $\theta_1,\theta_2,...,\theta_q$, ϕ and θ_1' $\theta_2',...,\theta_q',\phi'$. Let $F(\Theta)$ be a function depending only on the angle Θ between the two vectors. Then

$$\Lambda_\mu Y_\mu(\theta_1,\theta_2,...,\theta_q,\phi) = \int_{\substack{\text{unit}\\ \text{hypersphere}}} F(\cos\Theta) \, Y_\mu(\theta_1',\theta_2',...,\theta_q',\phi') \, d\Omega'_{(q+1)} , \tag{A9}$$

where

$$\Lambda_\mu = \frac{2\pi^{(1/2)(q+1)}}{\Gamma[\frac{1}{2}(q+1)]} \frac{\mu!(q-1)!}{(\mu+q-1)!} \times \int_{-1}^{1} F(x) C_\mu^{q/2}(x)(1-x^2)^{(1/2)(q-1)} \, dx . \tag{A10}$$

The integral

$$I_\mu = \int_{-1}^{1} F(x) C_\mu^{q/2}(x)(1-x^2)^{(1/2)(q-1)} \, dx \tag{A11}$$

in (A10) can be evaluated[11] by using the generating function for Gegenbauer polynomials,[18]

$$(1-2xr+r^2)^{-q/2} = \sum_{\mu=0}^{\infty} C_\mu^{q/2}(x) r^\mu . \tag{A12}$$

Multiplying (A11) by r^μ, summing over μ, and using (A12),

$$\sum_{\mu=0}^{\infty} r^\mu I_\mu = \int_{-1}^{1} F(x)(1-2xr+r^2)^{-q/2}(1-x^2)^{(1/2)(q-1)} \, dx . \tag{A13}$$

The integral on the right-hand side of (A13) can be carried out for the desired value of q and function $F(x)$. Then I_μ can be found by comparing coefficients of r^μ on the left- and right-hand sides of the equation.

Beginning with the differential equation for Gegenbauer polynomials[18] and using the definition (A8), $P^{(s)}_{p,r}(z)$ is found to satisfy the second-order differential equation

$$0 = (1-z^2)\frac{d^2}{dz^2} P^{(s)}_{p,r}(z) - (s+1)z \frac{d}{dz} P^{(s)}_{p,r}(z) + p(p+s) P^{(s)}_{p,r}(z) - \frac{r(r+s-1)}{1-z^2} P^{(s)}_{p,r}(z) . \tag{A14}$$

With the help of (A8) and (A14), the two following differentiation formulas are obtained:

$$\sqrt{1-z^2}\frac{d}{dz} P^{(s)}_{p,r}(z) = \frac{-rz}{\sqrt{1-z^2}} P^{(s)}_{p,r}(z) + P^{(s)}_{p,r+1}(z) , \tag{A15}$$

$$\sqrt{1-z^2}\frac{d}{dz} P^{(s)}_{p,r}(z) = \frac{(s+r-1)z}{\sqrt{1-z^2}} \times P^{(s)}_{p,r}(z) + (r-p-1) \times (r+p+s-1) P^{(s)}_{p,r-1}(z) . \tag{A16}$$

Using (A8) and identities[18] satisfied by Gegenbauer polynomials, $P^{(s)}_{p,r}(z)$ is found to satisfy the following identities:

$$(r-p)\sqrt{1-z^2} P^{(s)}_{p,r}(z) = P^{(s)}_{p-1,r+1}(z) - z P^{(s)}_{p,r+1}(z) , \tag{A17}$$

$$(p+r+s)\sqrt{1-z^2} P^{(s)}_{p,r}(z) = P^{(s)}_{p+1,r+1}(z) - z P^{(s)}_{p,r+1}(z) , \tag{A18}$$

$$\sqrt{1-z^2} P^{(s)}_{p,r+1}(z) - r(p+r+s-1)\sqrt{1-z^2} P^{(s)}_{p,r-1}(z) = (2r+p+s)z P^{(s)}_{p,r}(z) - (p+1) P^{(s)}_{p+1,r}(z) , \tag{A19}$$

$$\sqrt{1-z^2} P^{(s)}_{p,r+1}(z) + r(p-r+1)\sqrt{1-z^2} P^{(s)}_{p,r-1}(z) = (2r-p)z P^{(s)}_{p,r}(z) + (p+s-1) P^{(s)}_{p-1,r}(z) . \tag{A20}$$

By combining (A17)–(A20) in various ways, five additional useful identities are obtained:

$$\sqrt{1-z^2} P^{(s)}_{p,r}(z) = (p+r+s-1)z P^{(s)}_{p,r-1}(z) - (p-r+2) P^{(s)}_{p+1,r-1}(z) , \tag{A21}$$

$$\sqrt{1-z^2} P^{(s)}_{p,r}(z) = (r-p-1)z P^{(s)}_{p,r-1}(z) + (p+r+s-2) P^{(s)}_{p-1,r-1}(z) , \tag{A22}$$

$$(2p+s)z\,P_{p,r}^{(s)}(z)$$
$$= (p+r+s-1)\,P_{p-1,r}^{(s)}(z) + (p-r+1)\,P_{p+1,r}^{(s)}(z)\,, \qquad \text{(A23)}$$

$$(2p+s)\sqrt{1-z^2}\,P_{p,r}^{(s)}(z) = P_{p+1,r+1}^{(s)}(z) - P_{p-1,r+1}^{(s)}(z)\,, \qquad \text{(A24)}$$

$$(2p+s)\sqrt{1-z^2}\,P_{p,r}^{(s)}(z)$$
$$= (p+r+s-1)(p+r+s-2)\,P_{p-1,r-1}^{(s)}(z) - (p-r+1)(p-r+2)\,P_{p+1,r-1}^{(s)}(z)\,. \qquad \text{(A25)}$$

The identities are valid if none of the $P_{p,r}^{(s)}$ are zero.

[1]C. D. Anderson and S. H. Neddermeyer, Phys. Rev. **51**, 884 (1937); J. C. Street and E. C. Stevenson, Phys. Rev. **52**, 1033 (1937).

[2]A. O. Barut and J. Kraus, J. Math. Phys. 17, 506 (1976); A. O. Barut, J. Math. Phys. **21**, 568 (1980); A. O. Barut, M. Berrondo, and G. Garcia-Calderón, J. Math. Phys. **21**, 1851 (1980).

[3]A. O. Barut, *American Institute of Physics Conference Proceedings*, edited by T. Seligman (AIP, New York, 1981), Vol. 71, p. 76; A. O. Barut, *Surveys in High Energy Physics* (AIP, New York, 1980), Vol. 1, p. 121.

[4]G. B. Mainland and D. M. Scott, Nuovo Cimento A **74**, 198 (1983); **82**, 357 (1984).

[5]E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).

[6]V. Fock, Z. Phys. **98**, 145 (1935).

[7]M. Lévy, Proc. R. Soc. London Ser. A **204**, 145 (1950).

[8]A. Bassetto, S. Ciccariello, and M. Tonin, Nuovo Cimento **35**, 292 (1964).

[9]J. S. Goldstein, Phys. Rev. **91**, 1516 (1953).

[10]G. C. Wick, Phys. Rev. **96**, 1124 (1954).

[11]R. E.Cutkosky, Phys. Rev. **96**, 1135 (1954).

[12]R. Sugano and Y. Munakata, Prog. Theor. Phys. **16**, 532 (1956).

[13]Our notation is that of J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw–Hill, New York, 1965). We set $\hbar = c = 1$.

[14]The eigensolutions of the Coulomb potential in the Dirac equation were first given by C. G. Darwin, Proc. R. Soc. London Ser. A **118**, 654 (1928) and W. Gordon, Z. Phys. **48**, 11 (1928). Here we follow Ref. 13.

[15]A. Bohm, *Quantum Mechanics* (Springer, New York, 1979).

[16]E. Hecke, Math. Ann. **78**, 398 (1918).

[17]P. Appell and J. Kampé de Fériet, *Fonctions Hypergéométriques et Hyperspheriques* (Gauthier–Villars, Paris, 1926).

[18]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).

Glauber's coherent states and the semiclassical propagator

H. G. Solari[a)]

Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

(Received 28 March 1985; accepted for publication 3 January 1986)

This paper studies the classical propagator on Glauber's coherent states in a generalized P-form. A full class of equivalent expressions labeled by a complex number is found through a direct WKB expansion. The study of matrix elements of the propagator between coherent states shows that they match Klauder's classical expression but are in some disagreement in the semiclassical order with expression derived through path integral procedures. Arguments are given that favor the present results. The properties of the time propagator with respect to symmetry operations are stated. Possible extensions are briefly discussed.

I. INTRODUCTION

Semiclassical-like approaches to quantum problems have received renewed attention in the last years because of their ability to obtain useful descriptions of large systems like heavy ions[1] and also due to the efforts towards understanding the role of chaos in quantum dynamics.[2] Furthermore semiclassical approaches have their proper place in mathematical physics.

The harmonic oscillator coherent states, or Glauber coherent states (GCS's),[3] are well known as the most classical quantum states, i.e., they minimize the Heisenberg uncertainty relationship. Because of this fact, it is expected that they may make semiclassical approaches easier. From another point of view, they are related to the coset representatives of the special nilpotent group[4,5] and form an overcomplete basis, which makes room for a Bargmann[6] space. This property provides a way for generalizations and a great amount of structure.

The earlier works on the semiclassical propagator (SP)[7,8] were not very successful in finding a direct relation between the SP and the classical trajectories in phase space, because of the necessity of complexifying the position and momentum in order to fulfill the boundary conditions, or alternatively because of the necessary introduction of an auxiliary field.[9] In contrast with these earlier works, which concern the matrix elements of the SP between GCS (or more general coherent states), Suzuki[10] proposes, on an heuristic basis, an approach to the classical kernel of the propagator in the P-form. We recall that Q- and P-forms relating operators with coherent states were proposed from the beginning by Glauber[3] (see also Sec. II and Ref. 5). Suzuki's approach is fully developed with respect to the classical trajectories in phase space.

In this work we study Suzuki's propagator through a direct WKB method.[11] This procedure allows us to obtain a generalized form of the classical expression and also go further towards the semiclassical one, which does not coincide with that conjectured by Suzuki[10] (Sec. II). The equivalence of these expressions with Klauder's[7] is shown in Sec. III, where we also compare our semiclassical formula with the equivalent one by Kuratsuji and Mizobuchi[12,13] for spin systems. Section IV is devoted to showing the invariance property of the SP related to the Hamiltonian symmetries (this property was observed in some special situations by Levit[9] and Levit *et al.*[14]). In Sec. V we develop some examples and finally Sec. VI is reserved for the conclusions and perspectives.

II. THE SEMICLASSICAL PROPAGATOR

A. *P*-form of the classical propagator

The matrix elements of an operator with respect to a GCS is a function called the Q-form of the operator $\hat{O}$,

$$\mathscr{O}_Q(\alpha,\beta^*)\equiv\langle\beta|\hat{O}|\alpha\rangle\,, \tag{2.1}$$

where $|\alpha\rangle$ is a GCS,

$$|\alpha\rangle\equiv\exp(-\alpha\alpha^*/2)|\alpha)\,, \tag{2.2}$$

$$|\alpha)=\sum_n(n!)^{-1/2}\alpha^n|n\rangle\,. \tag{2.3}$$

We distinguish here between unnormalized states (normal parentheses) and normalized ones (brackets), the states $|n\rangle$ are the Fock basis states of the harmonic oscillator and α is a complex number, $\alpha=(q+ip)/\sqrt{2}$. The GCS $|\alpha\rangle$ also have the following properties:

$$|\alpha\rangle=\exp(\alpha a^\dagger-\alpha^* a)|0\rangle=\exp(-\alpha\alpha^*/2+\alpha a^\dagger)|0\rangle\,, \tag{2.4a}$$

$$a^\dagger|\alpha\rangle=\exp(-\alpha\alpha^*/2)\frac{\partial|\alpha)}{\partial\alpha}\,, \tag{2.4b}$$

$$a|\alpha\rangle=\alpha|\alpha\rangle\,, \tag{2.4c}$$

$$\exp(-i\epsilon a^\dagger a)|\alpha\rangle=|\alpha\exp(-i\epsilon)\rangle\,, \tag{2.4d}$$

where $a^\dagger$ (a) is the creation (destruction) operator and $|0\rangle$ the ground state of the harmonic oscillator, the expression (2.4c) shows that the GCS's are eigenvectors of the destruction operator.

In addition to the Q-form some operators may be written in a diagonal from,[3] for example, density operators representing statistical mixtures of pure coherent states. This representation is called the P-form and deals with operators like

$$\hat{O}=\int[d\alpha\wedge d\alpha^*\,\mathscr{O}_P(\alpha,\alpha^*)|\alpha)(\alpha|]\,.$$

[a)] Fellow of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) of Argentina.

 0022-2488/86/051351-07$02.50

We note that this P-form does not include Suzuki's propagators and seems to be too restrictive for our proposals. We are going to consider here operators $\hat{B}$ of the form

$$\hat{B} = \int [dg \wedge d\alpha^* B(g,\alpha^*)|g)(\alpha|] , \tag{2.5}$$

where $g = g(\alpha, \alpha^*)$.

It is necessary to demand, for a faithful generalization, that the matrix elements of $\hat{B}$ between coherent states be finite figures, which simply means

$$\left| \int [dg \wedge d\alpha^* B(g,\alpha^*) \exp(g\phi^* + \alpha^* \xi)] \right| < \infty ,$$

for arbitrary but fixed ϕ and ξ.

It is important to note that several different (in appearance) expressions may represent the same operator. This situation will happen at least each time that the integration path ($-\infty < \mathrm{Re}\,\alpha$, $\mathrm{Im}\,\alpha \leqslant \infty$) may be deformed in a complex way. We have established enough conditions in the Appendix in order that the equivalence holds. The essential requirements are that $B(g, \alpha^*)$ be a holomorphic function of g and α^*, and $g(\alpha, \alpha^*)$ be extensible in a holomorphic way to $g(z, \alpha^*)$. If this hypothesis holds, the operator $\hat{B}$ (2.5) may be written in the form

$$\hat{B} = \int \left[dz \wedge d\alpha^* \frac{\partial g}{\partial z}\bigg|_{\alpha^*} B(g(z,\alpha^*),\alpha^*)|g)(\alpha| \right] , \tag{2.6}$$

where $i(z - \alpha^*)$ and $(z + \alpha^*)$ go from $-\infty + i0$ to $\infty + i0$.

As useful examples we write the identities

$$\hat{I} = \frac{1}{\pi} \int_{-\infty}^{\infty} d(\mathrm{Re}\,\alpha) \int_{-\infty}^{\infty} d(\mathrm{Im}\,\alpha) e^{-z\alpha^*}|z)(\alpha| , \tag{2.7}$$

where $z = \delta + \alpha$ for any complex number δ (this formula has been proved in the Appendix), and the evolution operator for the Hamiltonian $\hat{H} = \epsilon a^\dagger a$, which reads [applying (2.4d) and (2.7)]

$$\exp(-i\hat{H}t) = \exp(-i\hat{H}t)\hat{I} = \frac{1}{\pi} \int \int d(\mathrm{Re}\,\alpha) d(\mathrm{Im}\,\alpha) e^{-z\alpha^*}|z_t)(\alpha| ,$$

with $z_t = e^{-i\epsilon t} z$.

This last formula suggests the starting point of our approach for the evolution operator $\hat{U}(t)$, $t = t_f - t_i$, for a time-independent Hamiltonian. We are going to look for a semiclassical approach to $\hat{U}(t)$ assuming that the propagator may be written in the form

$$\hat{U}(t) = \int d\alpha \wedge d\alpha^* \, \mathscr{U}(\gamma_t,\alpha^*,t)|\gamma_t)(\alpha| , \tag{2.8}$$

where γ_t is related with α and α^* by a time-dependent mapping.

The semiclassical approach centers attention on the states with large mean values of position and momentum. We then introduce an order parameter N, which will help us to identify terms in the asymptotic series and is supposed to be large, in the form

$$\alpha = \sqrt{N}\beta . \tag{2.9}$$

The parameter N plays the role of $1/\hbar$ and later will be taken as unity.

The Hamiltonian is assumed to be of the form

$$\hat{H} = N \sum_{nm} b_{nm} N^{-(n+m)/2} (a^\dagger)^n a^m \tag{2.10}$$

and, taking into account Eq. (2.4b), it acts over a GCS as

$$\hat{H}|\alpha) = N \sum_{nm} b_{nm} N^{-(n+m)/2} \alpha^m \frac{\partial^n}{\partial \alpha^n} |\alpha) . \tag{2.11}$$

The starting point for establishing $\mathscr{U}(\gamma_t, \alpha^*, t)$ is the Schrödinger evolution equation

$$\dot{\hat{U}} = -i\hat{H}\hat{U} . \tag{2.12}$$

Taking into account the expression (2.8) for $\hat{U}(t)$ and the action of $\hat{H}$ on a coherent state (2.11), we obtain

$$\begin{aligned} 0 &= \left(\frac{\partial}{\partial t} + i\hat{H}\right)\hat{U}(t) \\ &= \int \left[\frac{d\mathscr{U}}{dt}|\gamma_t)(\alpha| + \mathscr{U}\left(\frac{\partial}{\partial t} + iH\right)|\gamma_t)(\alpha| \right] d\alpha \wedge d\alpha^* , \\ 0 &= \int \left(\frac{\partial \mathscr{U}}{\partial t} + \frac{\partial \mathscr{U}}{\partial \gamma_t}\dot{\gamma}_t + \mathscr{U}\dot{\gamma}_t \frac{\partial}{\partial \gamma_t} \right. \\ &\quad \left. + i\mathscr{U} N \sum_{nm} b_{nm} N^{-(n+m)/2} \gamma_t^m \frac{\partial^n}{\partial \gamma_t^n} \right) \\ &\quad \times |\gamma_t)(\alpha| d\alpha \wedge d\alpha^* , \end{aligned} \tag{2.13}$$

where $\dot{\gamma}_t = d\gamma_t/dt$. Introducing the Jacobian

$$J \equiv \left(\frac{\partial \alpha}{\partial \gamma_t}\right)_{\alpha^*,t} , \tag{2.14}$$

integrating by parts (assuming that the surface at $|\alpha| = \infty$ do not contribute), and considering

$$\mathscr{U}(\gamma_t,\alpha^*,t) = \exp(iNS(\gamma_t,\alpha^*,t))\tilde{\mathscr{U}} , \tag{2.15a}$$

where $\tilde{\mathscr{U}}$ is expanded in the form

$$\tilde{\mathscr{U}} = \sum_{n=1} N^{-n} \tilde{\mathscr{U}}_{n+1} , \tag{2.15b}$$

we are able to write the following expression:

$$\begin{aligned} N \int & \left(iJN\mathscr{U}\frac{\partial S}{\partial t} + Je^{iNS}\frac{\partial \tilde{\mathscr{U}}}{\partial t} - \mathscr{U}\frac{\partial}{\partial z_t}(\dot{z}_t J) \right. \\ & + iN \sum_{nm} \left[b_{nm}(-N)^{-n} \right. \\ & \left. \left. \times \sum_{k=0}^{n} \binom{n}{k} \left(\frac{\partial^k e^{iNS}}{\partial z_t^k} \right) \frac{\partial^{n-k}}{\partial z_t^{n-k}} (\tilde{\mathscr{U}} J z_t^m) \right] \right) \\ & \times |\sqrt{N} z_t)(\sqrt{N}\beta| dz_t \wedge d\beta^* , \end{aligned} \tag{2.16}$$

where we have performed a change of coordinates in the form of Eq. (2.9) and adopted the notation that partial derivatives are considered to be taken with respect to the variables (z_t, β^*, t) except when another situation is explicitly indicated.

Identifying the coefficients of N^2 we get the classical equations

$$-\frac{\partial S}{\partial t} + \mathscr{H}\left(z_t, -i\frac{\partial S}{\partial z_t}, t\right) = 0 , \tag{2.17a}$$

where

$$\mathscr{H}(z,\beta^*) = \sum_{nm} b_{nm} z^n \beta^{*m} = \frac{(\beta|\hat{H}|z)}{(\beta|z)} . \qquad (2.17b)$$

Equations (2.17) are highly undetermined as soon as $z_t(\beta, \beta^*, t)$ is an undetermined function. This situation reflects in part the overcompleteness of the coherent states.

A solution for (2.17) may be found requiring that S be an action in a Hamiltonian set of equations, i.e., we require that

$$-i\frac{\partial S}{\partial z_t} = y_t^* , \qquad (2.18a)$$

$$i\dot{z}_t = \frac{\partial \mathscr{H}}{\partial y_t^*}(z_t, y_t^*)\Big|_{z_t} . \qquad (2.18b)$$

Demanding that $\hat{H}\hat{U} = \hat{U}\hat{H}$, we obtain, in an equivalent form, Eq. (2.18), the last formulas of the Hamiltonian system

$$-i\frac{\partial S}{\partial \beta^*} = w_{t_i} , \qquad (2.18c)$$

$$i\frac{d\beta^*}{dt_i} = \frac{\partial \mathscr{H}}{\partial w_{t_i}}(w_{t_i}, \beta^*)|_{\beta^*} , \qquad (2.18d)$$

where t_i is the starting time of the evolution.

As soon as the Hamiltonian has been considered to be time independent, it is natural to require $\hat{U}(t_f, t_i)$ to be only a function of $t = (t_f - t_i)$, and then it follows that

$$\mathscr{H}(z_t, y^*) = \mathscr{H}(w_{t_i}, \beta^*) . \qquad (2.18e)$$

The set of equations [(2.17) and (2.18)] allows us to integrate S, which turns out to be

$$S = \int_{t_i}^{t_f} (iy^*\dot{z}_t - \mathscr{H}(z_t, y^*))dt + \beta^* z_{t_i} + \text{const} , \qquad (2.19)$$

where $z_t \equiv w_t$ and $y_t^* \equiv \beta_t^*$ fulfill the Hamiltonian equation (2.18b).

The action S reflects some undetermination related to the initial conditions for z_t and y^*. This problem may be studied comparing the classical propagator at $t = 0$ with the identity (2.7), which implies that

$$z_{t_i} = \beta + \delta , \qquad (2.20)$$

with δ an arbitrary complex number. The full set of equations are summarized in Table I.

The classical propagator reads

$$\hat{U}_{\rm cl} = \int \exp(iS(z_t, \beta^*, t))/i\pi |z_t)(\beta| d\beta \wedge d\beta^* \qquad (2.21)$$

and the particular choice of $\delta = 0$ in (2.20) brings us to the Suzuki propagator.[10]

B. The semiclassical propagator

We are now going to analyze the first correction to the classical approach, which comes from the coefficient of N^1 in (2.16). The correction $\tilde{\mathscr{U}}_2$ must satisfy the following equation:

$$0 = -\tilde{\mathscr{U}}_2 J\frac{\partial \dot{z}_t}{\partial z_t} + J\frac{d\tilde{\mathscr{U}}_2}{dt} - i\tilde{\mathscr{U}}_2 J\frac{\partial^2 \mathscr{H}(z_t, \beta_t^*)}{(\partial z_t\, \partial \beta_t^*)} - 0.5\tilde{\mathscr{U}}_2 J\frac{\partial^2 \mathscr{H}(z_t, \beta_t^*)}{(\partial \beta_t^*)^2}\frac{\partial^2 S}{(\partial z_t)^2} . \qquad (2.22)$$

Considering $\tilde{\mathscr{U}}_2 = e^{iS_2}$ with the initial condition $S_2(z_0, \beta^*, 0) = 0$, the second-order action must satisfy the equation

$$\frac{i\,dS_2}{dt} = \left(i\frac{\partial^2 \mathscr{H}(z_t, \beta_t^*)}{\partial z_t\, \partial \beta_t^*} + 0.5\frac{\partial^2 \mathscr{H}(z_t, \beta_t^*)}{\partial \beta_t^{*2}}\frac{\partial^2 S}{\partial z_t^2} + \frac{\partial \dot{z}_t}{\partial z_t}\right) , \qquad (2.23)$$

which, after some handling, is transformed to

$$i\frac{dS_2}{dt} = -0.5\frac{d}{dt}\left[\ln\left(\frac{\partial^2 S}{(\partial \beta^*\, \partial z_t)}\right)\right] + 0.5i\frac{\partial^2 \mathscr{H}(z_t, \beta_t^*)}{\partial z_t + \partial \beta_t^*} . \qquad (2.24)$$

Time integration of (2.23) and (2.24), taking into account the initial condition, yields the first correction to the classical kernel of the time propagator as

$$\tilde{\mathscr{U}}_2 = \pm\left(i\frac{\partial^2 S}{\partial z_t\, \partial \beta^*}\right)^{1/2} \exp\left[0.5i\int_{t_i}^{t_f} \frac{\partial^2 \mathscr{H}(z_t, \beta_t^*)}{\partial z_t\, \partial \beta_t^*}dt\right] . \qquad (2.25)$$

It is interesting to note here that while the first term may be identified with the square root of the Jacobian J, Eq. (2.14),

$$-i\frac{\partial^2 S}{\partial z_t\, \partial \beta^*} = \frac{\partial \beta_t^*}{\partial \beta^*} = \frac{\partial z_0}{\partial z_t} = J , \qquad (2.26)$$

an additional exponential term appears. The semiclassical propagator then reads

$$\hat{U}_{\rm sc} = \int \frac{d\beta \wedge d\beta^*}{i\pi} J^{-1/2} \exp\left(0.5i\int_{t_i}^{t_f} dt \frac{\partial^2 \mathscr{H}(z_t, \beta_t^*)}{\partial z_t\, \partial \beta_t^*}\right) \times \exp(iS(z_t, \beta^*, t))|z_t)(\beta| . \qquad (2.27)$$

In this formula we also have the freedom to choose $\delta = 0$ in (2.20) identifying $z_t = \beta_t = (\beta_t^*)^*$.

The expression for $\tilde{\mathscr{U}}_2$ (2.25) shows that it is in special situations, as the one considered in Sec. V, that the correction depends only on time. This last assertion was Suzuki's conjecture[10] (however suppossed to be valid in every case). That is, Suzuki wrote his semiclassical propagator in the form

TABLE I. Full set of equations of the Hamiltonian classical system.

$\frac{\partial S}{\partial t_f} = -\frac{\partial S}{\partial t_i} = \mathscr{H}(z_{t_f}, \beta_{t_f}^*)$	(2.17)
$-i\frac{\partial S}{\partial z_{t_f}} = y^* \equiv \beta_{t_f}^*$	(2.18a)
$-i\frac{\partial S}{\partial \beta^*} = z_{t_i} \equiv z_0$	(2.18c)
$i\dot{z}_t = \frac{\partial \mathscr{H}(z_t, \beta_t^*)}{\partial \beta_t^*}$	(2.18b)
$i\dot{\beta}_t^* = \frac{-\partial \mathscr{H}(z_t, \beta_t^*)}{\partial z_t}$	(2.18d)
$z_0 = \beta + \delta$	(2.20)
$S = \int_{t_i}^{t_f} (i\beta_t^* \dot{z}_t - \mathscr{H}(z_t, \beta_t^*))dt + \beta^* z_0 + \text{const}$	(2.19)

$$\hat{U} = n(t)\int d\beta \wedge d\beta^* \exp(iS(\beta_t,\beta_t^*,t))|\beta_t)(\beta| , \qquad (2.28)$$

where the secondary kernel $\tilde{\mathscr{U}}_2(z_t,\beta^*,t)$ does not depend on the trajectory and is partially represented by the time-dependent c-number $n(t)$.

III. MATRIX ELEMENTS OF THE SEMICLASSICAL PROPAGATOR

The matrix elements of the semiclassical propagator with respect to GCS have been studied by Klauder[7] and later by Levit[9] and Blaizot and Orland.[15] It is useful to make contact with this approach from the semiclassical propagator of Eq. (2.27); this proposal only requires the evaluation of the integral by a method consistent with the semiclassical approximation like the saddle point method.[16]

Taking into account the order parameter N again, we have to evaluate

$$(\phi|\hat{U}_{sc}|\xi) = \int \frac{d\beta\wedge d\beta^*}{i\pi}\exp\{N(iS + z_t\phi^* + \xi B^*)\}. \qquad (3.1)$$

The saddle point is characterized by the set of equations

$$\phi^* = \beta_t^* , \qquad (3.2a)$$

$$\xi = z_0 . \qquad (3.2b)$$

These equations cannot be fulfilled for an arbitrary δ, Eq. (2.20), because the equations of motion (2.18) only admit one complex initial condition; the satisfaction of (3.2) would require complexifying the position and momentum as in Refs. 7 and 15. We rather fix the value of δ in order to make possible the solution of (3.2), noting that the same problem and also the same solutions appear when one evaluates the matrix elements of the identity, (2.7b), by the saddle point method in the form

$$(\phi|I|\xi) = \exp(\phi^* g + \xi\alpha^* - g\alpha^*) ,$$

where

$$\alpha^* = \phi^*, \quad g = \xi ,$$

which fix the value of δ in the form $\delta = \xi - \phi$, producing the semiclassical evaluation of the matrix elements of the identity by

$$(\phi|\xi) = \exp(\phi^*\xi) .$$

The classical evaluation of (3.1) reads

$$(\phi|\hat{U}_{sc}|\xi) = \exp(iS + \xi\beta^* + \phi^* z_t) \equiv \exp(iF) , \qquad (3.3a)$$

where we have defined the function

$$F(\xi,\phi^*) = \int_{t_i}^{t_f} dt\,(i\beta_t^*\dot{z}_t - \mathscr{H}(z_t,\beta_t^*)) - iz_{t_f}\phi^* , \qquad (3.3b)$$

whose partial derivatives are

$$i\frac{\partial F(\xi,\phi^*,t)}{\partial\xi} = z_{t_f}, \qquad (3.3c)$$

$$i\frac{\partial F(\xi,\phi^*,t)}{\partial\phi^*} = \beta_{t_i}^* . \qquad (3.3d)$$

The equations of motion that arise from the variation

$$\delta(F) = 0 \qquad (3.4a)$$

are

$$i\dot{z}_t = \frac{\partial\mathscr{H}(z_t,\beta_t^*)}{\partial\beta_t^*}, \quad z_{t_i} = \xi , \qquad (3.4b)$$

$$-i\dot{\beta}_t^* = \frac{\partial\mathscr{H}(z_t,\beta_t^*)}{\partial z_t}, \quad \beta_t^* = \phi^* . \qquad (3.4c)$$

Equations (3.3) and (3.4) were also found by Klauder[7] with boundary conditions in both ends (t_i and t_f) and are in contrast with Eqs. (2.18) and (2.19) because of the difference in these conditions.

Expanding the exponent in Eq. (3.1) around the saddle point up to the second order and evaluating $\tilde{\mathscr{U}}_2$ at this point, we obtain the next-order approximation. Following this procedure and after some handling we get

$$(\phi|\hat{U}_{sc}|\xi) \simeq \exp\left(iF + 0.5i\int_{t_i}^{t_f} dt\frac{\partial^2\mathscr{H}(z_t,\beta_t^*)}{\partial z_t\,\partial\beta_t^*}\right) \times\left(-i\frac{\partial^2 F}{\partial z_t\,\partial\beta^*}\right)^{1/2} , \qquad (3.5)$$

which must be considered together with (3.3) and (3.4).

We may compare this last expression with the one obtained in Refs. 12 and 13 for spin systems by means of a path integral approach. We note at a first glance that the present expression is a little more closed because it is not expressed in terms of the eigenvalues of a Sturm–Liouville problem.[12,13] Despite this small difference we note that for $t_f = t_i$ we obtain

$$(\phi|\hat{\mathscr{U}}_{sc}|\xi) \simeq \exp(-\phi^*\xi) = (\phi|\hat{I}|\xi) = (\phi|\xi) , \qquad (3.6)$$

which is the correct figure, while in a simple example presented in Ref. 13 (considering the J_z Hamiltonian) the matrix elements for the semiclassical propagator evaluated at $t_f = t_i$ diverges. It has been suggested (Ref. 17, p. 248) that this behavior may be due to a too restricted election of the paths.

This procedure completes the evaluation of the semiclassical matrix elements, which also presents an extra phase coming from the second-order term. The formulas (2.27) and (3.5) are easily found as the exact figures for the harmonic oscillator, nontrivial examples are shown in Sec. V.

IV. PERIODIC HAMILTONIANS

The expression (2.27) of the SP is very well suited for studying the way in which symmetries are dealt with in the semiclassical approach. Symmetries are important by themselves and also may be reflected in the requantization formulas obtained from the coherent states-SP as found by Levit *et al.* in the frame of the Lipkin model.[14] Approximate formulas for the energies of a system may be obtained by the Gutzwiller quantization method[18–21] that assigns a relevant role to the closed orbits and their neighborhoods.[18] We are not developing this point in the present work, but let us mention that the approach to the SP here developed may allow a better understanding of chaotic motion and its relationship with quantal energies just because this approach is based on actual classical trajectories in phase space (and not only on stationary paths).

Turning to the symmetry problem we are going to suppose that the Hamiltonian $\hat{H}$ does commute with the operator $\hat{G}$,

$$\hat{G} = \exp(\alpha a^\dagger - \alpha^* a) , \tag{4.1}$$

for some fixed α, not necessarily infinitesimal,

$$[\hat{G},\hat{H}] = 0 . \tag{4.2}$$

The action of the operator G on the SP, $\hat{U}_{sc}$, reads

$$\hat{G}\hat{U}_{sc} = \int \frac{d\beta\wedge d\beta^*}{i\pi} \tilde{U}_2 e^{iS}\hat{G}\,|z_t)(\beta| , \tag{4.3}$$

while the action of $\hat{G}$ on an (unnormalized) GCS is

$$\hat{G}\,|z) = \exp(-\alpha\alpha^*/2 - z\alpha^*)|z+\alpha) . \tag{4.4}$$

From this equation and (4.2) we get

$$\mathscr{H}(z_t,\beta_t^*) = \frac{(\beta_t|\hat{H}|z_t)}{(\beta_t|z_t)} = \frac{(\beta_t+\alpha|\hat{H}|z_t+\alpha)}{(\beta_t+\alpha|z_t+\alpha)}$$
$$= \mathscr{H}(z_t+\alpha,\beta_t^*+\alpha^*) \equiv \mathscr{H}(u_t,v_t) . \tag{4.5}$$

It follows that the pair of functions

$$u_t \equiv z_t + \alpha , \tag{4.6a}$$

$$v_t^* \equiv \beta_t^* + \alpha , \tag{4.6b}$$

is also a solution of Eqs. (2.18b) and (2.18d) with starting points at

$$u_{t_i} = \beta + \alpha + \delta \tag{4.7a}$$

and

$$v^* = \beta^* + \alpha^* . \tag{4.7b}$$

From the periodicity of $\mathscr{H}$, (4.5), the periodicity of the phase portrait with respect to the shifting by α also follows. It is easy to verify by straightforward integration from the expression of the action S in (2.20) that its simultaneous evaluation at the points (z_t,β^*) and (u_t,v^*) gives the relation

$$iS(z_t,\beta^*,t) = iS(u_t,v^*,t) - \alpha\alpha^* + \alpha v^* + \alpha^* u_t , \tag{4.8}$$

which in turn allows us to verify the following relationship between the second derivatives of S:

$$\frac{\partial^2 S(u_t,v^*,t)}{\partial u_t\,\partial v^*} = \frac{\partial^2 S(z_t,\beta^*,t)}{\partial z_t\,\partial\beta^*} . \tag{4.9}$$

Equations (4.8) and (4.9) allow us to express $\hat{G}\,\hat{U}_{sc}$ in (4.3) through a change to the new variables $v = \beta + \alpha$ and $v^* = (\beta+\alpha)^*$, as

$$\hat{G}\hat{U}_{sc} = \int \frac{dv\wedge dv^*}{i\pi} \tilde{\mathscr{U}}_2(u_t,v^*,t)\exp(iS(u_t,v^*,t))$$
$$\times |u_t)(v|\hat{G} = \hat{U}_{sc}\hat{G} , \tag{4.10}$$

which proves that both the classical and the semiclassical propagators do commute with the shifting operator $\hat{G}$, Eq. (4.1), if the Hamiltonian $\hat{H}$ does. Let us remark that the proof that we develop here is based on group theoretical arguments and may be directly extended to Suzuki's propagators.

V. AN EXAMPLE: THE QUADRATIC HAMILTONIAN

We are going to consider the SP for a system with a Hamiltonian of the form

$$\hat{H} = \epsilon a^\dagger a + 0.5(v(a^\dagger)^2 + v^*(a)^2) . \tag{5.1}$$

Associating with $\hat{H}$ the classical Hamiltonian $\mathscr{H}$,

$$\mathscr{H}(z,\beta^*) = \epsilon z\beta^* + 0.5(vz^2 + v^*\beta^{*2}) , \tag{5.2}$$

and the equations of motion

$$i\dot{z}_t = \epsilon z_t + v\beta_t^* = \left.\frac{\partial\mathscr{H}}{\partial\beta_t^*}\right|_{z_t} , \quad z_0 = \beta_0 + \delta , \tag{5.3a}$$

$$-i\dot{\beta}_t^* = \epsilon\beta_t^* + v^* z_t = \left.\frac{\partial\mathscr{H}}{\partial z_t}\right|_{\beta_t^*} , \quad \beta_0^* = \beta^* . \tag{5.3b}$$

The action (2.19) for this system simply reads

$$S = 0.5i(z_0\beta^* + z_t\beta_t^*) , \tag{5.4a}$$

where

$$z_t = z_0\cos(\omega t) - i/\omega(\epsilon z_0 + v\beta^*)\sin(\omega t) , \tag{5.4b}$$

and

$$\beta_t^* = \beta^*\cos(\omega t) + i/\omega(\epsilon\beta^* + v^* z)\sin(\omega t) , \tag{5.4c}$$

$$\omega = (\epsilon^2 - v^* v)^{1/2}$$

are the solutions of (5.3). The action in (5.5a) may be explicitly written in terms of (z_t,β^*,t) in the form

$$S = \frac{i(\beta^* z_t + 0.5i/\omega(v\beta^{*2} + v^* z_t^2)\sin(\omega t))}{\cos(\omega t) - i\epsilon/\omega - \sin(\omega t)} \tag{5.5}$$

and its second derivative with respect to z_t and β^* is

$$\frac{\partial^2 S}{\partial z_t\,\partial\beta^*} = \frac{1}{\cos(\omega t) - i\epsilon/\omega\,\sin(\omega t)}/1 = iJ , \tag{5.6}$$

which in this case does not depend on the variables z_t and β^*.

The semiclassical propagator in the P-form (2.27) reads in this particular model

$$\hat{U}_{sc} = \int d\beta\wedge d\beta^*$$
$$\times\exp(-(\beta^* z_t + 0.5i/\omega(v^* z_t^2 + v\beta^{*2})\sin(\omega t))/J)$$
$$\times J^{-1/2}/(i\pi)\exp(0.5i\epsilon t)\,|z_t)(\beta| . \tag{5.7}$$

It may be directly checked[10] that for the quadratic Hamiltonian (5.1) the SP matches the exact one as usual.[17]

VI. CONCLUSIONS

We have studied the classical propagator expressed in terms of the Glauber coherent states in the P-form. It has been shown in this context that one of them matches Suzuki's classical expressions. In this case, the kernel of the propagator is fully determined by the classical trajectories in phase space (labeled by their initial conditions). We have also developed explicitly the formulas to the next order, which contribute not only to the square root of the Jacobian but also to an additional time-dependent phase.

We have shown that the expression developed here is equivalent (to the same order in the asymptotic series) to the Klauder formulas for the matrix elements of the semiclassical propagator. Going to these matrix elements we have to match two complex boundary conditions, which forces us to fix the complex parameter δ that labels the elements of the class of the semiclassical propagators. This requirement as well as the existence of a complete class of SP (P-form) is fully related to the overcompleteness of the coherent states.

We have also shown that the SP preserves the dynamic symmetries of the Hamiltonian.

Generalization to systems of several particles (or to higher dimensions) seems to be immediate, and because the most important points are based on the overcompleteness of the coherent states and their group structure we believe that many expressions may be extended to more general coherent states, in particular the symmetry properties of the Suzuki's propagator are always valid. In addition to this common structure, we observe that the proofs developed in Sec. II make use of several special properties of GCS's, not shared by other sets of coherent states; because of this situation an actual generalization beyond GCS requires some work, which is now in progress.

ACKNOWLEDGMENT

I am grateful to Dr. E. Susana Hernández for helpful discussions and a careful reading of the manuscript.

APPENDIX: EQUIVALENCE RELATIONS FOR *P*-FORM OPERATORS

We wish to establish enough conditions for the equivalence of any two different generalized P-forms of an operator. In other words, we are going to sketch a proof for the following theorem.

Theorem: Let $\mathcal{O}(q,p)$ be a function such that

$$\int \mathcal{O}(q,p)\exp(A(q,p))dq\,dp=\gamma,\quad |\gamma|<\infty\,,$$

where the integration is extended to the whole plane (q,p) and A is a linear function of (q,p) with complex range. In addition, let $\mathcal{O}(q,p)$ be extensible to a holomorphic function of z_1 and z_2, $\mathcal{O}(z_1,z_2)$, in a domain D which includes the plane $(q+i0;p+i0)$. Then

$$\int dz_1\int dz_2\;\mathcal{O}(z_1,z_2)\exp(A(z_1,z_2))=\gamma\,,$$

where Γ_1 and Γ_2 are curves that belong to D and extend from $-\infty+i0$ to $\infty+i0$.

Proof: As $\mathcal{O}(z_1,z_2)$ is a holomorphic function in D,

$$\Delta(z_2)\equiv\int_{-\infty}^{+\infty}\mathcal{O}(q,z_2)\exp(A(q,z_2))dq$$
$$=\int_{\Gamma_1}\mathcal{O}(z_1,z_2)\exp(A(z_1,z_2))dz_1,$$

where the equality holds by virtue of Cauchy's theorem on holomorphic functions. The function $\Delta(z_2)$ will be also holomorphic because of the hypothesis on $\mathcal{O}(z_1,z_2)$ and assuming that the derivation under the integration sign is allowed. Applying Cauchy's theorem again, we obtain

$$\gamma=\int_{\Gamma_2}dz_2\int dz_1\;\mathcal{O}(z_1,z_2)\exp(A(z_1,z_2))\,,$$

which completes the proof.

The possibility of generalizing the complex paths for each integral depends upon the convergence conditions of $\mathcal{O}(z_1,z_2)$ for large $|\mathrm{Re}\,z_1|$ and $|\mathrm{Re}\,z_2|$. It is reasonable to expect that

$$\lim_{A\to\infty}\int_{\Gamma'(A)}\mathcal{O}(z_1,z_2)\exp(A(z_1,z_2))dz_1=0\,,$$

if $\Gamma'(A)$ is a curve which goes from A to $A+if(A)$ and $|\lim_{A\to\infty}f(A)|$ is finite, because $\mathcal{O}(q,p)$ goes to zero when q goes to infinity faster than any exponential.

We are going to show two corollaries of the theorem.

Corollary 1: $\mathcal{O}'(\alpha,\ \alpha^*)=\mathcal{O}(q,\ p)$, where $\alpha=(q+ip)/\sqrt{2}$, and we restrict ourselves to paths such that $z_1-iz_2=\alpha^*/\sqrt{2}$, we obtain the identity

$$\int\mathcal{O}'(\alpha,\alpha^*)\exp(A'(\alpha,\alpha^*))dq\,dp$$
$$=\int\mathcal{O}'(\alpha,\alpha^*)\exp(A'(\alpha,\alpha^*))\frac{d\alpha\wedge d\alpha^*}{i}$$
$$=\int_{\Gamma_1}\int_{\Gamma_2}\mathcal{O}'(g,\alpha^*)\exp(A'(g,\alpha^*))\frac{dg\wedge d\alpha^*}{i}\,,$$

where

$$A'(\alpha,\alpha^*)=A(q,p)\quad\text{and}\quad g=(z_1+iz_2)/\sqrt{2}\,.$$

Corollary 2: The identity may be written in terms of coherent states in the form

$$\hat{I}=\frac{1}{\pi}\int_{-\infty+ia}^{\infty+ia}dp\int_{-\infty+ib}^{\infty+ib}dq\exp(-z\alpha^*)|z)(\alpha|$$
$$=\frac{1}{\pi}\int_{-\infty}^{\infty}dp\int_{-\infty}^{\infty}dq\exp(-z\alpha^*)|z)(\alpha|\,,$$

where $z=\alpha+(a+ib)$.

The proof follows by applying the theorem and Corollary 1 to the matrix elements of $\hat{I}$ between coherent states and noting for the second equation that

$$0=\int_{-\infty}^{-\infty+ia}dp\exp(-z\alpha^*)|z)(\alpha|$$
$$=\int_{\infty}^{\infty+ia}dp\exp(-z\alpha^*)|z)(\alpha|$$

and

$$0=\int_{-\infty}^{-\infty+ib}dq\exp(-z\alpha^*)|z)(\alpha|$$
$$=\int_{\infty}^{\infty+ib}dq\exp(-z\alpha^*)|z)(\alpha|$$

because $\exp(-z\alpha^*+A(z,\alpha^*))$ (A linear) goes to zero as $|\alpha|$ goes to infinity as a quadratic exponential.

We note that this kind of deformation is necessary when we evaluate matrix elements by means of the saddle point method[16] and has been formally applied in Refs. 7, 12, 13, and 15 calling the procedure the complexification of the variables q and p.

[1]J. W. Negele, Rev. Mod. Phys. 54, 913 (1982).
[2]Y. Weissman and J. Jortner, J. Chem. Phys. 77, 1486 (1982).
[3]R. J. Glauber, Phys. Rev. 131, 2766 (1963).
[4]A. Perelomov, Commun. Math. Phys. 26, 222 (1972).
[5]R. Gilmore, Rev. Mex. Fis. 23, 143 (1974).

[6]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[7]J. Klauder, Phys. Rev. D **19**, 2349 (1979).
[8]J. Klauder, Ann. Phys. NY **11**, 123 (1960).
[9]S. Levit, Phys. Rev. C **21**, 1594 (1980).
[10]T. Suzuki, Nucl. Phys. A **398**, 557 (1983).
[11]A. Messiah, *Mechanica Cuantica* (Tecnos, Madrid, 1965).
[12]H. Kuratzuji and Y. Mizobuchi, Phys. Lett. A **82**, 279 (1981).
[13]H. Kuratzuji and Y. Mizobuchi, J. Math. Phys. **22**, 757 (1981).
[14]S. Levit, J. W. Negele, and L. Paltiel, Phys. Rev. C **21**, 1603 (1980).
[15]J. P. Blaizot and H. Orland, Phys. Rev. C **24**, 1740 (1981).
[16]E. T. Copson, *Asymptotic Expansions* (Cambridge U. P., Cambridge, 1965).
[17]L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[18]M. C. Gutzwiller, J. Math. Phys. **12**, 343 (1971).
[19]M. C. Gutzwiller, J. Math. Phys. **8**, 1979 (1967).
[20]M. C. Gutzwiller, J. Math. Phys. **10**, 1004 (1969).
[21]M. C. Gutzwiller, J. Math. Phys. **11**, 1791 (1970).

Lobachevskian quantum field theory in 2ν dimensions

G. A. Kerimov
Institute of Physics of the Academy of Sciences of the Azerbaijan SSR, Baku, U.S.S.R.

(Received 12 September 1985; accepted for publication 3 October 1985)

In this, the first of a series of papers on quantum field theory in a Lobachevskian space (a space-time, topologically Euclidean, of constant negative curvature), the scalar fields are considered. These Lobachevskian fields, which are related in the flat-space limit to the Euclidean fields, act in a space $\mathscr{H}$, which again has Fock space structure. A formalism is presented that is valid in an arbitrary number of dimensions and therefore allows the use of dimensional regularization.

I. INTRODUCTION

This paper is the first of two devoted to the study of quantum field theories in Lobachevskian space (LQFT). In the present paper we shall consider only self-interacting scalar field theories. The interacting systems involving spin-$\frac{1}{2}$ and spin-0 fields will be studied in the next paper of this series. The formalism presented here is valid in an arbitrary number of dimensions and therefore allows the use of dimensional regularization to investigate the renormalization of the theory.

The key to the construction of LQFT lies in the observation that Lobachevskian space is diffeomorphic to the Euclidean space, and its group of motions [the well-known de Sitter group $SO_0(2\nu, 1)$] contracts [with respect to $SO(2\nu)$] onto the Euclidean group $ISO(2\nu)$ in the sense of Inonu and Wigner.[1] This suggests the following assumption that the Euclidean quantum field theory (EQFT) must be in some sense a limiting case of LQFT. Our treatment is based on this chief assumption.

We note that the EQFT is unusual in two respects.[2] First, the Euclidean fields are completely commutative or anticommutative, as befits the statistics. Second, the Euclidean field does not satisfy a free-field equation even in the absence of interactions (or equivalently, the Euclidean one-particle space supports highly reducible representation of the Euclidean group[3]). The Lobachevskian fields constructed in this series of papers also satisfy these properties.

Our paper is organized as follows: We will construct Lobachevskian fields for scalar bosons by starting with creation and annihilation operators satisfying the canonical commutation relations and then defining the fields in terms of these operators. This Fock space construction is carried out in Sec. IV and goes quite parallel to ordinary Euclidean Fock space construction. For this purpose we recall the definition of an Euclidean free Bose field in Sec. II. In order to make the paper as self-contained and readable as possible, we briefly summarize the salient features of Lobachevskian space and give the definition of a Fourier transform on Lobachevskian space in Sec. III. In Sec. V we briefly discuss the interacting theory. A useful integral is calculated in the Appendix.

II. FREE EUCLIDEAN BOSON FIELDS

In this section we summarize the theory of a free Euclidean boson field of mass $m > 0$ (see Ref. 3).

The Euclidean one-particle boson space $\mathscr{E}^{(1)}$ is represented as $\mathscr{L}^2(R^{2\nu})$ and the boson Fock space is the Hilbert space completion of the symmetric tensor algebra over $\mathscr{E}^{(1)}$,

$$\mathscr{E} = \mathbb{C} \oplus \mathscr{E}^{(1)} \oplus (\mathscr{E}^{(1)} \otimes_S \mathscr{E}^{(1)}) \oplus \cdots .$$

The vacuum is denoted by $|0\rangle$. In the standard fashion we introduce annihilation and creation operators $a(\mathbf{p})$ and $a^*(\mathbf{p})$, $\mathbf{p} \in R^{2\nu}$, with the commutation relations

$$[a(\mathbf{p}), a^*(\mathbf{p}')]_- = \delta^{2\nu}(\mathbf{p} - \mathbf{p}') .$$

The Euclidean boson fields are then defined by

$$\varphi(\mathbf{x}) = \frac{1}{(2\pi)^\nu} \int \frac{d^{2\nu}p}{\sqrt{\mathbf{p}^2 + m^2}} (a(\mathbf{p})e^{i(\mathbf{p},\mathbf{x})} + a^*(\mathbf{p})\, e^{-i(\mathbf{p},\mathbf{x})})$$

[here and in the following we use the scalar product $(\mathbf{p}, \mathbf{x}) = \Sigma_{\alpha=1}^{2\nu} p_\alpha x_\alpha$] or, in polar coordinates $\mathbf{p} = p\mathbf{n}, p \geqslant 0$, and $\mathbf{n}$ a unit vector,

$$\varphi(x) = (2\pi)^{-\nu} \int_{R^+} \int_{S^{2\nu-1}} \frac{1}{\sqrt{p^2 + m^2}} \{a(p;\mathbf{n})\, e^{ip(\mathbf{n},\mathbf{x})} + a^*(p;\mathbf{n}) e^{-ip(\mathbf{n},\mathbf{x})}\} p^{2\nu-1}\, dp\, d\mathbf{n} , \tag{2.1}$$

where $R^+ = \{p \in R\colon p \geqslant 0\}$ and $d\mathbf{n}$ is the Euclidean measure on the unit sphere $S^{2\nu-1}$. We note that the operators $a(p; \mathbf{n})$ and $a^*(p; \mathbf{n})$ satisfy the commutation relations

$$[a(p;\mathbf{n}), a^*(p';\mathbf{n}')]_- = p^{1-2\nu}\delta(p - p')\delta^{2\nu-1}(\mathbf{n} - \mathbf{n}') , \tag{2.2}$$

where $\delta^{2\nu-1}(\mathbf{n} - \mathbf{n}')$ is the Dirac distribution on the $S^{2\nu-1}$,

$$\int_{S^{2\nu-1}} d\mathbf{n}\, \delta^{2\nu-1}(\mathbf{n} - \mathbf{n}') f(\mathbf{n}) = f(\mathbf{n}') .$$

We have a unitary representation of the inhomogeneous rotation group $ISO(2\nu)$ (here called the Euclidean group) on $\mathscr{E}$ defined by

$$U(\mathbf{a}, k)|0\rangle = |0\rangle , \tag{2.3}$$

$$U(\mathbf{a}, k)a^*(p;\mathbf{n})U^{-1}(\mathbf{a}, k) = e^{-ip(k\mathbf{n},\mathbf{a})}a^*(p; k\mathbf{n}) , \tag{2.4}$$

where $\mathbf{a} \in R^{2\nu}$, $k \in SO(2\nu)$. Therefore the field transforms as

$$U(\mathbf{a}, k)\varphi(\mathbf{x})U^{-1}(\mathbf{a}, k) = \varphi(k\mathbf{x} + \mathbf{a}) . \tag{2.5}$$

The two-point Green's function is given by

$$\begin{aligned} G_E(\mathbf{x}, \mathbf{x}') &\equiv \langle 0|\varphi(\mathbf{x})\varphi(\mathbf{x}')|0\rangle \\ &= (2\pi)^{-2\nu} \int_{R^+} \int_{S^{2\nu-1}} (p^2 + m^2)^{-1} \times e^{-ip(\mathbf{n},\mathbf{x}-\mathbf{x}')} p^{2\nu-1}\, dp\, d\mathbf{n} \\ &= (2\pi)^{-\nu}(m/|\mathbf{x} - \mathbf{x}'|)^{\nu-1}K_{\nu-1}(m|\mathbf{x} - \mathbf{x}'|) , \end{aligned} \tag{2.6}$$

 0022-2488/86/051358-05$02.50

where $K_{\nu-1}$ is the modified Bessel function of the third kind.

Obviously Euclidean boson fields are commutative, i.e.,

$$[\varphi(\mathbf{x}), \varphi(\mathbf{x}')]_- = 0, \quad \text{for all } \mathbf{x}, \mathbf{x}' \in R^{2\nu}.$$

This is in contrast to relativistic free boson fields, which commute only for spacelike separated points.

We have already recalled in the Introduction that the Euclidean one-particle boson space $\mathscr{E}^{(1)}$ supports a highly reducible representation of ISO(2ν). It follows from (2.4) that this representation is decomposed onto the direct integral[4]

$$U(\mathbf{a}, k) = \int_{R^+} T^p(\mathbf{a}, k) p^{2\nu-1}\, dp \tag{2.7}$$

of class I (or scalar) unitary irreducible representations (UIR's) $T^p(\mathbf{a}, k)$ of ISO(2ν) characterized by parameter p ($0<p<\infty$). Hence the Euclidean fields may be viewed as creating and annihilating virtual particles.

Let us note at this point that the class I UIR of ISO(2ν) is realized in the Hilbert space $\mathscr{L}^2(S^{2\nu-1})$ of square-integrable functions $f(\mathbf{n})$ over the unit sphere $S^{2\nu-1}$. The representations $T^p(\mathbf{a}, k)$ are defined by (see, e.g., Chap. 9 of Ref. 4)

$$T^p(\mathbf{a}, k) f(\mathbf{n}) = e^{-ip(\mathbf{n}, a)} f(k^{-1}\mathbf{n}). \tag{2.8}$$

It is also worth noting that the function

$$e_{p,\mathbf{n}}: \mathbf{x} \to e^{-ip(\mathbf{n}, x)} \tag{2.9}$$

has the following properties: (i) $e_{p,\mathbf{n}}$ is constant on each hyperplane perpendicular to $\mathbf{n}$ (i.e., $e_{p,\mathbf{n}}$ is a plane wave with the normal $\mathbf{n}$); and (ii) $e_{p,\mathbf{n}}$ is an eigenfunction of the Laplace operator on $R^{2\nu}$.

III. REPRESENTATIONS OF THE de SITTER GROUP

In our approach it is essential to know of the de Sitter transformation properties of creation and annihilation operators. For this purpose we give in this section a short description of class I (or scalar) principal series representation of the de Sitter group and the definition of Fourier transform on Lobachevskian space.

Let $R^{2\nu,1}$ be a $(2\nu+1)$-dimensional pseudo-Euclidean space with the bilinear form

$$[\xi, \eta] \equiv \xi^0\eta^0 - \xi^1\eta^1 - \xi^2\eta^2 - \cdots - \xi^{2\nu}\eta^{2\nu}. \tag{3.1}$$

The Lobachevskian space $\Lambda^{2\nu}$ (a space-time, topologically Euclidean, of constant negative curvature) may be realized as the 2ν-dimensional hypersurface

$$[\xi, \xi] = 1, \quad \xi^0 > 0, \tag{3.2}$$

in $R^{2\nu,1}$. On the hypersurface (3.2) we use the metric induced from $R^{2\nu+1}$. [We note that the metric on the hyperboloid (3.2) is positive definite.] It is clear from (3.2) that the group of motions for Lobachevskian space $\Lambda^{2\nu}$ is $SO_0(2\nu, 1)$ (here called the de Sitter group). The group $SO_0(2\nu, 1)$ acts transitively in $\Lambda^{2\nu}$. It is also worth noting that $\Lambda^{2\nu}$ is isomorphic to the coset space $SO_0(2\nu, 1)/SO(2\nu)$.

The Iwasawa decomposition of $SO_0(2\nu, 1)$ can be written uniquely as $g = nak$, that is,

$$g = \begin{pmatrix} 1 + (\lambda^2/2)\mathbf{x}^2 & \lambda\mathbf{x}^t & -(\lambda^2/2)\mathbf{x}^2 \\ \lambda\mathbf{x} & \mathbf{1} & -\lambda\mathbf{x} \\ (\lambda^2/2)\mathbf{x}^2 & \lambda\mathbf{x}^t & 1-(\lambda^2/2)\mathbf{x}^2 \end{pmatrix} \times \begin{pmatrix} \cosh\lambda x^{2\nu} & 0 & \sinh\lambda x^{2\nu} \\ 0 & \mathbf{1} & 0 \\ \sinh\lambda x^{2\nu} & 0 & \cosh\lambda x^{2\nu} \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & & \\ & k & \\ 0 & & \end{pmatrix}, \tag{3.3}$$

where $k \in K = SO(2\nu)$ is the maximal compact subgroup of $SO_0(2\nu, 1)$; $a(x^{2\nu}) \in A$, the Abelian subgroup of $SO_0(2\nu, 1)$; $n(\mathbf{x}) \in N$, the nilpotent subgroup of $SO_0(2\nu, 1)$, where $\mathbf{x}$ is the column vector $(x^1, x^2, \ldots, x^{2\nu-1})$, $\mathbf{x}^t$ its transpose, and $\mathbf{x}^2 = (x^1)^2 + (x^2)^2 + \cdots + (x^{2\nu-1})^2$; and λ is a universal constant of dimension $[\text{length}]^{-1}$. The quantity λ enters the theory as a fundamental constant in the commutation relations defining the de Sitter group, just like the speed of light enters within the Poincaré group. It is also known that the de Sitter group contracts [with respect to $SO(2\nu)$] onto the Euclidean group in the Inonu–Wigner contraction limit $\lambda \to 0$.

It is easy to see that geometrically the Iwasawa decomposition is merely telling us that the hyperboloid (3.2) can be parametrized by coordinates x^α as follows:

$$\begin{aligned} \xi^0 &= \cosh\lambda x^{2\nu} + (\lambda^2/2)\mathbf{x}^2 e^{-\lambda x^{2\nu}}, \\ \xi^i &= \lambda x^i e^{-\lambda x^{2\nu}}, \quad i = 1, 2, \ldots, 2\nu-1, \\ \xi^{2\nu} &= \sinh\lambda x^{2\nu} - (\lambda^2/2)\mathbf{x}^2 e^{-\lambda x^{2\nu}}, \end{aligned} \tag{3.4}$$

where $-\infty < x^\alpha < \infty$, $\alpha = 1, 2, \ldots, 2\nu$. This coordinate system we shall call the x or horospherical coordinate system.

Some of the useful geometrical properties of Lobachevskian space are to be found in Ref. 5, though we shall describe those that we will need as we go along.

Let us choose two points P and Q on $\Lambda^{2\nu}$ with coordinates ξ^a and η^a, $a = 0, 1, 2, \ldots, 2\nu$, respectively. The distance, $r(P, Q) > 0$, between P and Q is given by

$$\cosh\lambda r = [\xi, \eta]. \tag{3.5}$$

Let B be the boundary of $\Lambda^{2\nu}$, i.e., $B = \{\zeta \in \Lambda^{2\nu}: [\zeta, \zeta] = 0\}$. The parallel geodesics in $\Lambda^{2\nu}$ are by definition geodesics originating from the same point ζ on the boundary B of $\Lambda^{2\nu}$.

A horosphere with normal $\zeta \in B$ is by definition an orthogonal hypersurface to the family of all parallel geodesics corresponding to ζ. Hence a horosphere in $\Lambda^{2\nu}$ is the non-Euclidean analog of a hyperplane in $R^{2\nu}$. The vector $\zeta \in B$ determines a whole family of horospheres in $\Lambda^{2\nu}$ that can be uniquely characterized by the horosphere through the point $\mathring{\xi} = (1, 0, 0, \ldots, 0)$ (i.e., through the origin 0 of the horospherical coordinate system), which is given by the equation $[\zeta, \xi] = 1$, $\xi \in \Lambda^{2\nu}$. The surfaces

$$[\zeta, \xi] = c = \text{const} \tag{3.6}$$

are, for each c, horospheres parallel to the one passing through $\mathring{\xi}$. We note that the distance τ from the origin 0 to the horosphere with normal $\zeta \in B$, passing through ξ is

$$\tau = \lambda^{-1} \ln[\zeta, \xi].\tag{3.7}$$

Let us now give a very sketchy account of the class I (or scalar) representations of $SO_0(2\nu, 1)$, emphasizing some points which are relevant for the discussion of Lobachevskian Fock space.

The class I UIR of $SO_0(2\nu, 1)$ are known[4] to form three series: principal, supplementary, and discrete. We restrict the discussion to the principal series of class I UIR of $SO_0(2\nu, 1)$, since not all the series of the class I (UIR) actually appear in the harmonic expansion of the quasiregular representations of $SO_0(2\nu, 1)$ in $\mathscr{L}^2(\Lambda^{2\nu})$ (see below). It is also known that (see, e.g., Ref. 6 and references to earlier work cited there) only the principal series of the class I UIR of $SO_0(2\nu, 1)$ goes over in the Inonu–Wigner contraction limit into the class I UIR of $ISO(2\nu)$.

The principal series of the class I UIR of $SO_0(2\nu, 1)$ characterized by the parameter $\rho (0<\rho<\infty)$ can be realized on the Hilbert space $\mathscr{L}^2(S^{2\nu-1})$ of square-integrable functions $f(\mathbf{n})$ over the unit sphere $S^{2\nu-1}$ (here and in the following $\mathbf{n}$ denotes a unit vector in $R^{2\nu}$). In this realization, we have

$$T^\rho(g) f(\mathbf{n}) = \left[\sum_{\alpha=1}^{2\nu} (g^{-1})^0_\alpha n^\alpha + (g^{-1})^0_0 \right]^{1/2-\nu+i\rho} f(\mathbf{n}_{g^{-1}}),\tag{3.8}$$

where $\mathbf{n}_{g^{-1}}$ is the unit vector defined by

$$n^\alpha_{g^{-1}} = \frac{\Sigma^{2\nu}_{\beta=1} (g^{-1})^\alpha_\beta n^\beta + (g^{-1})^\alpha_0}{\Sigma^{2\nu}_{\beta=1} (g^{-1})^0_\beta + (g^{-1})^0_0}, \quad \alpha = 1, 2, \ldots, 2\nu.\tag{3.9}$$

If λ and ρ converge to zero and infinity, respectively, in such a way that $\lambda\rho \to p$, the operator (3.8) will also converge to the representation operator (2.8) of the contracted group $ISO(2\nu)$.

We conclude this section with the definition of a Fourier transform on $\Lambda^{2\nu}$ (see Ref. 4).

Let $f \in \mathscr{L}^2(\Lambda^{2\nu})$. Set

$$\hat{f}(\rho;\mathbf{n}) = (2\pi)^\nu \int_{\Lambda^{2\nu}} f(\xi) [\zeta, \xi]^{1/2-\nu+i\rho} d\xi,$$

where $d\xi = d\xi^1 d\xi^2 \ldots d\xi^{2\nu}/\xi^0$ is the invariant measure on $\Lambda^{2\nu}$ and $\xi = (1, \mathbf{n}) \in B$. Then

$$f(\xi) = \frac{1}{(2\pi)^n} \int_{R^+} \int_{S^{2\nu-1}} \hat{f}(\rho;\mathbf{n}) [\zeta, \xi]^{1/2-\nu-i\rho} \times |\Gamma(\nu - \tfrac{1}{2} + i\rho)/\Gamma(i\rho)|^2 \, d\rho \, d\mathbf{n},\tag{3.10}$$

where $d\mathbf{n}$ is the Euclidean measure on $S^{2\nu-1}$ and Γ is a gamma function.

Furthermore, we have the Plancherel formula

$$\int_{\Lambda^{2\nu}} |f(\xi)|^2 d\xi = \int_{R^+} \int_{S^{2\nu-1}} |\hat{f}(\rho;\mathbf{n})|^2 \left| \frac{\Gamma(\nu - \frac{1}{2} + i\rho)}{\Gamma(i\rho)} \right|^2 d\rho \, d\mathbf{n},$$

and the decomposition of the quasiregular representation $T(g)$: $f(\xi) \mapsto f(g^{-1}\xi)$ onto the direct integral of irreducible representations (3.8) are given by

$$T(g) \to \hat{T}(g) = \int_{R^+} T^\rho(g) \left| \frac{\Gamma(\nu - \frac{1}{2} + i\rho)}{\Gamma(i\rho)} \right|^2 d\rho.\tag{3.11}$$

Note that the function

$$E_{\rho,\zeta}: \xi \to [\zeta, \xi]^{1/2-\nu+i\rho}, \quad \xi \in \Lambda^{2\nu}, \quad \zeta \in B,\tag{3.12}$$

has the properties of the plane waves on $R^{2\nu}$. Indeed, (i) $E_{\rho,\zeta}$ is constant on each horosphere (3.6) with normal ζ; and (ii) $E_{\rho,\zeta}$ is the eigenfunction of the Laplace–Beltrami operator on $\Lambda^{2\nu}$ (see Ref. 4).

It is useful to observe that the function $E_{\rho,\zeta}$ can be written in the form[5]

$$e^{(1/2-\nu+i\rho)\lambda\tau},$$

where τ is distance from the origin 0 to the horosphere with normal ζ, passing through ξ [see Eq. (3.7)]. This form is clearly analogous to the flat space-time plane waves (2.9). [In Eq. (2.9), of course, $(\mathbf{n},\mathbf{x})$ gives the distance from the origin to the hyperplane with normal $\mathbf{n}$, passing through $\mathbf{x}$.] Notice now that if we take the limit $\lambda \to 0$, $\rho \to \infty$ keeping $\lambda\rho = p$ fixed, we gain the flat Euclidean space-time result, with p interpreted as the momentum flowing through the plane wave.

That is,

$$[\zeta, \xi]^{1/2-\nu+ip/\lambda} \underset{(\lambda\to 0)}{\longrightarrow} e^{-ip(\mathbf{n},\mathbf{x})}\tag{3.13}$$

(in horospherical coordinates). Furthermore, it is not difficult, in fact, to see that in this limit the operators $\hat{T}(g)$ [see Eq. (3.11)] go over into the representation operators (2.7) of the contracted group $ISO(2\nu)$. Therefore we shall choose the Lobachevskian one-particle space $\mathscr{H}^{(1)}$ (see below) to be the Hilbert space supporting the unitary (highly reducible) representation $\hat{T}(g)$ of $SO_0(2\nu,1)$.

IV. FREE LOBACHEVSKIAN BOSON FIELDS

In this section we introduce free Lobachevskian boson fields, which will be related in the flat-space limit to the Euclidean fields constructed in the Sec. II. These Lobachevskian fields act in a space $\mathscr{H}$, which again has Fock space structure. This Fock space construction will be carried out in step-by-step analogy with the Euclidean Fock space construction.

We choose the Lobachevskian one-particle space $\mathscr{H}^{(1)}$ to be the Hilbert space of all functions $F(p;\mathbf{n})$ for which

$$\|F\| \equiv \int_{R^+} \int_{S^{2\nu-1}} |F(p;\mathbf{n})|^2 \times \left| \frac{\Gamma(\nu - 1/2 + ip/\lambda)}{\Gamma(ip/\lambda)} \right|^2 dp \, d\mathbf{n} < \infty.\tag{4.1}$$

The Hilbert space $\mathscr{H}^{(1)}$ carries a unitary (highly reducible) representation of $SO_0(2\nu,1)$ defined by

$$U(g)F(p;\mathbf{n}) = \left[\sum_{\alpha=1}^{2\nu} (g^{-1})^0_\alpha n^\alpha + (g^{-1})^0_0 \right]^{-\nu+1/2+ip/\lambda} F(p;\mathbf{n}_{g^{-1}}),\tag{4.2}$$

where $\mathbf{n}_{g^{-1}}$ is given by Eq. (3.8). [As is easily seen, (4.2) is nothing but the unitary representation $\hat{T}(g)$ of $SO_0(2\nu,1)$ defined by Eq. (3.11).] Then the Lobachevskian Fock space

$\mathscr{H}$ for scalar bosons is defined to be the Hilbert space completion of the symmetric tensor algebra over $\mathscr{H}^{(1)}$,

$$\mathscr{H} = \mathbb{C} \oplus \mathscr{H}^{(1)} \oplus (\mathscr{H}^{(1)} \otimes_S \mathscr{H}^{(1)}) \oplus \cdots .$$

The vacuum is denoted by $|0\rangle$. In the standard fashion we introduce boson annihilation and creation operators $A(p;\mathbf{n})$ and $A^*(p;\mathbf{n})$, satisfying the commutation relations

$$[A(p;\mathbf{n}), A^*(p';\mathbf{n}')]_- = \lambda^{1-2\nu}\left|\frac{\Gamma(\nu-\frac{1}{2}+ip/\lambda)}{\Gamma(ip/\lambda)}\right|^{-2} \times \delta(p-p')\delta^{2\nu-1}(\mathbf{n}-\mathbf{n}'), \quad (4.3)$$

all other commutators vanishing. Hence, one has a unitary representation of $SO_0(2\nu,1)$ on $\mathscr{H}$ defined by

$$U(g)|0\rangle = |0\rangle, \quad (4.4)$$

$$U(g)A^*(p;\mathbf{n})U^{-1}(g) = \left[\sum_{\alpha=1}^{2\nu} g_\alpha^0 n^\alpha + g_0^0\right]^{-\nu+1/2-ip/\lambda} A^*(p;\mathbf{n}_g). \quad (4.5)$$

We are now prepared to define the Lobachevskian scalar boson field. We want to form the free field by taking linear combinations of creation and annihilation operators. Equation (2.1) tells us to do this by setting the field equal to the de Sitter-invariant Fourier transform [see Eq. (3.10)] of these operators:

$$\Phi(\xi) = \frac{\lambda^{2\nu-1}}{(2\pi)^\nu}\int_{R^+}\int_{S^{2\nu-1}} \frac{1}{\sqrt{p^2+(m+\lambda/2)^2}} \times [A(p;\mathbf{n})[\zeta,\xi]^{-\nu+1/2-ip/\lambda} + A^*(p;\mathbf{n})[\zeta,\xi]^{-\nu+1/2+ip/\lambda}] \times \left|\frac{\Gamma(\nu-\frac{1}{2}+ip/\lambda)}{\Gamma(ip/\lambda)}\right|^2 dp\, d\mathbf{n}. \quad (4.6)$$

It is easy to verify that the field transforms as

$$U(g)\Phi(\xi)U^{-1}(g) = \Phi(g\xi). \quad (4.7)$$

The verification of property (4.7) is based on the relation

$$d\mathbf{n}_g = \left[\sum_{\alpha=1}^{2\nu} g_\alpha^0 n^\alpha + g_0^0\right]^{1-2\nu} d\mathbf{n}.$$

Obviously Lobachevskian boson fields are commutative, i.e.,

$$[\Phi(\xi), \Phi(\xi')]_- = 0, \quad \text{for all } \xi,\xi' \in \Lambda^{2\nu}. \quad (4.8)$$

In verifying (4.8) one has to use the formulas (A5) and (A6) (see the Appendix).

The two-point Green's function $G_L(\xi,\xi')$ is given by

$$G_L(\xi,\xi') \equiv \langle 0|\Phi(\xi)\Phi(\xi')|0\rangle = \frac{\lambda^{2\nu-1}}{(2\pi)^{2\nu}}\int_{R^+}\int_{S^{2\nu-1}} \frac{1}{p^2+(m+\lambda/2)^2} \times [\zeta,\xi]^{-\nu+1/2-ip/\lambda} \times [\zeta,\xi']^{-\nu+1/2+ip/\lambda}\left|\frac{\Gamma(\nu-1/2+ip/\lambda)}{\Gamma(ip/\lambda)}\right|^2 dp\, d\mathbf{n}.$$

Let us now calculate the explicit form of $G_L(\xi,\xi')$. After integration over the angles we find that [see Eqs. (A5) and (A7)]

$$G_L(\xi,\xi') = (-)^{\nu-1}\lambda^{2\nu-2}(2\pi)^{-\nu}(\sinh\lambda r)^{1-\nu} \times \int_0^\infty \frac{p\tanh(\pi p/\lambda)}{p^2+(m+\lambda/2)^2} P^{\nu-1}_{-1/2+ip/\lambda}(\cosh\lambda r)dp, \quad (4.9)$$

where $\cosh\lambda r = [\xi,\xi']$. The remaining p integral can be computed using the formulas 7.213 of Ref. 7 and 3.6(3) and 3.6(4) of Ref. 8, so that the expression (4.9) reduces to

$$G_L(\xi,\xi') = (-)^{\nu-1}\frac{\lambda^{2\nu-2}}{(2\pi)^\nu\sinh^{\nu-1}\lambda r} Q^{\nu-1}_{m/\lambda}(\cosh\lambda r), \quad (4.10)$$

where $Q^{\nu-1}_\sigma$ is the Legendre function of the second kind.

Furthermore we can check that the Euclidean Green's function is indeed a limiting case of the Lobachevskian Green's function, i.e., that

$$G_L(\xi,\xi') \underset{(\lambda\to 0)}{\to} G_E(\mathbf{x},\mathbf{x}'),$$

since

$$\lim_{\lambda\to 0}(-\lambda)^{\nu-1}Q^{\nu-1}_{m/\lambda}(\cosh\lambda r) = m^{\nu-1}K_{\nu-1}(mr)$$

[see, e.g., Eq. 7.8(4) of Ref. 9].

It follows from (3.13) and

$$\lim_{\lambda\to 0}\lambda^{2\nu-1}\left|\frac{\Gamma(\nu-\frac{1}{2}+ip/\lambda)}{\Gamma(ip/\lambda)}\right|^2 = p^{2\nu-1}$$

[see, e.g., Eq. 1.14(6) of Ref. 8] that the Lobachevskian field (4.6) goes over in the flat-space limit $\lambda\to 0$ into the Euclidean field (2.1).

V. INTERACTING FIELDS

Our next step will be the more important (in field theory) attempt to construct the N-point Green's functions of interacting fields. In analogy to the Euclidean case, we define the Lobachevskian Green's functions by

$$G_L(\xi_1,\xi_2,\ldots,\xi_N) = \frac{\langle 0|\Phi(\xi_1)\Phi(\xi_2)\cdots\Phi(\xi_N)e^{-V[\Phi]}|0\rangle}{\langle 0|e^{-V[\Phi]}|0\rangle} \quad (5.1)$$

(ordinary operator multiplication!), where $V[\Phi]$ is the Lobachevskian action needed to describe the field interactions. For example, the Lobachevskian action for the cubic boson self-interaction is given by

$$V[\Phi] = \kappa\int_{\Lambda^{2\nu}} :\Phi^3(\xi):d\xi, \quad (5.2)$$

where κ is interaction constant and : : denotes Wick ordering.

The Green's functions in perturbation theory can be calculated from (5.1) by using Wick's theorem (which, of course, remains true for Lobachevskian fields) as usual to derive the Feynman rules.

(a) For each internal vertex, include a factor (-1) times whatever coefficients appear with the fields in $V[\Phi]$. For example, each vertex arising from (5.2) will contribute a factor $(-\kappa)$.

(b) For each line joining ξ to ξ', include a factor (4.10).

(c) Integrate over the position of each internal vertex.

But the numerical evaluation is much more complicated than in flat space, although the labor can be reduced by using the geometrical properties of the space, such as the group of motions. We hope to return to the questions of divergences and renormalization in a future publication.

APPENDIX

Here we present the calculation of the integral

$$I=\int_{S^{2\nu-1}}[\zeta,\xi]^{1/2-\nu+i\rho}[\zeta,\xi']^{1/2-\nu-i\rho}\,d\mathbf{n}, \tag{A1}$$

where $\xi,\xi'\in\Lambda^{2\nu}$, $\zeta=(1,\mathbf{n})\in B$, and $d\mathbf{n}$ is the Euclidean measure on $S^{2\nu-1}$.

To evaluate the integral (A1), we expand the Lobachevskian plane waves (3.12) into the harmonic polynomials $\Xi_M(\mathbf{n})$, $M=(m_0,m_1,m_2,\dots,m_{2\nu-2})$ on $S^{2\nu-1}$ (see Ref. 4), viz.,

$$[\zeta,\xi]^{1/2-\nu+i\rho}=\sum_M t^{\rho}_{M0}(g_\xi)\Xi_M(\mathbf{n}) \tag{A2}$$

[see Eq. 10.3.2(5) of Ref. 4], where $t^{\rho}_{M0}(g_\xi)$, $0=(0,0,\dots,0)$ is the matrix elements of the class I principal series representations of the $SO_0(2\nu,1)$ group and g_ξ is given by $\xi=g_\xi\mathring{\xi}$. The harmonic polynomials $\Xi_M(\mathbf{n})$ satisfy the orthogonality and completeness relations,

$$\int_{S^{2\nu-1}}\Xi_M(\mathbf{n})\overline{\Xi}_{M'}(\mathbf{n})d\mathbf{n}=\frac{2\pi^\nu}{\Gamma(\nu)}\delta_{MM'} \tag{A3}$$

and

$$\sum_M\Xi_M(\mathbf{n})\overline{\Xi}_M(\mathbf{n}')=\delta(\mathbf{n}-\mathbf{n}'). \tag{A4}$$

By expanding in harmonic polynomials, and performing the $\int d\mathbf{n}$ using (A3), we find

$$I=\frac{2\pi^\nu}{\Gamma(\nu)}t^{\rho}_{00}(g_{\xi'}^{-1}g_\xi).$$

The matrix elements t^{ρ}_{00} can be expressed in terms of a Legendre function of the first kind [see Eq. 10.3.3(4) of Ref. 4].

Finally then, we find

$$\int_{S^{2\nu-1}}[\zeta,\xi]^{1/2-\nu+i\rho}[\zeta,\xi']^{1/2-\nu-i\rho}\,d\mathbf{n}$$
$$=(2\pi)^\nu(\sinh\lambda r)^{1-\nu}P^{1-\nu}_{-1/2+i\rho}(\cosh\lambda r), \tag{A5}$$

where $\cosh\lambda r=[\xi,\xi']$ [see Eq. (3.5)].

To finish, we mention some properties of the Legendre function:

$$P^{\mu}_{\sigma}(z)=P^{\mu}_{-1-\sigma}(z), \tag{A6}$$

$$P^{\mu}_{\sigma}(z)=\frac{\Gamma(\sigma+\mu+1)}{\Gamma(\sigma-\mu+1)}P^{-\mu}_{\sigma}(z),\quad \text{if } \mu=1,2,\dots\ . \tag{A7}$$

[1]E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[2]J. Schwinger, Proc. Natl. Acad. Sci. USA **44**, 956 (1958); Phys. Rev. **115**, 721 (1959).
[3]K. Osterwalder and R. Schrader, Helv. Phys. Acta **46**, 277 (1973).
[4]N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968).
[5]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[6]M. K. F. Wong and H-Y. Yeh, J. Math. Phys. **16**, 800 (1975).
[7]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[8]*Higher Transcendental Functions*, Vol. 1 of Bateman Manuscript Project, edited by A. Erdelyi (McGraw–Hill, New York, 1953).
[9]*Higher Transcendental Functions*, Vol. 2 of Bateman Manuscript Project, edited by A. Erdelyi (McGraw–Hill, New York, 1953).

Spherically symmetric perfect fluid solutions in isotropic coordinates

Joseph Hajj-Boutros
Physics Department, Lebanese University, Mansourieh El-Metn, P. O. Box 72, Lebanon

(Received 21 February 1985; accepted for publication 3 December 1985)

A method of generating solutions of Einstein's field equations for static fluid spheres is presented. This new method has a very clear advantage since it bypasses the direct resolution of the field equations. One of the solutions obtained is then studied in some detail; it verifies an equation of state of the form $p \simeq \alpha\rho + \beta\rho^{1+1/n}$.

I. INTRODUCTION

The study of fluid spheres in general relativity requires the solutions of Einstein's field equations.

Due to the nonlinearity of those equations it becomes difficult to obtain analytical solutions, in particular the problem of constructing a model of a static sphere of perfect fluid is often solved by numerical methods using the Tolman–Oppenheimer–Volkoff equation.[1–4]

Though a large number of numerically computed solutions are available in the literature the efficiency of exact solutions for giving a clear understanding of the internal structure of a spherical star cannot be denied.

The present paper has a twofold purpose; in the first place a method for generating new solutions for a static spherically symmetric distributions of matter from existing solutions is presented in isotropic coordinates. It can be compared to that given by the author in usual coordinates[5] and largely differs from those already existing in the literature.[6–8]

This method reduces the field equations (in isotropic coordinates) to a Riccati equation in λ' or ν', where $' \equiv d/dr$, λ and ν being the metric functions used in the line element:

$$ds^2 = e^\nu dt^2 - e^\lambda d\sigma^2,$$

where λ and ν are functions of r only. It is well known that if we know a particular solution for a Riccati equation then this equation reduces to a linear first-order differential equation.

Therefore a more general solution can be easily obtained. As particular solutions we use those already existing in the literature.

We deal here with what is essentially a Riccati equation in ν', in this case the new solution generated has the same density ρ as the old one.

An example dealing with the second case (Riccati equation in λ') is given in the Appendix though the solution obtained has not been studied.

The generation technique used here leads directly to several new solutions.

By studying the properties of one of the solutions obtained it then appears that the equation of state contains a barytropic term added to a polytropic one.

It may be used as an analytical model for studying the physical properties of a spherical star.

Finally, we consider a solution to be physically reasonable if pressure and density are monotonic decreasing functions of r throughout the star and regular in all the interior region of the star and if pressure goes to zero at a finite radius R at which we must match in a proper manner[9,10] the internal solution to the standard Schwarzschild[11] external one.

Furthermore the strong energy conditions must be satisfied.

II. FIELD EQUATIONS AND GENERATION TECHNIQUE

We begin by examining the nonvacuum static spherically symmetric Einstein's field equations for a line element of the form

$$ds^2 = e^\nu dt^2 - e^\lambda d\sigma^2, \tag{2.1}$$

$$d\sigma^2 = dr^2 + r^2 d\Omega^2, \tag{2.2}$$

$$d\Omega^2 = d\theta^2 + \sin^2(\theta) d\varphi^2. \tag{2.3}$$

The metric (2.1) is written in isotropic coordinates ν and λ are functions of r only. In order that the interior solution joins properly to the exterior Schwarzschild solution, i.e.,

$$ds^2 = \frac{(1 - M/2r)^2}{(1 + M/2r)^2} dt^2 - \left(1 + \frac{M}{2r}\right)^4 d\sigma^2, \tag{2.4}$$

at the boundary of the star $r = R$, we will require

$$\nu(R) = 2\ln((1 - M/2R)/(1 + M/2R)), \tag{2.5}$$

$$\nu'(R) = (2M/R^2)(1 - M^2/4R^2)^{-1}, \tag{2.6}$$

$$\lambda(R) = 4\ln(1 + M/2R), \tag{2.7}$$

$$\lambda'(R) = -(2M/R^2)(1 + M/2R)^{-1}, \tag{2.8}$$

M and R being, respectively, the mass and radius of the star as measured by a distant observer.

The field equations in isotropic coordinates read

$$8\pi p = e^{-\lambda}\left(\frac{\lambda'^2}{4} + \frac{\lambda'\nu'}{2} + \frac{\lambda' + \nu'}{r}\right), \tag{2.9}$$

$$8\pi p = e^{-\lambda}\left(\frac{\lambda''}{2} + \frac{\nu''}{2} + \frac{\nu'^2}{4} + \frac{\lambda' + \nu'}{2r}\right), \tag{2.10}$$

$$8\pi\rho = e^{-\lambda}\left(\lambda'' + \frac{(\lambda')^2}{4} + \frac{2\lambda'}{r}\right) \tag{2.11}$$

$$\left(' = \frac{d}{dr}\right).$$

The energy-momentum tensor can be written as

$$T_{ij} = (\rho + p)u_i u_j + p g_{ij}, \quad u_i u^i = -1.$$

(In this paper units are chosen so that $G = c = 1$.) Hence we have

$$T_0^0 = \rho, \tag{2.12}$$

$$-T_1^1 = -T_2^2 = -T_3^3 = P, \tag{2.13}$$

 0022-2488/86/051363-04$02.50

$$T^i_j = 0, \quad \text{for } i \neq j \quad (i = 0,1,2,3), \tag{2.14}$$

where p is the pressure and ρ the mass density.

From the isotropy of pressure [(2.9) and (2.10)] we get

$$\lambda'' + \nu'' + \tfrac{1}{2}(\nu'^2 - \lambda'^2) - \lambda'\nu' - (1/r)(\lambda' + \nu') = 0. \tag{2.15}$$

Equation (2.15) is of a Riccati type either in λ' or ν'.

Now we proceed as follows.

Case II.A: $\lambda' = L$. Equation (2.15) becomes

$$L' - \tfrac{1}{2}L^2 - L(\nu' + 1/r) + \nu'' + \tfrac{1}{2}\nu'^2 - \nu'/r = 0. \tag{2.16}$$

Suppose now that $L(\lambda_0)$ and ν_0 are particular solutions for (2.16), then we have

$$L = L_0 + 1/H_1, \tag{2.17}$$

$$\frac{1}{H_1} = \frac{re^{(\lambda_0 + \nu_0)}}{[\int -\frac{1}{2}re^{(\lambda_0+\nu_0)}\,dr + C]}, \tag{2.18}$$

where L is the new solution for (2.16) with $\nu = \nu_0$ and C is a constant of integration.

From (2.17) and (2.18), we have

$$\lambda = \lambda_0 + \ln(C_1/B_1(r,c))^2. \tag{2.19}$$

Here

$$B_1(r,c) = \int -\frac{1}{2}re^{(\lambda_0+\nu_0)}\,dr + C. \tag{2.20}$$

Equation (2.19) yields

$$e^\lambda = e^{\lambda_0}[C_1/B_1(r,c)]^2, \tag{2.21}$$

C being a constant of integration. Thus from the couple of known metric functions (λ_0,ν_0) our generation technique allows us to obtain the new one (λ,ν_0). By using (2.9), (2.11), and (2.19) we can conclude that the pressure and density obtained for the new solution generated are different from those corresponding to the old solution.

Case II.B: $\nu' = N$. In this case Eq. (2.15) becomes

$$N' + \frac{1}{2}N^2 - \lambda'N - \frac{N}{r} + \lambda'' - \frac{\lambda'^2}{2} - \frac{\lambda'}{r} = 0. \tag{2.22}$$

If ν_0,λ_0 are known solutions for (2.22), then we have

$$N = N_0 + 1/H_2, \tag{2.23}$$

$$\frac{1}{H_2} = \frac{re^{(\lambda_0 - \nu_0)}}{[\int \frac{1}{2}re^{(\lambda_0-\nu_0)}\,dr + c]}, \tag{2.24}$$

where N is a new solution for (2.22) and c is a constant of integration.

Now integrating (2.23) we obtain

$$\nu = \nu_0 + \ln[B_2(r,c)/C_1]^2, \tag{2.25}$$

where

$$B_2(r,c) = \int \frac{1}{2}re^{(\lambda_0-\nu_0)}\,dr + C \tag{2.26}$$

and

$$e^\nu = e^{\nu_0}[B_2(r,c)/C_1]^2, \tag{2.27}$$

C_1 being a constant of integration.

Hence, starting from the couple (ν_0,λ_0) the generation technique allows us to obtain the new (ν,λ_0). We notice that the new mass density ρ obtained here is equal to the old one, i.e., $\rho = \rho_0$ (λ_0 stays invariable). We conclude by noting that in the application our choice between formulas (2.21) and (2.26) is motivated by the following arguments.

(1) Formula (2.26) is adopted if the density ρ_0 for the known solution is physically reasonable.

(2) Formula (2.21) is adopted if ρ_0 is not physically reasonable.

Finally we notice that if integration by means of (2.26) is difficult we can go to (2.21) and vice versa.

III. SOLUTION CONTAINING A POLYTROPIC TERM IN ITS EQUATION OF STATE

We use here as a particular solution the Bayin[12] solution VI. First we show that the pressure evaluated by Bayin contains an error, hence his equation of state cannot contain a polytropic term. In fact, applying formulas (2.11) and (2.9) for the solution obtained by Bayin, we get

$$8\pi\rho = -(C_0r^2 + C_1)^{-b/(1-c)} \times \left[\frac{(1-c)(6bC_0C_1 + 2bC_0^2r^2) + b^2C_0^2r^2}{(1-c)^2(C_0r^2 + C_1)^2}\right], \tag{3.1}$$

$$8\pi p = (C_0r^2 + C_1)^{-b/(1-c)}\left[\frac{C_0^2b^2r^2 - 2abC_0^2r^2 + 2C_0(b-a)(1-C)(C_0r^2 + C_1)}{(1-c)^2(C_0r^2 + C_1)^2}\right]. \tag{3.2}$$

Expression (3.2) differs from that given in (12). Now combining (3.1) and (3.2) we get ($C_1 = 0$)

$$p = \rho[(2ab - b^2) + 2(a-b)(1-c)]/[b^2 + 2b(1-c)]. \tag{3.3}$$

Hence the equation of state is of the form

$$p = \alpha_0\rho, \tag{3.4}$$

where

$$\alpha_0 = [(2ab - b^2) + 2(a-b)(1-c)]/[b^2 + 2b(1-c)].$$

This completely differs from the result already obtained by Bayin himself.

We now show that the application of our technique to Bayin's VI solution gives a new line element for which the equation of state may be nearly polytropic. Bayin's VI solution reads

$$ds^2 = (C_0r^2 + C_1)^{-a/(1-a)}\,dt^2 - (C_0r^2 + C_1)^{b/(1-a)}\,d\sigma^2. \tag{3.5}$$

Applying the formulas (2.27) and (2.26), we get

$$B_2(r,c) = \frac{(C_0r^2 + C_1)^{(a+b)/(1-\alpha)+1}}{4C_0[(a+b)/(1-\alpha)+1)} + C, \tag{3.6}$$

$$e^{\nu} = C_2^2(C_0r^2 + C_1)^{-a/(1-\alpha)} \times \left[\frac{(C_0r^2 + C_1)^{(a+b)/(1+\alpha)+1}}{4C_0[(a+b)/(1-\alpha)+1)} + C\right]^2. \tag{3.7}$$

Here α must satisfy the relation

$$\alpha = [\tfrac{1}{2}b^2 - a^2) - ab + b - a]/(b-a), \tag{3.8}$$

where C_2 and C are two constants of integration. Setting now

$$a_0 = \frac{a}{1-\alpha}, \quad b_0 = \frac{b}{1-\alpha}, \quad \alpha_0 = a_0 + b_0 + 1,$$
$$\alpha_1 = 4CC_0\alpha_0, \quad W = (C_0r^2 + C_1). \tag{3.9}$$

We obtain

$$ds^2 = \frac{C_2^2 W^{-a_0}}{16C_0^2\alpha_0^2}(W^{\alpha_0} + \alpha_1)^2 dt^2 - W^{b_0} d\sigma^2. \tag{3.10}$$

We can use formulas (2.9) and (2.11) for evaluating the pressure P and the energy density ρ and we obtain

$$8\pi P = w^{-b_0}\left[\frac{(b_0^2 - 2a_0b_0)w'^2 + 8C_0w(b_0 - a_0)}{4w^2} + \frac{b_0\alpha_0w'^2w^{\alpha_0-1} + 4\alpha_0C_0w^{\alpha_0}}{w(w^{\alpha_0} + \alpha_1)}\right], \tag{3.11}$$

$$8\pi\rho = w^{-b_0}[(4b_0rww'' - 3b_0w'^2 + 8b_0ww')/4rw^2]. \tag{3.12}$$

The pressure and the density have finite values at the center of the star ($r = 0$).

We look now for the equation of state verified by our solution; in order to compare the solution which we obtain with that of Bayin's. We must take the constant $C_1 = 0$. (Note in this case the pressure P and the density ρ become singular at the center $r = 0$.) In this case we have $w = C_0r^2$, $w'^2 = 4C_0w$, and $w'' = 2C_0$. Using formulas (3.11) and (2.12) we obtain

$$\frac{P}{\rho} = \frac{b_0(b_0 - 2a_0) + 2C_0(b_0 - a_0)}{3b_0C_0} + \frac{2\alpha_0(b_0 + 1)}{3b_0(1 + \alpha_1w^{-\alpha_0})}. \tag{3.13}$$

The density ρ is given by

$$\rho = (3b_0C_0/2\pi)w^{-b_0-1}. \tag{3.14}$$

From (3.14) we obtain

$$w = (2\pi\rho/3b_0C_0)^{-1/(b_0+1)}. \tag{3.15}$$

Setting

$$\gamma = [b_0(b_0 - 2a_0) + 2C_0(b_0 - a_0)]/3b_0C_0$$

and using formula (3.15) in (3.13), we get

$$\frac{P}{\rho} = \gamma + \frac{2\alpha_0(b_0+1)}{3b_0} \times \left[\frac{1}{1 + \alpha_1(2\pi\rho/3b_0C_0)^{[1/(b_0+1)]}}\right]. \tag{3.16}$$

Now choosing α_1 (c being an arbitrary constant at our disposal) in such a manner that

$$\alpha_1(2\pi\rho/3b_0C_0)^{1/(b_0+1)} \ll 1,$$

we obtain

$$\frac{P}{\rho} \simeq \gamma + \frac{2\alpha_0(b_0+1)}{3b_0} - \frac{2\alpha_0\alpha_1(b_0+1)}{3b_0}\left(\frac{2\pi\rho}{3b_0C_0}\right)^{1/(b_0+1)}. \tag{3.17}$$

We now can set

$$1/(b_0 + 1) = 1/n, \tag{3.18}$$

where n is a positive integer such that $n > 1$, and we obtain

$$P \simeq \left(\gamma + \frac{2\alpha_0(b_0+1)}{3b_0}\right)\rho - \frac{2\alpha_1\alpha_0(b_0+1)}{3b_0}\left(\frac{2\pi\rho}{3b_0C_0}\right)^{1/n+1}. \tag{3.19}$$

The choice $n > 1$ implies

$$b_0 + 1 > 1, \tag{3.20}$$

hence

$$b/(1-\alpha) > 0.$$

The above inequality can always be satisfied (b_0 and α being arbitrary constants) ($\alpha \neq 1$). For $b_0 \gg 2a_0$ and α_0 positive, and if we choose $\alpha_1 < 0$, the pressure and density ρ become positive. The term

$$\frac{2\alpha_1\alpha_0(b_0+1)}{3b_0}\left(\frac{2\pi\rho}{3b_0C_0}\right)^{1/n+1}$$

could be neglected compared to

$$\gamma + 2\alpha_0(b_0+1)/3b_0,$$

hence if we restrict ourselves to

$$0 < \gamma + \frac{2\alpha_0(b_0+1)}{3b_0} < 1,$$

we can satisfy the strong energy conditions

$$-\rho < p < \rho, \quad \rho > 0. \tag{3.21}$$

Going back now to the formula (3.11) we can choose α_1 in such a manner so as to drop P to zero at distance $r = R$. (Note here that in the case of Bayin's solution the pressure P is always nonzero except for $r \to \infty$.)

Thus the boundary of the star can be choosing for $r = R$ and we can match our internal solution to the external Schwarzschild one at this boundary ($r = R$) and we get

$$\frac{-2aC_2^2C_0R}{1-\alpha}(C_0R^2 + C_1)^{-a/(1-\alpha)-1} \times \left\{\frac{(C_0R^2 + C_1)^{(a+b)/(1-\alpha)}}{4C_0[(a+b)/(1-\alpha)+1)} + C\right\}^2 + C_2^2R(C_0R^2 + C_1)^{b/(1-\alpha)} \times \left\{\frac{(C_0R^2 + C_1)^{(a+b)/(1-\alpha)+1}}{4[(a+b)/(1-\alpha)+1)} + C\right\} = \frac{2M}{R^2}\left(1 - \frac{M^2}{4R^2}\right)^{-1}, \tag{3.22}$$

$$\frac{b}{(1-\alpha)}\ln(C_0R^2 + C_1) = 4\ln\left(1 + \frac{M}{2R}\right), \tag{3.23}$$

$$\frac{2bC_0R}{(1-\alpha)(C_0R^2+C_1)} = \frac{-2M}{R^2}\left(1+\frac{M}{2R}\right)^{-1}, \quad (3.24)$$

$$\frac{C_2^2(C_0R^2+C_1)^{-\alpha_0}}{16C_0^2\alpha_0^2}[(C_0R^2+C_1)^{\alpha_0}+\alpha_1]^2 = \left[\frac{1-M/2R}{1+M/2R}\right]^2. \quad (3.25)$$

The above equations determine the constants C_1, C_2, C_0, and α as functions of M and R.

IV. DISCUSSION

In a recent paper, Collins and Wainwright[13] enumerated all perfect fluid solutions to Einstein's field equations, which have the following properties: (1) the fluid has nonzero expansion ($\theta \neq 0$); (2) the fluid is irrotational ($\omega_{ij} = 0$); (3) the fluid is shear-free ($\sigma_{ij} = 0$); and (4) the equation of state that satisfies the above solutions is of the form $p = p(\rho)$; with $p + \rho \neq 0$.

As a result of their study[13] one can see that among the solutions satisfying conditions (1)–(4) there are some possessing spherical symmetry.[14]

The latter have the following line element:

$$ds^2 = \frac{1}{U^2} \times \left[-\frac{U'^2}{At+B}dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)\right]. \quad (4.1)$$

Formula (4.1) is obviously written in isotropic coordinates.

The metric function U depends only on the variable $U = t + r^2$, furthermore U' satisfies the relation

$$U'^2 = \tfrac{2}{3}U^3 - A/3, \quad (4.2)$$

where A and B are two constants such that

$$A^2 + B^2 \neq 0, \quad (4.3)$$

the prime denotes differentiation with respect to U. By introducing (4.2) into (4.1) and choosing $A = 0$, we obtain

$$ds^2 = \frac{2U}{3B}dt^2 + \frac{1}{U^2}[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)]. \quad (4.4)$$

We can now compare our solution (3.9) with that given by Collins and Wainwright.[13] Our function metrics depend on the variable $(C_0r^2 + C_1)$, where C_0 and C_1 are two arbitrary constants arising from a second-order differential equation in the variable r. Without loss of generality we can choose $C_1 = C_1(t)$, where t is the time; thus our solution can be related closely to that given by Wyman.[14]

APPENDIX: GENERALIZED BUCHDAHL SOLUTION

Using the Buchdahl[15] solution, i.e.,

$$ds^2 = ((1-f)/(1+f))^2dt^2 - (1+f)^4d\sigma^2, \quad (A1)$$

where

$$f = A/\sqrt{1+Kr^2}, \quad (A2)$$

as a particular solution for the Riccati equation (2.15), we get

$$B_1(r,c) = \frac{A^2}{2K}\int\frac{(1+f^4-2f^2)}{f^3}df. \quad (A3)$$

By integration we obtain

$$B_1(r,c) = \frac{A^2}{2K}\left[-\frac{1}{2f^2}+\frac{f^2}{2}-2\ln(f)\right]+C. \quad (A4)$$

Hence

$$e^\lambda = (1+f)^4 \times\left[\frac{C_1}{(A^2/2K)\{-1/2f^2+f^2/2-2\ln(f)\}+C}\right]^2 \quad (A5)$$

and $\nu = \nu_0$, where C and C_1 are two constants of integration. Then the new line element reads

$$ds^2 = ((1-f)/(1+f))^2dt^2 - (1+f)^4 \times\left\{\frac{C_1}{(A^2/2K)[-1/2f^2+f^2/2-2\ln(f)]+C}\right\}^2 \times d\sigma^2. \quad (A6)$$

[1] R. C. Tolman, *Relativity, Thermodynamics, and Cosmology* (Oxford U.P., New York, 1934).
[2] R. C. Tolman, Phys. Rev. **55**, 364 (1939).
[3] G. M. Volkoff, Phys. Rev. **44**, 413 (1939).
[4] J. R. Oppenheimer and G. M. Volkoff, Phys. Rev. **55**, 374 (1939).
[5] J. Hajj-Boutros, "A method for generating exact solutions of Einstein's field equation," in *Lecture Notes in Physics, Gravitation, Geometry and Relativistic Physics*, Vol. 212 (Springer, Berlin, 1984), p. 51.
[6] H. A. Buchdahl, Austral. J. Phys. **9**, 13 (1956).
[7] E. N. Glass and S. P. Goldman, J. Math. Phys. **19**, 856 (1978).
[8] B. W. Stewart, J. Phys. A **15**, 1799 (1982).
[9] A. Lichnerowicz (1955) *Theories relativistes de la gravitation et de l'electromagnetisme* (Masson, Paris, 1955), pp. 59–68.
[10] W. B. Bonnor and P. A. Vickers, Gen. Relativ. Gravit. **13**, 29 (1981).
[11] K. Schwarzschild, Sitz. Preuss. Akad. Wiss., 189 (1916).
[12] S. S. Bayin, Phys. Rev. D **18**, 2745 (1978).
[13] C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).
[14] M. Wyman, Phys. Rev. **70**, 396 (1946).
[15] H. A. Buchdahl, Astrophys. J. **140**, 1512 (1964).

Charged viscous fluids: Exact solutions

G. Papini and M. Weiss
Department of Physics and Astronomy, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

(Received 23 September 1985; accepted for publication 14 January 1986)

Exact solutions of Einstein's equations are given for some charged viscous fluids.

I. INTRODUCTION

The importance of viscosity in the theory of the early universe has been stressed repeatedly by various authors.[1] It stems largely from the possibility that viscosity affects the appearance of the initial singularity[2] and the expansion rate of the universe.[3] It can further be argued that since viscosity is virtually incompatible with an inflationary expansion,[4] at which time the universe behaves as a superfluid,[5] a transition from a preinflationary universe to an inflationary one involves a viscous fluid to superfluid phase change. This makes viscous fluid models interesting candidates for the preinflationary universe.

While a number of exact solutions of Einstein's equations are known for neutral fluids with bulk viscosity,[3,6] solutions involving shear viscosity are rather scanty.[7] The situation deteriorates rapidly with magnetoviscous fluids[8] and even more so with charged ones.[9]

The aim of this paper is to present additional solutions for charged fluids with bulk and shear viscosity. The energy momentum tensor for such systems is written in the form

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu - pg_{\mu\nu} + E_{\mu\nu} + \Lambda_{\mu\nu}, \tag{1}$$

where u_μ represents the fluid four-velocity

$$u_\alpha u^\alpha = 1, \tag{2}$$

and $\Lambda_{\mu\nu}$ is the viscosity tensor, defined by

$$\Lambda_{\mu\nu} = \zeta P_{\mu\nu} u^\lambda{}_{;\lambda} + \eta P_{\mu\alpha} P_{\nu\beta} (u^{\alpha;\beta} + u^{\beta;\alpha} - \tfrac{2}{3} g^{\alpha\beta} u^\lambda{}_{;\lambda}). \tag{3}$$

In (3), ζ and η are the coefficients of bulk and shear viscosity and $P_{\mu\alpha}$ is the projection tensor

$$P_{\mu\alpha} = g_{\mu\alpha} - u_\mu u_\alpha. \tag{4}$$

Finally $E_{\mu\nu}$ in (1) represents the Maxwell tensor for the electromagnetic field.

In the general spherically symmetric metric[10]

$$ds^2 = e^{2\nu} dt^2 - e^{2\lambda} dr^2 - Y^2 d\theta^2 - Y^2 \sin^2\theta \, d\phi^2, \tag{5}$$

which will be used, the unknown functions ν, λ, and Y depend exclusively on r and t. Since only the radial electric field F^{01} is nonzero because of spherical symmetry, Maxwell's equations reduce to

$$(e^\nu e^\lambda Y^2 F^{01})_{,1} = -4\pi Y^2 e^\nu e^\lambda J^0, \tag{6}$$

$$(e^\nu e^\lambda Y^2 F^{01})_{,0} = 4\pi Y^2 e^\nu e^\lambda J^1. \tag{7}$$

In the rest frame of the fluid, defined by $u_\mu = e^\nu \delta^0_\mu$, Eqs. (6) and (7) can be integrated to

$$F^{01} = (-e^{-\nu} e^{-\lambda}/Y^2) Q(r,t), \tag{8}$$

where

$$Q(r,t,) = 4\pi \int_0^r J^0 Y^2 e^\nu e^\lambda \, dr. \tag{9}$$

Einstein's equations become

$$\frac{-1}{Y^2} + \frac{2e^{-2\lambda}}{Y}\left(Y'' - Y'\lambda' + \frac{Y'^2}{2Y}\right) - \frac{2e^{-2\nu}}{Y}\left(\dot{Y}\dot{\lambda} + \frac{\dot{Y}^2}{2Y}\right) = -8\pi\left(\rho + \frac{Q^2}{8\pi Y^4}\right), \tag{10}$$

$$(2e^{-2\lambda}/Y)(-\dot{Y}' + Y'\dot{\lambda} + \dot{Y}\nu') = 0, \tag{11}$$

$$-\frac{1}{Y^2} + \frac{2}{Y} e^{-2\lambda}\left(Y'\nu' + \frac{Y'^2}{2Y}\right) - \frac{2e^{-2\nu}}{Y}\left(\ddot{Y} + \frac{\dot{Y}^2}{2Y} - \dot{Y}\dot{\nu}\right)$$
$$= -8\pi\left\{-p + \frac{Q^2}{8\pi Y^4} - e^{-\nu}\left[\left(-\frac{4\eta}{3} - \zeta\right)\dot{\lambda} + \left(\frac{4\eta}{3} - 2\zeta\right)\frac{\dot{Y}}{Y}\right]\right\}, \tag{12}$$

$$e^{-2\lambda}\left(\frac{Y''}{Y} + \nu'' + \nu'^2 - \frac{Y'\lambda'}{Y} - \lambda'\nu' + \frac{Y'\nu'}{Y}\right) + e^{-2\nu}\left(-\ddot{\lambda} - \dot{\lambda}^2 - \frac{\ddot{Y}}{Y} + \dot{\lambda}\dot{\nu} + \frac{\dot{Y}\dot{\nu}}{Y}\frac{\dot{Y}\dot{\lambda}}{Y}\right)$$
$$= -8\pi\left\{-p - \frac{Q^2}{8\pi Y^4} - e^{-\nu}\left[\left(\frac{2\eta}{3} - \zeta\right)\dot{\lambda} + \left(\frac{-2\eta}{3} - 2\zeta\right)\frac{\dot{Y}}{Y}\right]\right\}, \tag{13}$$

where $' = \partial/\partial r$ and $\cdot = \partial/\partial t$. The coefficients ζ and η are so far arbitrary.

II. SPHERICALLY SYMMETRIC SOLUTIONS

For the metric (5), acceleration, expansion, and shear[11] become

$$a_\mu = -\nu'\delta_\mu{}^1, \tag{14}$$

$$\theta = e^{-\nu}(\dot{\lambda} + 2\dot{Y}/Y), \tag{15}$$

$$\sigma^1_1 = \sigma^2_2 = -\sigma^3_3/2 = e^{-\nu}(\dot{Y}/Y - \dot{\lambda})/3. \tag{16}$$

Using these parameters one can determine some exact solutions of Eqs. (10)–(13).

 0022-2488/86/051367-03$02.50

(i) Charged viscous fluid with $\theta = a_\mu = 0$ and $p = \alpha\rho$. Equations (15) and (14) imply

$$(\dot{Y}/Y) = (-\dot{\lambda}/2), \quad \nu' = 0.$$

In particular the condition $\theta = 0$ eliminates all terms containing ζ from Eqs. (12) and (13). Assuming the metric depends on t only and rescaling the time so that $\nu = 0$, Eqs. (12) and (13) can be combined to yield

$$(1-3\alpha)Y\ddot{Y} + (5-3\alpha)\dot{Y}^2 + 16\pi\eta(1-3\alpha)Y\dot{Y} + 1 + \alpha = 0. \tag{17}$$

(ia) A particular solution of (17) can be easily determined when $\eta = \eta_1/t$, where η_1 is a positive constant. Then

$$Y^2 = \frac{(1+\alpha)t^2}{3\alpha - 5 + (3\alpha - 1)16\pi\eta_1}, \tag{18}$$

$$Q^2 = \frac{-(1+\alpha)(1+8\pi\eta_1)t^2}{(3\alpha - 5 - 16\pi\eta_1 + 48\pi\eta_1\alpha)^2}, \tag{19}$$

$$p = \frac{\alpha(1+8\pi\eta_1)}{2\pi(1+\alpha)}\frac{1}{t^2} = \alpha\rho, \tag{20}$$

$$e^\lambda = c_1/t^2. \tag{21}$$

Solution (18)–(21) holds for arbitrary α.

(ib) The particular case $\alpha = -1$, corresponding to a universe with tension, has also the nontrivial solution

$$\frac{Y^3}{3} = c_1 \int^t dt' \exp\left(-16\pi \int^{t'} \eta(t'')dt''\right) + c_2, \tag{22}$$

$$2Q^2 = Y^2 + 3Y^3\ddot{Y} - 3Y^2\dot{Y}^2 + 48\pi\eta Y^3\dot{Y}, \tag{23}$$

$$8\pi\rho = \frac{1}{2Y^2} - \frac{3}{2}\frac{\dot{Y}^2}{Y^2} - \frac{3}{2}\frac{\ddot{Y}}{Y} - 24\pi\eta\frac{\dot{Y}}{Y}. \tag{24}$$

One can also see from (17) that a solution of the type $Y \propto e^{\gamma t}$ with γ a positive constant requires either $\gamma = 0$ or $\eta < 0$, and from (24) that ρ tends exponentially to a negative value.

The choices $\eta = \text{const} > 0$ or $\eta = \eta_1/t$ with $\eta_1 = \text{const}$, at times made in the literature,[3] do not produce an exponential increase in Y. In the first case, in fact, one obtains

$$Y^2 = (-(c_1/16\pi\eta)e^{-16\pi\eta} + c_2)^{2/3}, \tag{25}$$

and the metric approaches flatness exponentially while Q tends to a finite value in the same limit. In the second instance

$$Y^3 = 3(c_1 \ln t + c_2). \tag{26}$$

To the extent that an exponentially increasing Y corresponds to inflation in an anisotropic universe, the above may be taken as an indication that the onset of inflation implies a phase transition with suppression of viscosity.

(ii) Charged viscous fluid with $\sigma_{\mu\nu} = 0$, $a_\mu = 0$ and $p = \alpha\rho$. For vanishing shear, one obtains from (16)

$$(\dot{Y}/Y) = \dot{\lambda}.$$

If again one assumes the metric to have only time dependence, Eqs. (12) and (13) give

$$4Y\ddot{Y} + 2(1+3\alpha)\dot{Y}^2 - 48\pi\zeta Y\dot{Y} + 1 + \alpha = 0. \tag{27}$$

(iia) If in addition $\zeta = \zeta_1/t$, where ζ_1 is a positive constant, then Eq. (27) has the solution

$$Y^2 = (1+\alpha)t^2/(48\pi\zeta_1 - 2 - 6\alpha), \tag{28}$$

which holds for arbitrary α.

Thus,

$$Q^2 = Y^2/2, \tag{29}$$

$$\rho = (12\pi\zeta_1 + 1)/4\pi(\alpha+1)t^2, \tag{30}$$

$$e^\lambda = c_3 t, \tag{31}$$

where c_3 is a constant, and $\theta = 3/t$.

(iib) The case $\alpha = -1$ yields the nontrivial solution

$$Y = \exp\left\{c_1 \int^t dt'\left(\exp\left\{-12\pi\int^{t'} \zeta(t'')dt''\right\}\right) + c_2\right\}, \tag{32}$$

with

$$Q^2 = Y^2/2, \tag{33}$$

$$8\pi\rho = (1/2Y^2) + (3\dot{Y}^2/Y^2). \tag{34}$$

Considerations entirely similar to those given under (18)–(21) apply to this case. Here $\sigma_{\mu\nu} = 0$ at all times and the roles of ζ and η are interchanged.

III. SOLUTIONS OF THE ROBERTSON–WALKER TYPE

The metric (5) acquires the Robertson–Walker form when

$$\nu = 0, \quad e^\lambda = R(t)G(r), \quad Y = R(t)G(r)r, \tag{35}$$

where $R(t)$ and $G(r)$ are arbitrary functions. This choice makes the shear vanish. One also has

$$\theta = 3(\dot{R}/R),$$

while the acceleration and vorticity vanish.

(i) If $p = 0$, the only surviving equations are

$$2R\ddot{R} + \dot{R}^2 - 24\pi\zeta R\dot{R} = \omega, \tag{36}$$

$$\frac{3}{2}\frac{G'}{rG^3} + \frac{G''}{2G^3} = \omega, \tag{37}$$

for constant ω.

Equation (36) can be solved when ζ has the form ζ_1/t with ζ_1 constant and positive. Then,

$$R = (\omega/(1 - 24\pi\zeta_1))^{1/2}t. \tag{38}$$

Two solutions have been found for Eq. (37). One is $G = 1/(Br^2 - \omega/4B)$, where B is constant, and can be reduced to the canonical Robertson–Walker form by a coordinate transformation. It requires $Q = 0$ and is discussed in Ref. 4. The second is

$$\text{(ia) } G = (1/\sqrt{-2\omega})(1/r), \tag{39}$$

$$Q = t/[2(24\pi\zeta_1 - 1)^{1/2}], \tag{40}$$

$$\rho = [(18\pi\zeta_1 - 1)/2\pi(1 - 24\pi\zeta_1)](1/t^2), \tag{41}$$

$$\theta = 3.$$

(ib) The special case $\omega = 0$ leads to the solution

$$\frac{2}{3}R^{3/2} = c_1 \int^t dt'\left\{\exp\int^{t'} 12\pi\zeta(t'')dt''\right\} + c_2, \tag{42}$$

valid for arbitrary $\zeta(t)$. If ζ is constant, one has in particular

$$R = \{(c_1/12\pi\zeta)\exp(12\pi\zeta t) + c_2\}^{2/3} \tag{43}$$

and

$$G = c_3/r^2 + c_4, \tag{44}$$

where all c's are constant. Then

$$Q = 2\sqrt{c_1 c_2}\{(c_1/12\pi\zeta)\exp(12\pi\zeta t) + c_2\}^{2/3}, \qquad (45)$$

$$\rho = \frac{-c_1^2}{6\pi} e^{24\pi\zeta t}\left\{\frac{c_1\exp(12\pi\zeta t)}{12\pi\zeta} + c_2\right\}^{-2} - \frac{c_1 c_2 r^4}{\pi(c_2 + c_2 r^2)^4}\left\{\frac{c_1\exp(12\pi\zeta t)}{12\pi\zeta} + c_2\right\}^{-4/3}, \qquad (46)$$

$$\theta = 2c_1 e^{12\pi\zeta t}/[(c_1/12\pi\zeta)\exp(12\pi\zeta t) + c_2].$$

Thus ρ and θ approach here a finite value for $t \to \infty$, while Q and R increase exponentially in the same limit.

IV. CONCLUDING REMARKS

The behavior of the universe described by (43)–(46) bears a close resemblance to those with bulk viscosity and Robertson–Walker metric discussed in Ref. (3). Though the metric (43), (44) is conformally flat, but not in general reducible to the canonical Robertson–Walker form, here, too, the expansion of $R(t)$ is increased by ζ, a fact that has no counterpart in classical cosmology, where one expects the expansion to slow down because of viscosity.

Some similarities worthy of note exist between the solutions with shear and bulk viscosity of Sec. II. The interest largely resides in the solutions with non-negligible shear viscosity, which, even in the chargeless case, are not numerous. The similarities are not only represented by (22) and (32), for which viscosity hinders the occurrence of inflation, here defined in an anisotropic context, but also by (18)–(21) and (28)–(31). In (18), viscosity tends to suppress or enhance the magnitude of Y^2 for $\eta_1 \gtrless (6 - 3\alpha)/16\pi(3\alpha - 1)$. Since $\eta_1 > 0$, one has $\frac{1}{3} < \alpha < 1$, where the upper limit is required by causality. For $\alpha = -1$ $(Y = 0)$ and for a radiation filled universe $(\alpha = \frac{1}{3})$, η does not affect the metric. Similarly for (28), suppression or enhancement of Y occur for $\zeta_1 \gtrless (1 + 2\alpha)/16\pi$, and ζ_1 remains positive for $\alpha > -\frac{1}{2}$. For both $\alpha = -1$ and $\alpha = 8\pi\zeta_1$, viscosity does not affect the metric.

Finally, the subtle interplay of viscosity and equations of state, as evidenced by the discussion of (18) and (28), is not only indicative of the variety of viscous universes one can have, but also of the direction research should take in order to reduce the freedom in the choice of parameters.

ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

[1]S. W. Hawking, Astrophys. J. **145**, 544 (1966); C. W. Misner, Astrophys. J. **151**, 431 (1968); W. Israel and J. N. Vardalas, Lett. Nuovo Cimento **4**, 887 (1970); S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), p. 469ff.

[2]M. Heller and L. Suszycki, Acta Phys. Pol. B **5**, 345 (1974); M. Heller and Z. Klimek, Astrophys. Space Sci. **33**, L37 (1975).

[3]G. Neugebauer and H. Strobel, Wiss. Z. Friedrich-Schiller-Univ. Jena Math. Naturwiss. Reihe **18**, 175 (1969); R. Treciokas and G. F. R. Ellis, Commun. Math. Phys. **23**, 1 (1971); J. D. Nightingale, Astrophys. J. **185**, 105 (1973); M. Heller, Z.Klimek, and L. Suszycky, Astrophys. Space Sci. **20**, 205 (1973).

[4]G. Papini and M. Weiss, Lett. Nuovo Cimento **44**, 612 (1985).

[5]G. Papini and M. Weiss, Phys. Lett. **89** A, 329 (1982); Lett. Nuovo Cimento **44**, 83 (1985); Ref. 4.

[6]M. Heller, Z. Klimek, and L. Suszycky, Astrophys. Space Sci. **20**, 205, (1973); M. Szydlowski and M. Heller, Acta. Phys. Pol. B **14**, 303, (1983); S. R. Roy and O. P. Tiwari, Indian J. Pure Appl. Math. **14**, 233 (1983); A. A. Coley and B. O. J. Tupper, Phys. Rev. D **29**, 2701 (1984); Astrophys. J. **288**, 418 (1985); N. O. Santos, R. S. Dias, and A. Banerjee, J. Math. Phys. **26**, 878 (1985).

[7]G. E. Tauber, Astrophys. Space Sci. **57**, 163 (1978).

[8]A. A. Coley and B. O. J. Tupper, Astrophys. J. **280**, 26 (1984); G. Prasad and B. B. Sinha, Nuovo Cimento B **54**, 401 (1979); R. Bali, Astrophys. Space Sci. **107**, 155 (1984).

[9]K. P. Singh and O. N. Singh, Proc. Natl. Acad. Sci. India A **53**, 135 (1983).

[10]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein Field Equations* (Deutscher, Berlin, 1980).

[11]G. F. R. Ellis, "Relativistic Cosmology," in *Proceedings of the E. Fermi Summer School Course* XLVII (Academic, New York, 1971).

A note on Gödel-type space-times

M. J. Rebouças[a] and J. E. Åman
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England
A. F. F. Teixeira
Centro Brasileiro de Pesquisas Físicas, 22290 Rio de Janeiro, Brazil

(Received 23 July 1985; accepted for publication 11 December 1985)

The Gödel-type space-times are examined under different conditions on the algebraic structure of the energy-momentum tensor. A generalization of the Bampi–Zordan theorem on Gödel-type models is given. Chakraborty's results are recast as a special case of the generalized Bampi–Zordan theorem. The Rebouças–Tiomno solution is shown to be the unique Gödel-type metric with an algebraic structure of the tachyon fluid type.

I. INTRODUCTION

Some years ago, Gödel[1] presented a new cosmological solution of Einstein's field equations of the form

$$ds^2 = [dt + H(x)dy]^2 - D^2(x)dy^2 - dx^2 - dz^2\,, \tag{1.1}$$

which is known as a Gödel-type metric.

The Gödel solution and its energy-momentum tensor can be given by

$$H = e^{mx}, \quad D = e^{mx}/\sqrt{2}\,, \tag{1.2}$$

and

$$T_{\mu\nu} = \rho V_\mu V_\nu, \quad V^\alpha = \delta^\alpha_0\,, \tag{1.3}$$

$$\kappa\rho = -2\Lambda = m^2 = 2\Omega^2\,, \tag{1.4}$$

where κ is the Einstein constant, Λ is the cosmological constant, ρ is the fluid density, V^α is the fluid four-velocity, and Ω is the rotation of the matter. The Gödel model is homogeneous in space and time (hereafter called ST-homogeneous). Actually it admits a five parameter group of motion (G_5) having an isotropy group of one dimension (H_1).

Besides its historical and philosophical importance, pointed out in many works,[2] the cosmological solution discussed by Gödel in 1949 has given rise to a noticeable stimulation of the investigation of rotating cosmological space-times. Moreover, since then the search for solutions to Einstein's equations of the Gödel-type has received notably more attention (see, for example, Rebouças and Tiomno[3] and references therein).

In 1965, Ozsváth[4] obtained a class of new ST-homogeneous solutions of Einstein's equations with dust by using a spinor technique. As a special result he proved an assertion, presented in Gödel's original paper without proof, which states that there exist only two ST-homogeneous solutions of Einstein's field equations with incoherent matter, namely Einstein static universe (null rotation, $\Omega = 0$) and Gödel's cosmos ($\Omega \neq 0$). His results were stated for dust with a cosmological constant, however, by a trivial change of variables ($\rho \to \rho + p$, $\Lambda \to \Lambda + \kappa p$) they can be reinterpreted as perfect fluid solutions without affecting any of Ozsváth's arguments. This is so because they are ST-homogeneous solutions and therefore ρ and p are constants.

Although one can consider the Gödel-type metric as a good starting point in the search for more general rotating models, as far as perfect fluid is concerned, however, this is not true, as was shown by Bampi and Zordan.[5] In other words, under the assumption that the energy momentum is that of a perfect fluid, viz.,

$$T_{\alpha\beta} = (\rho + p) V_\alpha V_\beta - p g_{\alpha\beta}\,, \tag{1.5}$$

by a straightforward integration of Einstein's field equations they have reduced the whole class of possible solutions to a set of three models; then, after a lengthy and rather cumbersome calculation, they have shown that the three remaining models are isometric to the Gödel universe.

One might argue that the Bampi–Zordan theorem is contained in the above well-known Gödel-Ozsváth theorem. However this is not quite true because not all Gödel-type models are ST-homogeneous.[6,7]

The null tetrad formalism developed by Newman and Penrose[8] (hereafter called NP) has been shown to be a not only useful, but also powerful tool when one has to deal with various aspects and calculations of general relativity. The concise and elegant proof of the Goldberg–Sachs theorem given by Newman and Penrose in 1962 is perhaps the best known example of the use of the spin coefficient method in general relativity. The most obvious advantages of null tetrad formalism is the fact that, since the NP tetrads are complex, the number of equations one may need to write down is greatly reduced. The NP formalism has also been very useful in the construction of exact solutions of Einstein's field equations[9] and provides a framework for the investigation of invariant properties of the gravitational field.[10] Moreover, tetrad transformations can be used to simplify the field equations, and the algebraic properties of the Weyl tensor can be easily discussed.

Using the NP null tetrad techniques we reexamine, in this paper, the Gödel-type space-times (1.1), under the assumption that the energy-momentum tensor has the algebraic Segré type structure of either a perfect fluid or tachyon fluid. As a special result we prove a generalization of the Bampi–Zordan theorem.[5] We show that a solution of Einstein's equations recently found by Hoenselaers and Vishveshwara[11] is nothing but Gödel space-time, in agreement with Chakraborty's[12] conclusion, although it is obtained in a

[a] On leave of absence from Centro Brasileiro de Pesquisas Físicas, with a fellowship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

 0022-2488/86/051370-03$02.50

context completely different from ours. We also show that all algebraic tachyon fluid type Riemannian space-times (1.1) are isometric to the Rebouças-Tiomno model.[3]

Section II is devoted to the structure of the energy-momentum tensor, together with some other important prerequisites, while in Sec. III we present our major results.

The calculations of this paper were checked by using the computer program CLASSI,[13] written in the symbolic manipulation language SHEEP.[14]

II. STRUCTURE OF THE ENERGY-MOMENTUM TENSOR AND PREREQUISITES

The algebraic classification of the Weyl conformal tensor has played a significant role in the study of various topics of general relativity. In particular, the Petrov classification, as it is known, has played a remarkable part in the investigation and classification of solutions of Einstein's field equations (see, for example, Kramer *et al.*,[9] Karlhede[10] and MacCallum[15]). It splits up gravitational fields into six different types, namely the *algebraically general*, type I and the *algebraically special*, types II, III, D, N, and 0. One can think of it as a classification related to the purely gravitational properties of space-time, whereas the matter content is represented by the energy-momentum tensor $T_{\mu\nu}$, which by virtue of Einstein's equations

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}\ , \tag{2.1}$$

clearly has the same algebraic structure of the Ricci tensor $R_{\mu\nu}$. Therefore, for the full characterization of nonvacuum space-times, besides the Weyl part of the Riemann tensor, one also has to consider the Ricci part.

Our major aim in this section is to present a brief summary of canonical types of the symmetric energy-momentum tensor $T_{\mu\nu}$, to define the notation and make our text as clear and self-contained as possible. A detailed and useful review of the classification of second-order symmetric tensors in general relativity can be found in Hall.[16]

Let M be a four-dimensional space-time endowed with a Lorentz metric of signature -2. Let $T_P(M)$ denote the tangent space to M at P. The algebraic classification of a symmetric second-order tensor $T_{\mu\nu}$ (say) at a point P consists of finding real vectors $v^\alpha \in T_P(M)$ and real numbers λ such that

$$(T_{\alpha\beta} - \lambda g_{\alpha\beta})v^\beta = 0\ . \tag{2.2}$$

In a space with positive definite metric, a real symmetric matrix can always be diagonalized. However, since the metric tensor has Lorentz signature the eigenvalue problem (2.2) is not the standard one well known in linear algebra. The Lorentzian character of our space-time together with the symmetry of $T_{\alpha\beta}$ lead to a more complicated algebraic structure, excluding certain Segré types. On the other hand, if $T_{\alpha\beta}$ is meant to represent the energy-momentum tensor in general relativity, further restriction is provided by the *dominant energy condition*,[17] which states that the local energy density as measured by any observer with four-velocity u^α is non-negative ($T_{\alpha\beta}u^\alpha u^\beta \geqslant 0$), and that the velocity of energy flow does not exceed the velocity of light ($q^\alpha = T^\alpha{}_\beta u^\beta$ is nonspacelike). The energy conditions are indeed very restrictive and rule out an additional set of Segré types.[18]

Taking into account the above restrictions on the energy-momentum tensor $T_{\alpha\beta}$, it turns out that the possible Segré types are given by one of the following classes.

Class (*i*): [1,111] and its specializations: [1,1(11)], [(1,1)11], [(1,1)(11)], [1,(111)], [(1,11)1], and [(1,111)].

Class (*ii*): [2,11] and its specializations: [2,(11)], [(2,1)1], and [(2,11)].

Where the individual digits refer to the multiplicity of the corresponding eigenvalues, equal eigenvalues are enclosed in a round bracket, and in each case the first digit refers to the timelike eigenvalue, which is separated from the spacelike ones with a comma.

In a recent paper, McIntosh *et al.*[19] have a given canonical set of nonzero NP quantities Φ_{AB} for each Segré type. Corresponding to the above Segré types, the nonvanishing components of the Ricci spinor Φ_{AB} can be, respectively, put into the form of one of the following cases.

Class (*i*): $\{\Phi_{00} = \Phi_{22}, \Phi_{11}, \Phi_{02} = \Phi_{20}\}$ and its specializations: $\{\Phi_{00} = \Phi_{22}, \Phi_{11}\}$, $\{\Phi_{11}, \Phi_{02} = \Phi_{20}\}$, $\{\Phi_{11}\}$, $\{\Phi_{00} = \Phi_{22} = 2\Phi_{11}\}$, $\{-2\Phi_{11} = \Phi_{00} = \Phi_{22}\}$, and $\Phi_{AB} = 0$.

Class (*ii*): $\{\Phi_{11}, \Phi_{22}, \Phi_{02} = \Phi_{20}\}$ and its specializations: $\{\Phi_{11}, \Phi_{22}\}$, $\{2\Phi_{11} = \Phi_{02}, \Phi_{22}\}$, and $\{\Phi_{22}\}$.

Finally, we mention that a survey of Segré types corresponding to energy-momentum tensor of several fields used in general relativity can be found by Kramer *et al.*[9] and in Hall.[20]

III. MAIN RESULTS AND DISCUSSIONS

Consider a four-dimensional Riemannian manifold M, endowed with a Gödel-type metric (1.1). At each point of M one can define a set of pseudo-orthonormal complex null tetrads Θ^A $(A = 0,1,2,3)$, such that the line element (1.1) can be put into the form

$$ds^2 = \eta_{AB}\Theta^A\Theta^B\ , \tag{3.1}$$

where

$$\eta_{AB} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{3.2}$$

The basis vectors Θ^A are defined up to a local Lorentz transformation. However, the set of Θ^A that turns out to be adequate for our discussions is

$$\Theta^0 = \frac{1}{\sqrt{2}}(dt + H\,dy + dz), \quad \Theta^1 = \frac{1}{\sqrt{2}}(dt + H\,dy - dz)\ ,$$
$$\Theta^2 = \frac{1}{\sqrt{2}}(D\,dy - i\,dx), \quad \Theta^3 = \frac{1}{\sqrt{2}}(D\,dy + i\,dx)\ . \tag{3.3}$$

A straightforward calculation (checked by using CLASSI[13]) gives the value of NP Ricci spinor Φ_{AB} in the frame (3.3). We find the following nonvanishing components:

$$\Phi_{00} = \Phi_{22} = \tfrac{1}{8}(H'/D)^2,$$
$$\Phi_{01} = \Phi_{12} = \tfrac{1}{8}(H'/D)', \qquad (3.4)$$
$$\Phi_{11} = \tfrac{1}{4}[\tfrac{3}{4}(H'/D)^2 - (D''/D)],$$

where a prime indicates a derivative with respect to x.

Now, since $\Phi_{02} = 0$, the necessary and sufficient conditions for Φ_{AB} to be of *algebraic perfect fluid type* reduce to[19,20]

$$\Phi_{01} = \Phi_{12} = 0, \quad \Phi_{00} = \Phi_{22} = 2\Phi_{11}, \qquad (3.5)$$

which, according to (3.4), imply

$$H'/D = \text{const} \equiv 2\Omega, \qquad (3.6)$$

and

$$D''/D = \text{const} = m^2, \qquad (3.7)$$

with

$$m^2 = 2\Omega^2. \qquad (3.8)$$

Equations (3.6)–(3.8) define the Gödel space-time. Thus, we have proved the following generalization of Bampi–Zordan theorem.

Theorem 1: All algebraic perfect fluid type (Segré type [1,(111)]) Riemannian manifolds endowed with a Gödel-type metric are isometric to the Gödel space-time.

It seems worth emphasizing that stress-energy tensors of some quite different matter distributions may, in fact, have precisely the same Segré type. In particular, it has been shown by Tupper[21,22] that the energy-momentum tensor of a magnetohydrodynamic fluid, with or without viscous terms, may have the Segré type [1,(111)] algebraic structure.

By a trivial coordinate transformation, the Hoenselaers–Vishveshwara solution[11]

$$\begin{aligned} ds^2 = {} & [1 + \tfrac{1}{2}\sinh^2(K^{1/2}x)]dt^2 \\ & - A\,[\cosh(K^{1/2}x) - 1]^2 dy\,dt \\ & - (A^2/2)[4\cosh(K^{1/2}x) \\ & - \cosh^2(K^{1/2}x) - 3]dy^2 - dz^2 \end{aligned} \qquad (3.9)$$

can be brought to the form (1.1), with

$$H = A\,[1 - \cosh(K^{1/2}x)] \qquad (3.10)$$

and

$$D = (A/\sqrt{2})\sinh(K^{1/2}x), \qquad (3.11)$$

where $A,K = \text{const}$. Now since it is a Segré type [1,(111)] solution, by Theorem 1, it is isometric to the Gödel model, in agreement with Chakraborty's conclusion.[12]

Again since $\Phi_{02} = 0$, the necessary and sufficient conditions for the Ricci spinor Φ_{AB} to be of *algebraic tachyon fluid type* reduce to

$$\Phi_{00} = \Phi_{22} = -2\Phi_{11}, \quad \Phi_{01} = \Phi_{12} = 0, \qquad (3.12)$$

which again leads to Eqs. (3.6) and (3.7). But now, instead of (3.8) we obtain

$$m^2 = 4\Omega^2. \qquad (3.13)$$

Equations (3.6), (3.7), and (3.13) define the Rebouças–Tiomno[3] metric. Thus, the following theorem holds.

Theorem 2: All algebraic tachyon fluid type (Segré type [(1,11)1]) Riemannian manifolds endowed with a Gödel-type metric are isometric to Rebouças–Tiomno space-time.

To conclude, we should like to mention that a systematic study of Riemannian Gödel-type space-times employing the "equivalence problem" theory[10,15,23] has been carried out and we hope to publish our results shortly elsewhere.

ACKNOWLEDGMENTS

J. E. Å. and M. J. R. would like to thank Dr. M. A. H. MacCallum for his kind hospitality at the School of Mathematical Sciences, Queen Mary College, University of London.

J. E. Å. acknowledges SERC for financial support, and M. J. R. is grateful to CAPES for the grant of a fellowship.

[1]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[2]See, for example, J. Pfarr, Gen. Relativ. Gravit. **13**, 1073 (1981), and references therein quoted on the Gödel model.
[3]M. J. Rebouças and J. Tiomno, Phys. Rev. D **28**, 1251 (1983).
[4]I. Ozsváth, J. Math. Phys. **6**, 590 (1965).
[5]F. Bampi and C. Zordan, Gen. Relativ. Gravit. **9**, 393 (1978).
[6]M. J. Rebouças and J. Tiomno, "A class of inhomogeneous Gödel-type models," Nuovo Cimento, in press.
[7]A. Banerjee and S. Banerji, J. Phys. A **1**, 188 (1968).
[8]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[9]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[10]A. Karlhede, Gen. Relativ. Gravit. **12**, 693 (1980).
[11]C. Hoenselaers and C. V. Vishveshwara, Gen. Relativ. Gravit. **10**, 43 (1979).
[12]S. K. Chakraborty, Gen. Relativ. Gravit. **12**, 925 (1980).
[13]J. E. Åman, "A Manual for CLASSI. Classification Programs for Geometries in General Relativity," Institute of Theoretical Physics, University of Stocklholm report, 1982.
[14]I. Frick, "SHEEP Users Guide," USIP Report 77-15, University of Stockholm, 1979.
[15]M. A. H. MacCallum, "Classifying Metrics in Theory and Practice," in *Unified Field Theories in More than 4 Dimensions, Including Exact Solutions*, edited by V. de Sabbata and E. Schmutzer (World Scientific, Singapore, 1983).
[16]G. S. Hall, Diff. Geom. **12**, 53 (1983); *Lectures at the Stefan Banach Centre* (PWN–Polish Scientific, Warsaw, 1979).
[17]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[18]The types that violate the dominant energy condition happen to violate even the weak energy condition (see, for example, Ref. 9).
[19]C. B. G. McIntosh, J. M. Foyster, and A. W.-C. Lun, J. Math. Phys. **22**, 2620 (1981).
[20]G. S. Hall, Arab. J. Sci. Eng. **9**, 87 (1984).
[21]B. O. J. Tupper, Gen. Relativ. Gravit. **15**, 47 (1983).
[22]B. O. J. Tupper, Gen. Relativ. Gravit. **15**, 849 (1983).
[23]É. Cartan, *Leçons sur la géométrie des espaces de Riemann* (Gauthiers–Villars, Paris, 1928).

Oscillating and absolute Maxwellians: Exact solutions for ($d > 1$)-dimensional Boltzmann equations

Henri Cornille
Service de Physique Théorique, CEN SACLAY, 91191 Gif-sur-Yvette Cedex, France

(Received 30 September 1985; accepted for publication 8 January 1985)

A study of ($d > 1$)-dimensional nonlinear Boltzmann equations is made for which both momentum and energy conservations hold. Maxwell particles in the presence of outside forces are assumed. For either linear spatial force or linear velocity force plus source term, a class of exact solutions for the homogeneous and inhomogeneous distributions is determined. In particular, a look is taken at oscillating external forces with a varying parameter and distributions relaxing towards oscillating Maxwellians are found. On the other hand, exact inhomogeneous distributions relaxing towards absolute Maxwellians are obtained. The exact solutions have asymptotic regimes that are solutions of the linear part of the Boltzmann equations and the relaxations towards these regimes are studied. Furthermore when equilibrium absolute Maxwellian states exist, whether the relaxation towards these states are from above or from below is investigated and the possibility of finding an overpopulation of high velocity particles at intermediate times is looked at.

I. INTRODUCTION

This work and the one presented in a companion paper[1] try to answer different questions.

(i) It is of current interest to study the asymptotic behavior of nonlinear physical systems when external forces with varying parameters are present. What about the nonlinear Boltzmann equation (BE)?

(ii) Exact solutions, when the particles are of Maxwell type with present external spatially dependent forces, exist and have been known for a long time.[2,3] However, they were obtained with a vanishing collision term, which means that the distributions are solutions of the linear part of the BE. For instance, Boltzmann[2] for linear spatial $\mathbf{x}$ dependent forces, had found exact distributions with an asymptotic oscillatory behavior. Can we still hope to find exact solutions when the collision term does not vanish? What happens if the outside forces, instead of being spatially $\mathbf{x}$ dependent, are velocity $\mathbf{v}$ dependent?

(iii) Can we enlarge the class of known exact solutions when we add external forces? I recall that two classes of Boltzmann models, with Maxwell particles, have been considered. First, those for which both energy and momentum conservations hold. The so-called BKW exact solution is an homogeneous distribution found by Bobylev[4] and Krook and Wu[5] for the ($d = 3$)-dimensional BE, which has been generalized[6,7] to other d-dimensional models. *Unfortunately, the inhomogeneous BKW solution*[4–8] *obtained by* Bobylev,[4] *with the help of the Nikolskii*[9,10] *transform, goes to zero when $t \to \infty$ so that up to now no exact inhomogeneous solution relaxing towards to an absolute Maxwellian is known.* In the second class of models, the momentum conservation has been dropped. The oldest is the $d = 1$ Kac[11] model while[6] more recently the $d = 2$ Tjon and Wu[12] model and others have been presented. Besides the BKW distribution,[6] other exact homogeneous[13] and inhomogeneous[14] solutions of the Kac model have been found. In particular, inhomogeneous distributions relaxing towards absolute Maxwellians exist for the $d = 1$ Kac model. *Consequently we would like to find in the first class of models* (the ones studied here) *examples of exact inhomogeneous solutions with absolute Maxwellians* [$\exp(-\text{const}|\mathbf{c}|^2)$, $\mathbf{c}$ peculiar velocity].

Here we study the ($d > 1$)-dimensional BE with momentum and energy conservations holding, while in the companion[1] paper we restrict to the $d = 1$ Kac model. A brief summary of the present results, without proofs, has been presented.[1–15] We start with a d-dimensional BE ($d = 3$ is the standard BE) and assume Maxwellian particles:

$$Lf = \mathrm{Col}(f), \quad L = \partial_t + \mathbf{v}\cdot\partial_{\mathbf{x}} + \mathbf{A}_0(t)\cdot\partial_{\mathbf{v}} + \Lambda(\mathbf{v},\mathbf{x},t),$$
$$\Lambda = A(\mathbf{x},t)\cdot\partial_{\mathbf{v}}$$

or

$$\Lambda = a_1(t)\partial_{\mathbf{v}}\cdot\mathbf{v} + a_2(t) = a_1\mathbf{v}\cdot\partial_{\mathbf{v}} + da_1 + a_2,$$

$$\mathrm{Col}(f) = S_d^{-1}\int d\Omega_d\,\sigma^{(d)}(x)d\mathbf{w} \times [f(\mathbf{v}')f(\mathbf{w}') - f(\mathbf{v})f(\mathbf{w})], \tag{1.1}$$

$$\begin{pmatrix} v'^2 \\ w'^2 \end{pmatrix} = \begin{pmatrix} v^2 \\ w^2 \end{pmatrix} + \left((w^2 - v^2)\sin^2\frac{\chi}{2} + |\mathbf{v}|\,|\mathbf{w}|\sin\chi\sin\theta\cos\epsilon\right)\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

$$S_d = \int d\Omega_d;$$

$$d\Omega_d = (\sin\chi)^{d-2}(\sin\epsilon)^{d-3}(\sin\epsilon_1)^{d-4} \cdots \sin\epsilon_{d-4}\,d\chi\,d\epsilon\,d\epsilon_1\cdots,$$

$d\epsilon_{d-3}$ is the d-dimensional solid angle expressed as a function of the $d-2$ polar angle $\chi,\epsilon,\epsilon_1,\ldots,\epsilon_{d-3}\in[0,\pi]$ and $\epsilon_{d-2}\in[0,2\pi]$ azimuthal angle, of the two velocities $\mathbf{v}'-\mathbf{w}'$, $\mathbf{v}-\mathbf{w}\cdot\sigma^{(d)}(\chi)$ is the cross section and θ the angle between $\mathbf{v}$ and $\mathbf{w}$.

It turns out that our assumptions (Secs. II–III) lead finally to the following results: Let us define $\mathbf{c}$, the relative

 0022-2488/86/051373-14$02.50

velocity and $\langle \mathbf{v}\rangle$ the mean velocity with $\mathbf{c}=\mathbf{v}-\langle \mathbf{v}\rangle$. The distributions constructed f depend in fact on two variables (similarity solutions) $\boldsymbol{\eta}=\gamma(t)\mathbf{c}$ and t. The spatial $\mathbf{x}$ dependence is entirely contained in the mean velocity. Here f has the analytic structure of the BKW solutions, and f is even in $\boldsymbol{\eta}$ and $\mathscr{L}\boldsymbol{\eta}^2=0$, where $\mathscr{L}$ is the differential part of L. We adopt the following strategy.

In a first part, in order to have an idea of the possible asymptotic behavior we assume $\mathrm{Col}(f)\simeq 0$ for t large and the linear part of the BE defines asymptotic solutions f_{as}. We find

$$f_{\mathrm{as}}\simeq\nu(t)\exp(-\tfrac{1}{2}\boldsymbol{\eta}^2),\quad \boldsymbol{\eta}=\gamma(t)\mathbf{c}, \tag{1.2}$$

a product of a pure time factor (which can be a constant) by a Gaussian. We are interested in two possibilities: either $\gamma(t)$ is oscillating and $f_{\mathrm{as}}(\mathbf{c},t)$ can lead to oscillating Maxwellians or $\gamma(t)\to$const and f_{as} can correspond to absolute Maxwellians $\exp(-\text{const }\mathbf{c}^2)$ multiplied by $\nu(t)$. In this last case we seek the condition on the forces such that $\nu(t)$ is a constant. For the oscillating possibility, we choose the time dependence of the force in such a way that the $\gamma(t)$ couple two circular functions of periods 2π and $2\pi/q$ (q integer) and introduce a verying parameter. We find that the $(\mathbf{c},t)$ Gaussian in (1.2) can oscillate between two, three, four,... extremal Maxwellians. We do not observe the doubling of periods but instead the appearance of harmonics, the reason being that these f_{as}, obtained with vanishing collision terms, are solutions of linear differential systems (linear part of the BE).

In a second part we build up the exact solutions having asymptotic behaviors provided by (1.2). Starting with our original assumptions, Eq. (2.2) of Sec. II, different methods exist. The only possible ansatz compatible with the original assumptions being the product of a Gaussian in the $|\mathbf{c}|$ variable with time-dependent width by a first-order $|\mathbf{c}|^2$ polynomial, we can directly substitute it into (1.1) and solve the relations coming from the coefficients of like power $|\mathbf{c}|^{2p}$ in both sides. We can also check some of these relations by the macroscopic continuity equations. Finally we can use a generalization of the Nikolskii[9,10] method where at some stage we replace the inhomogeneous BE by an equivalent homogeneous one. All these methods are used here.

An interesting by-product is the possibility of obtaining ($d>1$)-dimensional inhomogeneous distributions relaxing towards absolute Maxwellians:

$$f_{\mathrm{abs\,Max}}=(2\pi K)^{-d/2}\rho_0(0)\exp[-(\gamma(\infty)\mathbf{c})^2/2], \tag{1.3}$$

K being a constant and ρ_0 the local density. We find two possible forces leading to such equilibrium states: either $\Lambda=\mathbf{A}(\mathbf{x},t)\cdot\partial_{\mathbf{v}}$ if $\mathbf{A}$ is linearly $\mathbf{x}$ dependent or $a_1(t)\mathbf{v}\cdot\partial_{\mathbf{v}}$, a particular mixing of velocity-dependent force plus sources. Although recently[14] such equilibrium distributions have been obtained for the $d=1$ Kac model, *to my knowledge*, for $d>1$ and models not violating momentum conservation, *these are the first explicit examples of inhomogeneous distributions having that property.*

In Sec. II we study spatially dependent forces, while in Sec. II velocity force and uniform source are investigated.

In Sec. IV we study the relaxation towards the asymptotic regimes. On the one hand, for the exact solutions f, written with the variables $\boldsymbol{\eta},t$ we define reduced distributions $\bar{F}=f(\boldsymbol{\eta},t)/f_{\mathrm{as}}(\boldsymbol{\eta},t)$ and find that they approach their limit 1^- from below. This is the same result as for homogeneous BKW distributions without external force. The pertinence of this study is justified only if f_{as} is itself really an asymptotic regime as it is for oscillating Maxwellians. However, for distributions relaxing towards absolute Maxwellians, $f_{\mathrm{as}}(\mathbf{c},t)\to f_{\mathrm{abs\,Max}}$, it is worthwhile to define another reduced distribution $F=f(\mathbf{c},t)/f_{\mathrm{as}}(\mathbf{c})$ and study the approach towards 1. *The new fact is that we find either* $F\to 1^-$ (from below) or $F\to 1^+$ (from above). We find examples of distribution having at intermediate times, *a population of high relative velocities* $|\mathbf{c}|$ *larger than the ones present at initial time or at equilibrium* (Tjon effect[16]).

II. SPATIALLY DEPENDENT EXTERNAL FORCES WITH INHOMOGENEOUS DISTRIBUTIONS

We assume Maxwell particles and study different classes of asymptotic f_{as} and exact solutions of the BE

$$Lf=\mathrm{Col}(f),\quad L=\partial_t+\mathbf{v}\cdot\partial_{\mathbf{x}}+(\mathbf{A}_0(t)+\mathbf{A}(\mathbf{x},t))\cdot\partial_{\mathbf{v}}. \tag{2.1}$$

A. The different assumptions and their consequences

(i) $f=\tilde{f}(\boldsymbol{\eta}^2,\mathbf{x},t),\quad \boldsymbol{\eta}=\gamma(t,\mathbf{x})(\mathbf{v}-\mathbf{v}_0(t,\mathbf{x}))$, (2.2a)

(ii) $\mathrm{Col}(f_{\mathrm{as}})\simeq 0$

or

$$f\simeq f_{\mathrm{as}}=\nu(\mathbf{x},t)\exp(-\boldsymbol{\eta}^2/2),\quad \text{for } t \text{ large.} \tag{2.2b}$$

From (i) follows that the local density is

$$\rho_0=\int f\,d\mathbf{v}=\gamma^{-d}\int\tilde{f}\,d\boldsymbol{\eta}$$

and for the mean velocity

$$\langle\mathbf{v}\rangle\rho_0=\int\mathbf{v}f\,d\mathbf{v}=\mathbf{v}_0\gamma^{-d}\int\tilde{f}\,d\boldsymbol{\eta}$$

or $\langle\mathbf{v}\rangle=\mathbf{v}_0$. It follows that $\boldsymbol{\eta}=\gamma\mathbf{c}$, $\mathbf{c}$ being the peculiar velocity. From (ii) we study $Lf_{\mathrm{as}}=0$ or $(\partial_t+\mathbf{v}\cdot\partial_{\mathbf{x}})\log\nu=L\boldsymbol{\eta}^2/2$. On the right-hand side (rhs) we have powers $\mathbf{v}^2v_i, v_i^2, v_iv_j$ not present on the lhs. Equating to zero the coefficients of these powers we find both $\partial_{x_i}\gamma=0$ or $\gamma=\gamma(t)$ and $\gamma^{-1}L\boldsymbol{\eta}^2/2=\Sigma\eta_i\zeta_i(\mathbf{x},t)$ [see (A3) in Appendix A]. At this stage, $L\boldsymbol{\eta}^2$ is linear in η_i, we could go on and find the whole class[2,3] of Boltzmann's solutions. On the contrary we stop here. Further restrictions, coming from another assumption, will appear later on, leading to $L\boldsymbol{\eta}^2=0$ and only a subclass of possible f_{as}.

We introduce a third assumption

$$\text{(iii)}\quad (2\pi\Delta(\mathbf{x},t))^{d/2}f=\exp\left(\frac{-\boldsymbol{\eta}^2}{2\Delta}\right)\sum_{p=0}^{n}\alpha_{2p}(\mathbf{x},t)\left(\frac{\boldsymbol{\eta}^2}{2}\right)^p, \tag{2.2c}$$

and show that the only possible ansatz solution (iii) is provided with $n=1,\Delta,\alpha_0,\alpha_2$ only t dependent, and $L\boldsymbol{\eta}^2=0$. Let us rewrite (2.1) for $\tilde{f}$:

TABLE I. Compatible forces and $\boldsymbol{\eta} = \gamma(t)(\mathbf{v} - \langle \mathbf{v} \rangle)$.

$$\mathscr{L} = \partial_t + \mathbf{v} \cdot \partial_{\mathbf{x}} + \mathbf{A}_0(t) \cdot \partial_{\mathbf{v}} + \Lambda, \quad \mathbf{A} = \sum_1^d \hat{\imath}_j A_l, \quad \mathbf{A}_0 = \sum \hat{\imath}_j A_{0j}, \quad \langle \mathbf{v} \rangle = \sum \hat{\imath}_j \langle \mathbf{v} \rangle_j,$$

$$\boldsymbol{\alpha} = \sum \hat{\imath}_j \alpha_j, \quad \left(\frac{\gamma_t}{\gamma} + \partial_t\right)\alpha_i - \sum_j \alpha_j \omega_{ij} = A_{0i}, \quad \omega_{ij} + \omega_{ji} = 0.$$

(a) $\Lambda = \mathbf{A}(\mathbf{x},t) \cdot \partial_{\mathbf{v}}, \quad \mathscr{L}\boldsymbol{\eta}^2 = 0 \quad \langle \mathbf{v} \rangle_i = \gamma^{-1}\gamma_t x_i + \sum \omega_{ij} x_j + \alpha_i,$

$$A_i = x_i\left(a_1(t) - \sum_j \omega_{ij}^2\right) + \sum_{j \neq i} x_j \left(\frac{d\omega_{ij}}{dt} + \sum_l \omega_{il}\omega_{lj} + 2\frac{\gamma_t}{\gamma}\omega_{ij}\right), \quad a_1(t) = \gamma_{tt}\gamma^{-1}.$$

(i) $\omega_{ij} = 0, \quad \mathbf{A} = a_1(t)\mathbf{X}, \quad \langle \mathbf{v} \rangle \gamma = \gamma_t \mathbf{X} + \int_0^t \mathbf{A}_0(t')\gamma(t')dt' + \boldsymbol{\alpha}(0), \quad a_1(t)\gamma = \gamma_{tt},$

if $\mathbf{A} \equiv 0$, $\gamma = \text{const}\ t \to \infty$ Bobylev–Muncaster sol.

(ii) $\gamma^2\omega_{ij} = \text{const}, \quad d = 3, \quad \langle \mathbf{v} \rangle = \boldsymbol{\alpha} + \gamma_t \gamma^{-1}\mathbf{x} + \boldsymbol{\omega} \wedge \mathbf{x}, \quad \boldsymbol{\omega} = (\omega_{32}, \omega_{12}, \omega_{21}),$
conservative force.

(iii) $\gamma^2\omega_{ij}(t), \quad \partial_{x_j}A_i - \partial_{x_i}A_j = \frac{2}{\gamma^2}\frac{d}{dt}\gamma^2\omega_{ij} \neq 0$ nonconservative force.

(a′) $\mathbf{A}_0 \equiv 0, \quad \Lambda_0 \equiv 0, \quad \omega_{ij}(t) \neq 0, \quad \omega_{ij} = \omega_{ij}(0)\gamma^{-2}, \quad \sum_j \omega_{ij}^2 = \omega^2(t),$

$\gamma = \sqrt{\omega^2(0)t^2 + 1}, \quad \gamma(0) = 1, \quad \langle \mathbf{v} \rangle_i = \gamma^{-1}\gamma_t x_i + \sum \omega_{ij} x_j + \alpha_i,$

$d = 2; \quad d = 3,$ no solution; $d = 4, \quad \omega_{24} = -\eta\omega_{13}, \quad \omega_{23} = \eta\omega_{14}, \quad \omega_{34} = \eta\omega_{12}, \quad \eta^2 = 1.$

(b) $\Lambda = a_1(t)\partial_{\mathbf{v}} \cdot \mathbf{v} + a_2(t), \quad (\mathscr{L} - da_1 - a_2)\boldsymbol{\eta}^2 = 0, \quad \langle \mathbf{v} \rangle_i = \left(\frac{\gamma_t}{\gamma} + a_1\right)x_i + \sum_l \omega_{il}x_l + \alpha_i.$

(i) $\omega_{ij} = 0, \quad \gamma = \exp\left(-\int_0^t a_1\, dt'\right)\left(1 + \mu_0 \int_0^t \exp\left(\int_0^{t'} a_1\, dt''\right)dt'\right),$

$a_i = \mu_0\gamma^{-1} + \partial_t \log\gamma^{-1}, \quad \langle \mathbf{v} \rangle = \boldsymbol{\alpha} + \mu_0\mathbf{X}\gamma^{-1}, \quad \gamma\boldsymbol{\alpha} = \boldsymbol{\alpha}(0) + \int_0^t \mathbf{A}_0(t')\gamma(t')dt'.$

(ii) $\omega_{ij} \neq 0, \quad \sum_{l,j \neq i} \omega_{il}\omega_{lj} = 0, \quad \omega_{ij} = \exp\left(\int_0^t a_1\, dt'\right)\omega_{ij}(0), \quad \left(\partial_t + \frac{\gamma_t}{\gamma}\right)a_1 + \frac{\gamma_{tt}}{\gamma} + \sum_\gamma \omega_{ij}^2 = 0,$

$d = 2; d = 3$ no solution; $d = 4 \quad \omega_{24} = -\eta\omega_{13}, \quad \omega_{23} = \eta\omega_{14}, \quad \omega_{34} = \eta\omega_{12}, \quad \eta^2 = 1, \quad \omega^2 = \sum_j \omega_{ij}^2.$

$$L\tilde{f} = \gamma^{-d}\,\text{Col}\,\tilde{f},$$

$$\text{Col}\,\tilde{f} = S_d^{-1}\int d\Omega_d\, \sigma^{(d)}\, d\boldsymbol{\eta}_w \times (f(\boldsymbol{\eta}_{v'})f(\boldsymbol{\eta}_{w'}) - f(\boldsymbol{\eta}_v)f(\boldsymbol{\eta}_w)), \tag{2.3}$$

with $\boldsymbol{\eta} = \boldsymbol{\eta}_v = \gamma\mathbf{c}$, and $\boldsymbol{\eta}_v, \boldsymbol{\eta}_{v'}, \boldsymbol{\eta}_w, \boldsymbol{\eta}_{w'}$, having the same collision law velocity as those for $\mathbf{v}, \mathbf{v}', \mathbf{w}, \mathbf{w}'$ written down in (1.1). The ansatz (iii) substituted into Col $\tilde{f}$ gives terms like $\boldsymbol{\eta}_{v'}^{2p}\boldsymbol{\eta}_{w'}^{2q} - \boldsymbol{\eta}_v^{2p}\boldsymbol{\eta}_w^{2q}$, which, after integration over $d\Omega_d$, lead to only powers $\boldsymbol{\eta}_v^{2m}$. Consequently the lhs of (2.3) must contain only even powers $\boldsymbol{\eta}_v$ terms. At this lhs $L\tilde{f}$, the highest power is $\boldsymbol{\eta}^{2n+2}\boldsymbol{\eta} \cdot \partial_{\mathbf{x}}\Delta$ and so $\Delta = \Delta(t)$. At this stage, the rhs and lhs require the highest powers $\boldsymbol{\eta}_v^{4n} = \boldsymbol{\eta}_v^{2n+2}$ or $n = 1$. In (iii) the possibility for α_0, α_2 to be $\mathbf{x}$ dependent remains. However the equation for $\boldsymbol{\eta}_v^4$ gives $\alpha_2 = \gamma^d(\sigma_2^{(d)}\Delta)^{-1}\partial_t \log\Delta$ [$\sigma_2^{(d)}$ being the moment of $\sin^2\chi\sigma^{(d)}(\chi)$] or $\alpha_2 = \alpha_2(t)$. Due to the fact that $L\boldsymbol{\eta}^2$ is linear in η_i (see above), the lhs contains odd η_i powers whose coefficients must vanish. For instance, $\boldsymbol{\eta}^2\eta_i$ has two terms: one, proportional to $\partial_{x_i}\alpha_2$, is identically zero; the other, coming from $\boldsymbol{\eta}^2 L\boldsymbol{\eta}^2$, shows that $L\boldsymbol{\eta}^2 = 0$ or $\zeta_i = 0$ [ζ_i defined above and in (A3)]. Finally the power η_i has also two terms: one, proportional to $L\boldsymbol{\eta}^2$, is zero; and the other, proportional to $\partial_{x_i}\alpha_0$, gives the last condition $\alpha_0(t)$. We have shown that the assumptions (2.2) lead to the only ansatz

$$(2\pi\Delta(t))^{-d/2}f(\boldsymbol{\eta}^2,t) = (\exp(-\boldsymbol{\eta}^2/2\Delta))(\alpha_0(t) + \alpha_2(t)\boldsymbol{\eta}^2/2),$$

$$\boldsymbol{\eta} = \gamma\mathbf{c}, \tag{2.4}$$

and in the supplement $L\boldsymbol{\eta}^2 = 0$. This last result is very important, coming back to $Lf_{as} = 0$ it gives (see Appendix A) $\nu(\mathbf{x},t) = \text{const}$ and

$$f_{as} \simeq \text{const}\exp(-\boldsymbol{\eta}^2/2), \quad \boldsymbol{\eta} = \gamma(t)\mathbf{c}, \quad \mathbf{c} = \mathbf{v} - \langle \mathbf{v} \rangle. \tag{2.2'}$$

B. Determination of $\gamma, \langle \mathbf{v} \rangle, \mathbf{A}, f_{as}$ from $L\boldsymbol{\eta}^2 = 0$

The study is done in Appendix A. Introducing Cartesian coordinates $(\hat{\imath}_j)$, $j = 1,...,d$, and components $\langle \mathbf{v} \rangle_j, A_{0j}, A_j, x_j$, we find

$$\langle \mathbf{v} \rangle = \boldsymbol{\alpha}(t) + \gamma_t \gamma^{-1} \mathbf{x} + \sum_j \hat{i}_j \sum_{l \neq j} \omega_{jl} x_l, \qquad (2.5)$$

$$\mathbf{A} + \mathbf{A}_0 = (\gamma_t \gamma^{-1} + \partial_t) \langle \mathbf{v} \rangle - \sum_j \hat{i}_j \sum_l \langle \mathbf{v} \rangle_j \omega_{lj}, \qquad (2.6)$$

$\omega_{ij}(t)$ being an antisymmetric tensor $\omega_{ij} + \omega_{ji} = 0$, and (2.6) expressing $L\eta^2 = 0 = \Sigma \eta_i \zeta_i$.

Substituting (2.5) into (2.6) we find that $\mathbf{A}(\mathbf{x},t)$, linear in the x_j spatial variables, corresponds to a conservative force, but can contain a nonconservative part if $(d/dt)\gamma^2\omega_{ij}(t) \neq 0$. Some simple classes of solutions $\mathbf{A}, \langle \mathbf{v} \rangle, \gamma$ established in Appendix A are quoted in Table I(a). In the most simple (i) $\omega_{ij} = 0$ case, if we further assume $\mathbf{A} = 0$ (force without spatial dependence), then $\gamma = \text{const } t \to \infty$ when $t \to \infty$, $f_{as} \to 0$ and we recover the asymptotic inhomogeneous Bobylev distribution[4] for $d = 3$. In the following we discard this uninteresting solution. For the spatially dependent other classes (i) $\omega_{ij} = 0$, $\mathbf{A} \neq 0$, (ii) $\gamma^2 \omega_{ij} = \text{const}$, and (iii) $(d/dt)\gamma^2\omega_{ij} \neq 0$, we find $\partial^2_{t^2}\gamma^2 = h(t)\gamma$ and, depending upon the choices of $h(t)$, we can either have oscillating γ (oscillating f_{as}) or $\gamma \to$const (absolute Maxwellians $f_{\text{abs Max}}$). If the forces are absent, $\mathbf{A}_0 = \mathbf{A} = 0$ but $(d/dt)\gamma^2\omega_{ij}(t) \neq 0$, we still find for $d = 2,4$ (no solution for $d = 3$) that $\gamma_{t \to \infty} \simeq t$ leading to $f_{as} \to 0$ when $t \to \infty$.

C. Exact solutions

In the next section, we establish a generalization of the Nikolskii[9] method when homogeneous source terms are present. In this method, homogeneous solutions are obtained in a first stage and associated inhomogeneous ones are deduced. Here we want to explain why the ansatz (2.4) leads naturally to an homogeneous formalism at an intermediate stage. We substitute (2.4) into the BE (2.3a) and take advantage of its dependence on two variables $\boldsymbol{\eta}(\mathbf{v},\mathbf{x},t)$ and t. In (2.3a) we can equate in both sides the coefficients of like $|\boldsymbol{\eta}|^{2p}$ powers and solve the resulting equations for $\alpha_0, \alpha_2, \Delta$. As a check of the results presented in this section we establish these relations in Appendix A. Here we use an equivalent method, taking into account both pseudoconservation laws for $\tilde{\rho}_i = \int \boldsymbol{\eta}^i \tilde{f}\, d\boldsymbol{\eta}$, $i = 0$ and 2, and equating the coefficients of $\boldsymbol{\eta}_v^4$ in both sides of (2.3a).

We first remark that from their definitions, there exist relations between the $\tilde{\rho}_i$'s and the coefficients α_i's, Δ of (2.4) in such a way that we can rewrite the ansatz (2.4) as

$$\tilde{\rho}_0 = \alpha_0 + \alpha_2\, d\Delta/2, \quad \tilde{\rho}_2 = d\alpha_0 \Delta + \alpha_2\, d(d+2)\Delta^2/2,$$

$$(2\pi\Delta)^{d/2}\tilde{f}(\boldsymbol{\eta}^2,t) = (\exp(-\boldsymbol{\eta}^2/2\Delta)) \times \tilde{\rho}_0[1 + (\tilde{\rho}_2/d\Delta\tilde{\rho}_0 - 1)(\boldsymbol{\eta}^2/2\Delta - 1)]. \qquad (2.4')$$

Second, we use the results of the above subsections (II A and II B). We notice that $L\boldsymbol{\eta}^2 = 0$ implies $\gamma^{-d}\,\text{Col}(\tilde{f}) = L\tilde{f} = \partial_t f(\boldsymbol{\eta},t)$, where $\tilde{f}$ must be considered as a function of the two independent variables $\boldsymbol{\eta}$, and t and ∂_t operates only on the t variable and not on $\boldsymbol{\eta}$. Except for the γ^{-d} factor, this is a homogeneous BE for a distribution with velocity $\boldsymbol{\eta}$. Consequently we have the conservation laws $\tilde{\rho}_i$

TABLE II. d-dimensional BE.

Homogeneous distribution $(\partial_\phi + a(\phi))F(\boldsymbol{\eta}_v,\phi) = \text{Col}(F) = S_d^{-1}\int d\Omega_d\, \sigma^d(x) d\boldsymbol{\eta}_w [F(\boldsymbol{\eta}_{w'})F(\boldsymbol{\eta}_{v'}) - F(\boldsymbol{\eta}_v)F(\boldsymbol{\eta}_w)]$, $d > 1$.

$$F = (2\pi\Delta)^{-d/2} e^{-\boldsymbol{\eta}^2/2\Delta}\left(\alpha_0 + \alpha_2 \frac{\boldsymbol{\eta}^2}{2}\right), \quad N_0 = \int F\, d\boldsymbol{\eta} = N_0(0)\exp\left(-\int_0^\phi a(\phi')d\phi'\right),$$

$$N_2 = \int F\boldsymbol{\eta}^2\, d\boldsymbol{\eta} = \frac{N_2(0)N_0}{N_0(0)}, \quad \alpha_2 = \Delta^{-2}(N_2 d^{-1} - N_0\Delta), \quad \alpha_0 = \Delta^{-1}(N_0\Delta(1 + d/2) - N_2/2),$$

$$\sigma_2^d = S_d^{-1}\int \sin^2\chi\,\sigma^d\, d\Omega_d, \quad \Delta_t + \sigma_2^d N_0 \Delta = \sigma_2^d N_2 d^{-1}, \quad \Delta = K(1 - \varphi).$$

$$F(\mathbf{w},\phi) = \frac{e^{-\mathbf{w}^2/2(1-\varphi)}}{(2\pi K(1-\varphi))^{d/2}} N_0\left[1 + \frac{\varphi}{2(1-\varphi)}\left(\frac{\mathbf{w}^2}{1-\varphi} - d\right)\right], \quad \mathbf{w} = \boldsymbol{\eta}K^{-1/2}, \quad K = N_2(0)d^{-1}(N_0(0))^{-1},$$

$$\varphi = \varphi(0)\exp\left(-\int_0^\phi \sigma_2^d N_0(\phi')d\phi'\right), \quad 0 < \varphi(0) < \left(1 + \frac{d}{2}\right)^{-1}.$$

Inhomogeneous distribution $f = (\partial_t + \mathbf{v}\cdot\partial_\mathbf{x} + \mathbf{A}_0(t)\cdot\partial_\mathbf{v} + \Lambda)f = \text{Col}(f)$, $d > 2$.

$$f(\mathbf{w},t) = \frac{e^{-\mathbf{w}^2/2(1-\varphi)}}{(2\pi K(1-\varphi))^{d/2}} \nu(t)\rho_0(0)\left[1 + \frac{1}{2}\frac{\varphi}{(1-\varphi)}\left(\frac{\mathbf{w}^2}{1-\varphi} - d\right)\right], \quad \mathbf{w} = \frac{\mathbf{c}\gamma(t)}{\sqrt{K}}, \quad \mathbf{c} = \mathbf{v} - \langle \mathbf{v} \rangle,$$

$$f_{as} = \frac{e^{-\mathbf{w}^2/2}}{(2\pi K)^{d/2}}\rho_0(0)\nu(t), \quad \rho_0(t) = \frac{\rho_0(0)\nu(t)}{\gamma^d} = \int f\, d\mathbf{v}, \quad \frac{\varphi(t)}{\varphi(0)} = \exp\left(-\sigma_2^{(d)}\int_0^t \rho_0(t')dt'\right),$$

$$\sigma_2^{(d)} = \frac{\int \sin^2\chi\,\sigma^{(d)}(\chi)d\Omega_d}{\int d\Omega_d}, \quad f > 0 \ \text{ if } \ 0 < \varphi(0) < \left(1 + \frac{d}{2}\right)^{-1},$$

$$\rho_2 = \rho_0[\gamma^{-2}Kd + \langle \mathbf{v} \rangle^2] = \int \mathbf{v}^2 f\, d\mathbf{v}\ \rho_0(0), \varphi(0), K \text{ constants } \boldsymbol{\eta} = \gamma\mathbf{c}.$$

(a) $\Lambda = \mathbf{A}(\mathbf{X},t)\cdot\partial_\mathbf{v}$, $\nu = 1, \gamma, \langle \mathbf{v} \rangle$ and $\mathbf{A}$ given by $\mathscr{L}\boldsymbol{\eta}^2 = 0$,

(b) $\Lambda = a_1(t)\partial_\mathbf{v}\cdot\mathbf{v} + a_2(t)$, $\nu = \exp\left(-\int_0^t (da_1 + a_2)dt'\right)$, γ and $\langle \mathbf{v} \rangle$ given by $(\mathscr{L} - a_1 d - a_2)\boldsymbol{\eta}^2 = 0$.

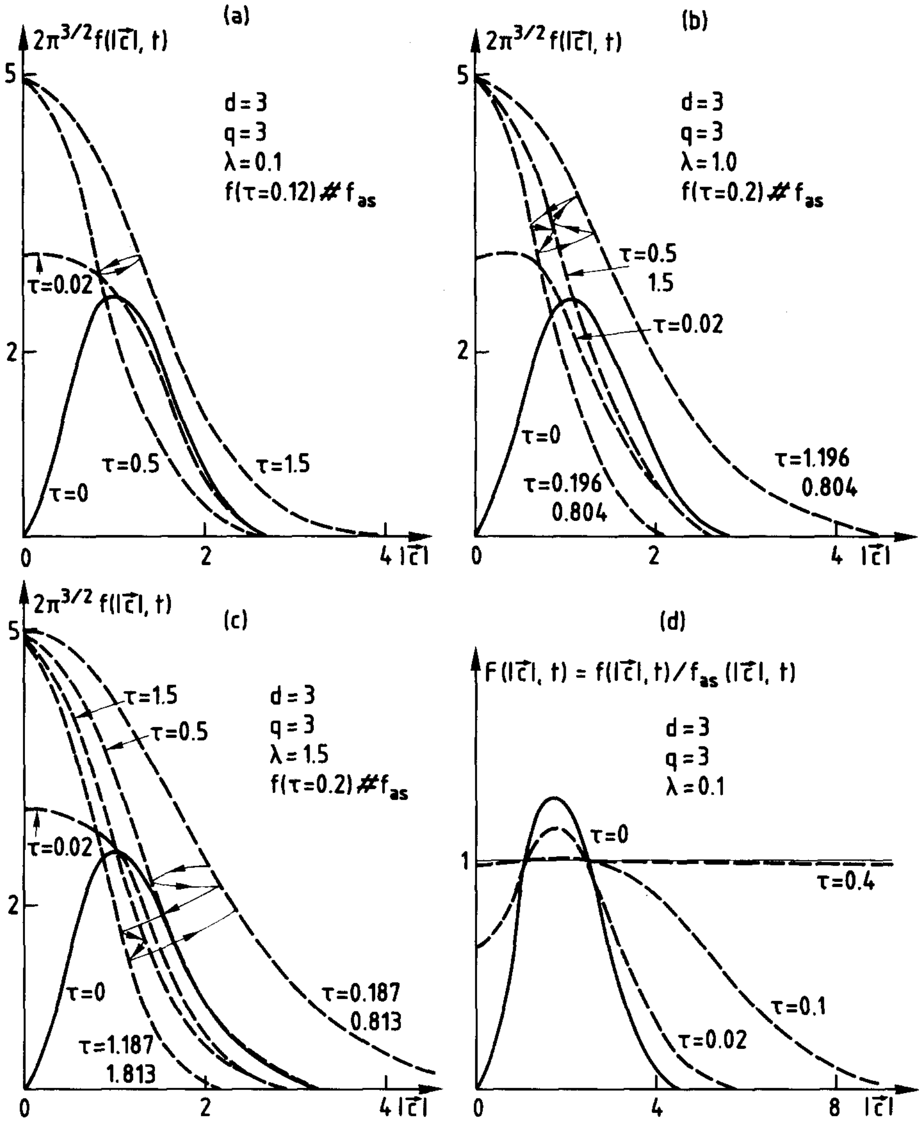

FIG. 1. Plot of $f(|\mathbf{c}|,t)$ against $|\mathbf{c}|$ for $\gamma = 1 + r(\sin t + \lambda \sin 3t)$ and either $\Lambda = a_1(t)\mathbf{x}\cdot\partial_\mathbf{v}$ or $a_1(t)\mathbf{v}\cdot\partial_\mathbf{v}$; $d = 3$, $r = 0.2$, $\rho_0(0) = 5$, $\varphi(0) = 0.4$, $K = 1$, $t = \tau\pi$: (a) $\lambda = 0.1$, (b) $\lambda = 1.0$, (c) $\lambda = 1.5$, (d) plot of the reduced $F(|\mathbf{c}|,t)$ for $\lambda = 0.1$.

$= \tilde{\rho}_i(0)$. Third, we substitute $\mathbf{v} = \gamma^{-1}\boldsymbol{\eta} + \langle \mathbf{v}\rangle$ into the true local energy and density $\int \mathbf{v}^i f\, d\mathbf{v} = \rho_i$, $i = 0$ and 2, notice that the integration over $d\boldsymbol{\eta}$ of odd $\boldsymbol{\eta}$ powers gives zero, and find

$$\rho_0(t) = (\gamma(t))^{-d}\tilde{\rho}_0(0), \quad \rho_0(0) = \tilde{\rho}_0(0) \quad \text{if } \gamma(0) = 1,$$
$$\rho_2(t,\mathbf{x}) = \rho_0(0)(\tilde{\rho}_2(0)/\gamma^2\rho_0(0) + \langle \mathbf{v}\rangle^2). \tag{2.7}$$

Fourth, only $\Delta(t)$ remains unknown. In $\partial_t \tilde{f} = \gamma^{-d}\,\mathrm{Col}(\tilde{f})$ we equate the coefficients of $\boldsymbol{\eta}_v^4$ and find $\partial_t \Delta = \alpha_2 \Delta^2 \sigma_2^{(d)} \gamma^{-d}$, $\sigma_2^{(d)}$ being a moment of $\sin^2\chi\sigma^d(\chi)$ defined in Table II. From (2.4′) we find $\alpha_2 = (\tilde{\rho}_2 - d\Delta\tilde{\rho}_0)/d\Delta^2$ and obtain a differential equation for Δ that is easily solved:

$$\partial_t \Delta = \sigma_2^{(d)}\rho_0(t)(\tilde{\rho}_2(0) d\rho_0(0) - \Delta),$$
$$\Delta(t) = K(1 - \varphi(t)),$$
$$\varphi = \varphi(0)\exp\left(-\sigma_2^{(d)}\int_0^t \rho_0(t')dt'\right),$$
$$\varphi(0) = 1 - \Delta(0)K^{-1}, \quad K = \tilde{\rho}_0(0)/d\rho_0(0). \tag{2.8}$$

Finally, substituting (2.7) and (2.8) into (2.4′) we obtain the explicit solution written down in Table II(a), where we have defined a new variable $\mathbf{w} = \boldsymbol{\eta}K^{-1/2}$ proportional to $\boldsymbol{\eta}$:

$$(2\pi K(1-\varphi))^{d/2} f(\mathbf{w},t) = \left(\exp\frac{-\mathbf{w}^2}{2(1-\varphi)}\right)\rho_0(0) \times\left[1 + \frac{\varphi}{2(1-\varphi)}\left(\frac{\mathbf{w}^2}{1-\varphi} - d\right)\right]. \tag{2.4″}$$

In this paper we are interested in nonvanishing asymptotic local density and assume $\rho_0(t) > \inf_t \rho_0(t)$ is a finite constant. It follows that φ in (2.8) goes to zero when $t\to\infty$ and $f \to f_{as}$,

$$(2\pi K)^{d/2} f_{as}(\mathbf{w}) = \rho_0(0)\exp(-\mathbf{w}^2/2), \quad \sqrt{K}\,\mathbf{w} = \boldsymbol{\eta} = \gamma\mathbf{c}, \tag{2.2″}$$

and the results found for f_{as} can be applied here. The discussion is in fact not trivial and in Sec. IV we thoroughly study the relaxation towards asymptotic regimes.

D. Oscillating Maxwellians

In (2.4″), (2.2′), or (2.2″) we discuss the Gaussian term $g(\mathbf{c}^2,t) = \exp[-(\gamma(t)\mathbf{c})^2/2]$ and recall [see Table I(a)] that, for $\mathbf{A}(x,t) \neq 0$, the equation for γ is of the type $\gamma_{tt} = h(t)\gamma(t)$. Clearly if γ oscillates, g will oscillate too, let us discuss periodic solution (the discussion is performed in Appendix A) and, for instance,

$$\gamma(t) = 1 + r(\sin t + \lambda \sin qt) > 0, \quad h(t) = \gamma_{tt}/\gamma, \tag{2.9}$$

with r,q,λ constants such that $\gamma > 0$ [we recall that the local density $\rho_0 = \rho_0(0)\gamma^{-d} > 0$]. If $\lambda = 0$, $\gamma(t)$ is periodic with period $T = 2\pi$ and the Gaussian oscillates between two Maxwellians. If $\lambda \neq 0$, q integer, we have competition between a circular function $T_1 = 2\pi$ and a harmonic $T_2 = 2\pi/q$, the important point being that *γ is a linear superposition* of circular functions: (i) $q = 2$, $|\lambda| < 0.5$, the Gaussian oscillates between two Maxwellians and four for $|\lambda| > 0.5$; (ii) $q = 3$, $-\frac{1}{3} < \lambda < \frac{1}{3}$, we have two Maxwellians and four for either $\lambda < -\frac{1}{3}$ or $\lambda > \frac{1}{3}$ (except three for $\lambda = 1$); and (iii) $q = 4$, $|\lambda| < 0.17$, we find two Maxwellians, six for $0.17 < |\lambda| < 0.25$, eight for $|\lambda| > 0.25$ (except seven for $|\lambda| = 0.92$). And so on for $q = 5,...$. For $q = Q^{-1}$, Q integer, we find an inverse situation when λ is varying; for $Q = 2$, $\lambda = 0$, we find two Maxwellians, four for $|\lambda| < 1/0.5 = 2$ (except three for $Q = 1$) and two for $|\lambda| = 2,...$. For q irrational, the Gaussian term is quasiperiodic with a countable set of extremal Maxwellians.

These features are different from those of the presently extensively studied nonlinear systems. We do not observe the cascade of doubling periods, and so on, when the parameter is varying, but, for instance, for q integer, the appearance of possible harmonics. Here, *the asymptotic behaviors of the distributions* being obtained with negligible collision terms (or the nonlinear part of the BE) *are provided by the linear part of the BE.*

In Figs. 1(a)–1(c) we present the inhomogeneous relaxation curves $f(|\mathbf{c}|,t)$ for $d = 3$, $q = 3$, $\lambda = 0.1,1,1.5$. We observe first the preasymptotic regime and second the asymptotic one with oscillations between two, three, and four Maxwellians. In Fig. 1(d) for the reduced distribution $F(|\mathbf{c}|,t) = f(|\mathbf{c}|,t)/f_{as}(|\mathbf{c}|,t)$ we observe that the relaxation towards 1 is from below.

E. Absolute Maxwellians

As said above, we exclude the trivial case $\mathbf{A} \equiv 0$, where the force identical to $\mathbf{A}_0(t)$ is not spatially dependent and for which, in the inhomogeneous formalism, $\gamma = t \to \infty$, $f \to 0$. When the force is really $\mathbf{x}$ dependent, as in the three possibilities quoted in Table I(a) (i) $\omega_{ij} = 0$, $\mathbf{A} \neq 0$; (ii) $\gamma^2\omega_{ij} = \text{const}$; and (iii) $(d/dt)\gamma^2\omega_{ij} \neq 0$ satisfies $\gamma_{tt} = h(t)\gamma$ with $h = a_1(t)$. In potential scattering theory, this equation is the S wave Schrödinger equation for potential $h(t)$ and wave function at zero momentum $\gamma(t)$. The asymptotic γ behaviors are either $\gamma \simeq t$ or $\gamma \simeq \text{const}$. In order to find absolute Maxwellians, we look at $\gamma \to \gamma(\infty) = \text{const}$. Then the solution, called the Jost function, satisfies a Volterra integral equation

$$\gamma(t) = \gamma(\infty) + \int_t^\infty (t' - t)\gamma(t')h(t')dt'. \tag{2.10}$$

This representation holds if the "potential" satisfies $t^{2+\epsilon}h(t) \underset{t\to\infty}{\to} 0$, $\epsilon > 0$. We want also $\gamma(0)$ finite or a "regular potential at the origin," i.e., $t^2h(t) \underset{t\to 0}{\to} 0$. Further $\rho_0 > 0$ requires $\gamma > 0$. For this whole class of potentials (here forces) exist $h(t) \to h(\infty)$ and $f \to f_{abs}(\mathbf{c})_{Max}$ written down in (1.3). There exist few cases for which the wave function (here γ) is known in closed form. Let us choose such a possibility where γ can be expressed in terms of Bessel functions:

$$h(t) = h(0)e^{-a_1t} \to \begin{cases} \gamma = I_0(2\sqrt{h(0)}e^{-a_1t/2})/I_0(2\sqrt{h(0)}), & \text{if } h(0) > 0, \\ \gamma = J_0(\sqrt{2|h(0)|}e^{-a_1t/2})/J_0(2\sqrt{|h(0)|}), & \text{if } h(0) < 0, \quad |h(0)| < 1.44, \end{cases} \tag{2.11}$$

the last restriction for J_0 coming from $\gamma > 0$ and a_1 is a constant.

III. LINEAR VELOCITY-DEPENDENT FORCES AND UNIFORM SOURCES

We still assume Maxwell particles and in $Lf = \text{Col}(f)$ write down both the linear operator and its differential part

$$L = \partial_t + \mathbf{v}\cdot\partial_\mathbf{x} + \mathbf{A}_0(t)\partial_\mathbf{v} + a_1(t)\partial_\mathbf{v}\cdot\mathbf{v} + a_2(t),$$
$$L - da_1 - a_2 = \partial_t + \mathbf{v}\cdot\partial_\mathbf{v} + (\mathbf{A}_0(t) + a_1(t)\mathbf{v})\partial_\mathbf{v}. \tag{3.1}$$

A. Assumptions and their consequences

We start with the same assumptions (2.2) as in Sec. II A and follow closely the lines of proofs. We first find $\gamma = \gamma(t)$, $(L - da_1 - a_2)\boldsymbol{\eta}^2$ linear in the η_i [see (B3) in Appendix B] and for the family of ansatz (iii) in (2.2), $\text{Col}(f)$ having only even $\boldsymbol{\eta}$ powers we must require $n = 1$. In Eq. (2.3) the vanishing of $\boldsymbol{\eta}^4\mathbf{v}\cdot\partial_\mathbf{x}\Delta$ still requires $\Delta(t)$, the equations for the $\boldsymbol{\eta}^4$ terms gives $\alpha_2(t)$, the vanishing of the $\boldsymbol{\eta}\eta_i$ and η_i terms leading at the end to $\alpha_0(t)$ and $(L - da_1 - a_2)\boldsymbol{\eta}^2 = 0$. This last result is the only change from Sec. II; only the differential part of the L operator operates for the constraints on $\boldsymbol{\eta}$. Finally (2.4) is the only possible ansatz.

B. f_{as} and $(L - da_1 - a_2)\boldsymbol{\eta}^2 = 0$

The study is done in Appendix B and we find, for $f_{as}(\boldsymbol{\eta},t)$, $\boldsymbol{\eta} = \gamma(t)\mathbf{c}$, $\mathbf{c} = \mathbf{v} - \langle\mathbf{v}\rangle$,

$$f_{as} = \text{const}\,\nu(t)\exp(-\boldsymbol{\eta}^2/2),$$
$$\nu = \exp\left(-\int_0^t (da_1 + a_2)dt'\right). \tag{3.2}$$

The different classes of $\langle\mathbf{v}\rangle$, γ, $\mathbf{A}_0$, a_1 are obtained in Appendix B. (Note that a_2 does not come into the discussion, which is independent of the source term.) However, the force being spatially independent there exists a smaller number of possibilities. Equation (2.5) for the mean velocity $\langle\mathbf{v}\rangle$ is still valid. We find two classes that are quoted in Table I(b): (i) with the antisymmetric tensor $\omega_{ij} = 0$ and $\gamma_t + a_1\gamma = \mu_0$ (μ_0 being constant); and (ii) with $\omega_{ij} \neq 0$, $(d/dt)\omega_{ij} \neq 0$, and $(\partial_t + \gamma_t\gamma^{-1})a_1 + \gamma_{tt}/\gamma - \Sigma\omega_{ij}^2 = 0$. Unfortunately this last class includes examples with $d = 2,4,...$ but not with the physical $d = 3$ dimensions.

We restrict our study to the $\omega_{ij} = 0$, (i) class for which

$$\gamma(t) = \exp\Big(- \int_0^t a_1 \, dt' \Big) \times \Big(1 + \mu_0 \int_0^t \exp\Big(\int_0^{t'} a_1 \, dt'' \Big) dt' \Big),$$

$$a_1 = (\mu_0 - \gamma_t)/\gamma. \quad (3.3)$$

For simplicity in the discussion of this subsection we assume here that

$Lf_{as} \simeq 0, \quad \text{for } t \gg 0.$

Let us first assume a pure velocity force with $a_2 = 0$ and $\nu = \exp(-\int_0^t da_1 \, dt')$. In order to have nontrivial ($\neq 0$ or $\neq \infty$) f_{as} when $t \to \infty$, both ν and the Gaussian terms must remain nontrivial. For instance, if $\alpha_{\inf} < \nu^{-1/d} = \exp(\int_0^t a_1 \, dt') < \alpha_{\sup}$, $\alpha_{\sup}$ and $\alpha_{\inf}$ being finite positive constants, then ν is nontrivial, but

$$(1 + \mu_0 t \alpha_{\inf}) \alpha_{\sup}^{-1} < \gamma(t) < (1 + \mu_0 t \alpha_{\sup}) \alpha_{\inf}^{-1}.$$

Consequently for inhomogeneous ($\mu_0 \neq 0$) solution $|\gamma \mathbf{c}| \to \infty$ and $f \to 0$. We have neither absolute nor oscillating Maxwellians.

Second, we add the source term a_2 in such a way that $a_1 d + a_2 = 0$ or $L = \partial_t + \mathbf{v} \cdot \partial_{\mathbf{x}} + (\mathbf{A}_0 + a_1 \mathbf{v}) \cdot \partial_{\mathbf{v}}$. Then the pure time factor $\nu(t) \equiv 1$ and only $\gamma(t)$ remains into the discussion of nontrivial f_{as}. For instance, assuming $a_1(t) = a_1(0) > 0$, then $\gamma \to$ const and $f_{as} \to \exp(-\text{const } \mathbf{c}^2)$, an absolute Maxwellian. We also can find oscillating Maxwellians. We can choose, for instance,

$$a_1(t) = a_1 + \partial_t \log(\lambda_0 + \lambda_1 \sin t + \lambda_2 \sin qt)$$

or

$$\gamma(t) = 1 + r(\sin t + \lambda \sin qt) > 0$$

and deduce a_1 from $a_1 = (\mu_0 - \gamma_t)\gamma^{-1}$. Note that for the other choices of the source term such that $a_1 d + a_2 \neq 0$, if we factorize the pure time factor $\nu(t)$ and define $q_{as} = f_{as} \nu^{-1}$, then as above we can find absolute or oscillating Gaussians for g_{as}.

C. Exact solutions

Instead of the direct substitution method used in Sec. II, we define a generalized Nikolskii[9,10] transform method. Essentially we begin by the determination of homogeneous solutions and by transforms, we deduce inhomogeneous ones. We start with an associate homogeneous problem when a source term is present. We consider d-dimensional BE's for distributions $F(\boldsymbol{\eta},\phi)$, $\boldsymbol{\eta}$ being the velocity, ϕ the time and assume $\boldsymbol{\eta}_v + \boldsymbol{\eta}_w = \boldsymbol{\eta}_{v'} + \boldsymbol{\eta}_{w'}$, $|\boldsymbol{\eta}_v|^2 + |\boldsymbol{\eta}_w|^2 = |\boldsymbol{\eta}_{v'}|^2 + |\boldsymbol{\eta}_{w'}|^2$:

$$(\partial_\phi + a(\phi))F(\boldsymbol{\eta}_v,\phi) = \text{Col}(F),$$

$$\text{Col}(F) = S_d^{-1} \int d\Omega_d \, \sigma^d(\chi) d\boldsymbol{\eta}_w (F(\boldsymbol{\eta}_{v'})F(\boldsymbol{\eta}_{w'}) - F(\boldsymbol{\eta}_v)F(\boldsymbol{\eta}_w)). \quad (3.4)$$

We determine the corresponding BKW solution. (See the first part of Table II.) Let us assume the ansatz solution (2.4), $f \to F$, $t \to \phi$, $\boldsymbol{\eta} \leftrightarrow \boldsymbol{\eta}$. We write both the particle, energy conservation law and the relations between the α_2's and the macroscopic quantities for F (local density N_0, local energy N_2). From the vanishing of the coefficients of $\boldsymbol{\eta}^4$ in (3.3) we obtain the last necessary relation, which is a differential equation for Δ. We find for $F(\mathbf{w},\phi)$ the solutions written down in Table II, $\mathbf{w}$ being proportional to $\boldsymbol{\eta}$ and ϕ appearing only through the local density $N_0(\phi)$ associated with the homogeneous distribution F.

We come back to our inhomogeneous solution $f(\mathbf{v},\mathbf{x},t)$, which, from the assumptions (3.1), can be rewritten as $\tilde{f}(\boldsymbol{\eta}(\mathbf{v},\mathbf{x},t),\phi(t))$, $(L - da_1 - a_2)\boldsymbol{\eta}^2 = 0$ and satisfies the BE

$$(\phi_t \, \partial_\phi + a_1 + da_2)\tilde{f} = \gamma^{-d} \, \text{Col}(\tilde{f}). \quad (3.4')$$

$\text{Col}(\tilde{f})$ written down in (2.3a) is the same as (3.4) with $\tilde{f} \to F$. If $\gamma^d \phi_t = 1$, $\gamma^d(a_1 + da_2) = a(\phi)$, or

$$\exp\Big(- \int_0^\phi a(\phi') d\phi' \Big) = \exp\Big(- \int_0^t (da_1 + a_2) dt' \Big) = \nu(t)$$

we can identify both homogeneous and inhomogeneous formalisms leading to the solution $f = \tilde{f} = F$ with $\mathbf{w} = \gamma(t)\mathbf{c}K^{-1}$. The mean velocity $\langle \mathbf{v} \rangle$ and $\gamma(t)$ are those determined from $(L - da_1 - a_2)\boldsymbol{\eta}^2 = 0$ and quoted in Table I(b). However, it remains to express the macroscopic quantities: local density and energy. For instance,

$$\rho_0 = \int f d\mathbf{v} = \gamma^{-d} N_0(\phi) = \rho_0(0) \nu \gamma^{-d}, \quad \text{if } \gamma(0) = 1;$$

$$\rho_2 = \int \mathbf{v}^2 f d\mathbf{v} = \gamma^{-d} \int (\boldsymbol{\eta}^2 \gamma^{-2} + \langle \mathbf{v} \rangle^2) f d\eta = \gamma^{-d}(N_2 \gamma^{-2} + \langle \mathbf{v} \rangle^2 N_0) = \rho_0(K d\gamma^{-2} + \langle \mathbf{v} \rangle^2).$$

The resulting solution $f(\mathbf{w},t)$, $\mathbf{w} = K^{-1/2}\boldsymbol{\eta}$ is written down in Table II(b).

As in Sec. II, assuming $\rho_0(t)$ larger than a finite constant, then $\varphi \to 0$ and for fixed $\mathbf{w}$, $f(\mathbf{w},t) \to f_{as}(\mathbf{w},t)$, which is (2.2″) multiplied by a time factor $\nu(t)$,

$$(2\pi K)^{d/2} f_{as}(\mathbf{w},t) = \rho_0(0)\nu(t)\exp(-\mathbf{w}^2/2),$$

$$\mathbf{w} = K^{-1/2}\gamma \mathbf{c}. \quad (3.2')$$

The previous results for f_{as} can be applied, while a complete discussion of the relaxation of f is performed in Sec. IV.

D. Velocity force plus source term $da_1 + a_2 = 0$ or $\Lambda = a_1(t)\mathbf{v} \cdot \partial_{\mathbf{v}}$

Let us compare the solutions $f(\mathbf{c},t)$ of Tables I(a), II(a) and I(b), II(b) and choose the same $\gamma(t)$ either for spatial force or for velocity plus source term $a_1 d + a_2 = 0$. In both cases, $\nu = 1$, $\rho_0 = \rho_0(0)\nu^{-d}$ and $f(\mathbf{c},t)$ are the same although the forces $a_1 = \gamma_{tt}/\gamma$ [Table I(a)] or $a_1 = (\mu_0 - \gamma_t)\gamma^{-1}$ [Table I(b)] as well as $\langle \mathbf{v} \rangle$ are different. Consequently, we can associate in both cases the same distributions $f(\mathbf{c},t)$ or reduced distributions $F(\mathbf{c},t) = f(\mathbf{c},t)/f_{as}(\mathbf{c},t)$.

As a first consequence, if we choose $\gamma = 1 + r(\sin t + \lambda \sin qt)$ and deduced $a_1(t)$ from (3.3), then Fig. 1 represents also the *oscillating Maxwellians* for this mixing of velocity force plus source term.

Second, we look at the *absolute Maxwellians* written down in (1.3) obtained when $\gamma(t) \to \gamma(\infty)$, a constant. *A priori* we could choose any functional $\gamma(0) > 0$, $\gamma \to$const, and deduce the force $a_1(t) = (\mu_0 - \gamma_t)\gamma^{-1}$. Here also the same $f(\mathbf{c},t) \to f_{\text{abs Max}}$ can be associated either with spatial force or with spatial plus source term. For instance we can choose the Bessel functions $\gamma(t)$ of (2.11) and deduce the velocity force. In general the expressions for $a_1(t)$ become complicated and for simplicity we choose a simple solution $a_1(t) = a_1 = \text{const} > 0$,

$$\gamma(t) = \gamma(\infty) + e^{-a_1 t}(1 - \gamma(\infty), \gamma(\infty)) = \mu_0/a_1. \tag{3.5}$$

E. Pure velocity force ($a_2 = 0$): Table I(b) and γ in (3.3)

In the homogeneous formalism, the above difficulty (Sec. III C), of both nontrivial γ and ν, does not occur: $\mu_0 = 0$, $a_1 = \gamma_t \gamma^{-1}$, $\rho_0(t) = \rho_0(0)$, and $\nu = \gamma^d$. If γ oscillates, then ν oscillates too and if $\gamma(\infty) = \text{const}$, then $\nu \to$const and $f \to$a Maxwellian. If γ oscillates (for instance, $\gamma = \exp[-(\sin t + \lambda \sin qt)]$ or $\gamma = 1 + r(\sin t + \gamma \sin qt))$, we can define $g = f\nu^{-1}$ and oscillating behaviors occur for g. For instance, we find oscillations between two, three, four,... Gaussians $\exp[-(\gamma \mathbf{c})^2/2]$. For the true distribution $f = g\nu$, there exist more complicated oscillations because the extrema have supplementary $|\mathbf{c}|,t$ values. All these features have been studied in the companion paper[1] for the $d = 1$ Kac model and they are practically unchanged here in higher dimensions (the main difference being that the Kac model distributions can have an odd velocity part).

In the inhomogeneous $\mu_0 \neq 0$ formalism, the situation is different, the difficulty with both ν,γ nontrivial remaining. All the discussion can be done with $\gamma(t)$, which gives other quantities

$$a_1 = (\mu_0 - \gamma_t)\gamma^{-1},$$

$$\nu\gamma^{-d} = \frac{\rho_0(t)}{\rho_0(0)} = \exp\left(-d\mu_0 \int_0^t \gamma^{-1}\, dt'\right)$$

$$= \left[1 + \mu_0 \int_0^t \exp\left(\int_0^{t'} a_1\, dt''\, dt'\right)\right]^{-d}, \tag{3.6}$$

assuming $\gamma(t) > 0$ (or < 0) in order to avoid problems in the integration of γ^{-1}.

(i) For oscillating γ, $0 < \gamma_{\text{inf}} < \gamma < \gamma_{\text{sup}}$, γ_{inf} and γ_{sup} being finite; then $\int_0^t \gamma^{-1}\, dt' \to \infty$ while the Gaussian $\exp[-(\mathbf{c}\gamma)^2/2]$ and $g = f\nu^{-1}$ have oscillating behaviors. If $\mu_0 > 0$ (or < 0), we find ρ_0 and $\nu \to 0$ (or ∞), $f \to 0$ (or ∞), for $|\mathbf{c}|$ fixed. These properties can be checked with $\gamma = 1 + r(\sin t + \lambda \sin qt)$.

(ii) For $\gamma \to \gamma(\infty)$, a constant, we find

$$\mu_0 \int_0^t \gamma^{-1}\, dt' \simeq t\mu_0(\gamma(\infty))^{-1} \to \pm\infty.$$

It follows that if $\mu_0\gamma(\infty)^{-1} > 0$ (or < 0), ρ_0 and $\nu \to \infty$ (or 0) and $f \to \infty$ (or 0). We can check these results with the constant force $a_1 = a_1(0)$ for which $\mu_0\gamma(\infty)^{-1} = a_1(0)$ and $\gamma = \gamma(\infty)\ [1 - \exp(-a_1(0)t)] + \exp[-a_1(0)t]$. The breakdown of the inhomogeneous formalism may be due here to the lack of boundary condition, compensated in Sec. III D by the introduction of a particular source term.

(iii) γ [see (3.3)] $\to \infty$ and $\rho_0(\infty)$ [see (3.6)] exists if, for instance, either $a_1(t) = -\bar{a}_1 < 0$ or $a_1(t) = -\bar{a}_1(1+t)^{-1}$, $\bar{a}_1 > 1$. Then we can give a meaning to the asymptotic singular time behavior of f. Let us rewrite f_{as} (Table II) as a local Maxwellian

$$f_{\text{as}} = (2\pi T)^{-d/2}\rho_0(t)\exp[-(\mathbf{v} - \langle \mathbf{v}(t)\rangle)^2/2T],$$
$$T(t) = K\gamma^{-2}. \tag{3.7}$$

If $T \to 0$ (or $\gamma \to \infty$) when $t \to \infty$, then $f_{\text{as}} \simeq \rho_0(\infty)\delta(\mathbf{v} - \langle \mathbf{v}(t = \infty)\rangle)$. Notice that similar singular asymptotic behaviors are needed for strong shocks profiles, t being replaced by one of the $\mathbf{x}$ space components.[3]

IV. RELAXATION TOWARDS ASYMPTOTIC REGIMES

If $\int_0^t \rho_0(t')dt' \to +\infty$ [in particular if $\inf_t \rho_0(t)$ is finite], then $\varphi \simeq 0$ when t becomes large and $f(|\mathbf{w}|,t) \simeq f_{\text{as}}(|\mathbf{w}|,t)$ given by (2.2″) or (3.2′). We define a reduced distribution $\bar{F}(|\mathbf{w}|,t) = f/f_{\text{as}}$,

$$\bar{F}(|\mathbf{w}|,t) = (1-\varphi)^{-d/2}\left[1 + \frac{\varphi}{2(1-\varphi)}\left(\frac{\mathbf{w}^2}{1-\varphi} - d\right)\right]$$
$$\times \exp\left[-\left(\frac{\varphi \mathbf{w}^2}{2(1-\varphi)}\right)\mathbf{w}\right] = \gamma \mathbf{c} K^{-1/2}, \tag{4.1}$$

and study its relaxation towards 1 for $|\mathbf{w}|$ fixed and t going to infinity. We notice that while f_{as} is a Gaussian $\exp(-\mathbf{w}^2/2)$ multiplied eventually by a time factor $\nu(t)$, this ν term disappears in the ratio $\mathbf{F}$. The study of $\mathbf{F}(|\mathbf{w}|,t)$ is quite justified if f_{as} is itself an asymptotic regime, like it is for oscillating Maxwellians. On the contrary when the equilibrium state is an absolute Maxwellian, then $f_{\text{as}}(|\mathbf{c}|,t)$ relaxes towards $f_{\text{abs Max}}$ written down in (1.3) and it is more convenient to define another reduced distribution

$$F(|\mathbf{c}|,t) = f(|\mathbf{c}|,t)/f_{\text{abs Max}}(|\mathbf{c}|),$$
$$F(|\mathbf{c}|,t) = (1-\varphi)^{-d/2}$$
$$\times \left[1 + \frac{\varphi}{2(1-\varphi)}\left(\frac{(\gamma \mathbf{c})^2}{K(1-\varphi)} - d\right)\right]$$
$$\times \exp\left[\frac{\mathbf{c}^2}{2K}\left(\frac{\gamma^2(t)}{1-\varphi} - \gamma^2(\infty)\right)\right]. \tag{4.2}$$

A. Relaxation towards $f_{\text{as}}(|\mathbf{w}|,t)$ and study of $\bar{F}$

First a previous argument,[7] given for the homogeneous reduced BKW distribution $\bar{F}(|\mathbf{v}|,t)$ without external forces, can be applied for $\bar{F}(|\mathbf{w}|,t)$. We want to prove that $\bar{F} < 1$ if $|\mathbf{w}|^2 > 2d + 6$. Setting $|\mathbf{w}| = d + 4 + \tilde{w}^2$ we rewrite $\bar{F}$ as

$$F(|\mathbf{w}|,t) = \left[\frac{\exp[-(\varphi/1-\varphi)(d/2+2)]}{(1-\varphi)^{d/2+2}}\right]$$
$$\times \left[\left(1 + \varphi^2\left(\frac{d}{2} + 1\right) + \varphi\frac{|\tilde{w}|^2}{2}\right)\right.$$
$$\left.\times \left(\exp\left(-\frac{\varphi}{1-\varphi}\frac{|\tilde{w}|^2}{2}\right)\right)\right], \quad \varphi > 0. \tag{4.3}$$

The first bracket is always less than (or equal to) 1, and so is the second one for $|\tilde{w}|^2 > d + 2$; whence the result that means that the relaxation towards f_{as} is always from below.

Second, we can investigate the large t (small φ), fixed

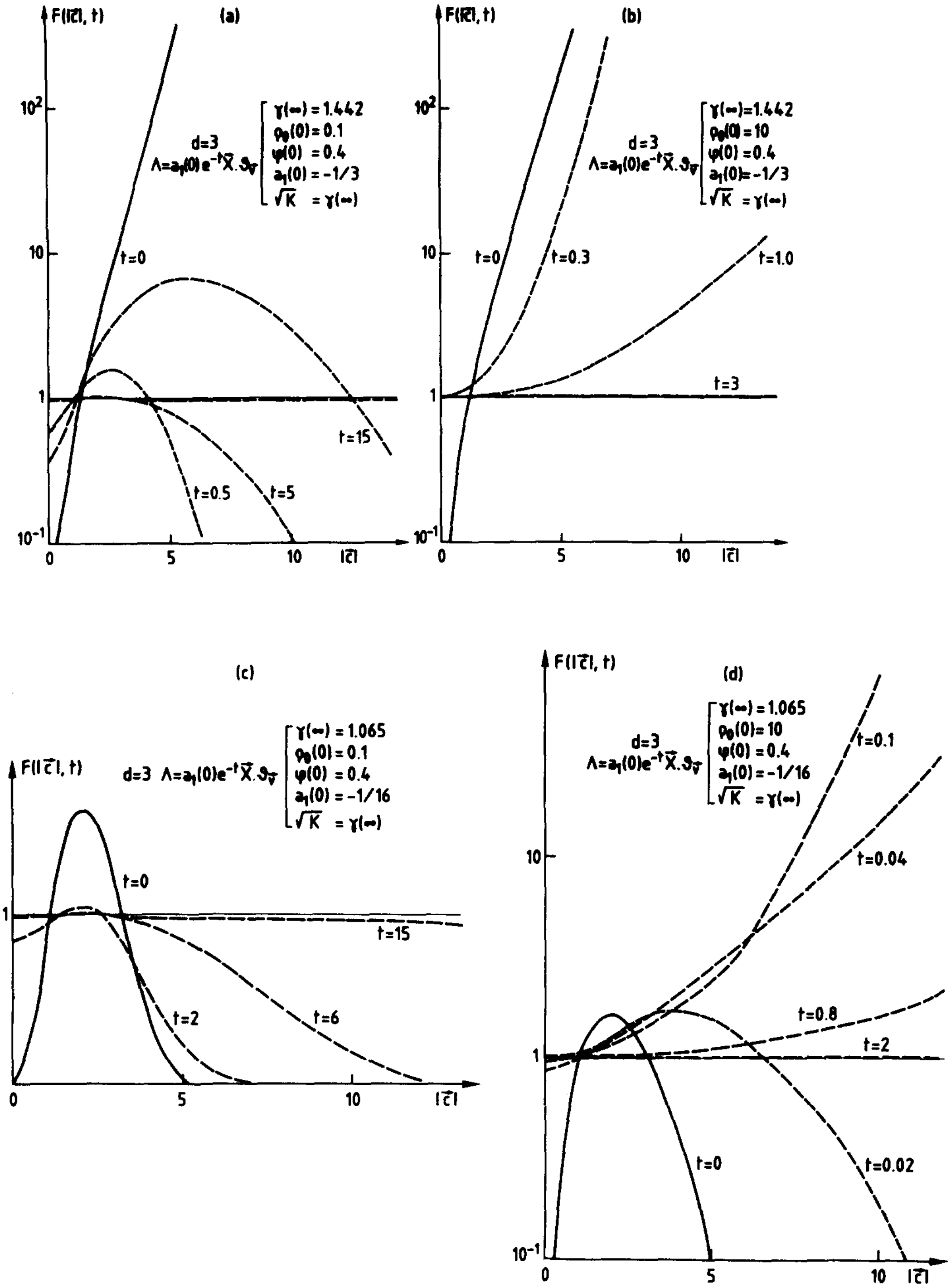

FIG. 2. Plot of the reduced $F(|c|,t)$ against $|c|$, given by (2.11) for $\Lambda = a_1(0)e^{-t}\,\mathrm{x}\cdot\partial_{\mathrm{v}}$, $d=3$, $\sqrt{K}=\gamma(\infty)$, $\varphi(0)=0.4$: (a) $a_1(0) = -\frac{1}{3}$, $\gamma(\infty)=1.442$, $\rho_0(0)=0.1$; (b) the same as (a) but $\rho_0(0)=10$; (c) $a_1(0) = -\frac{1}{16}$, $\gamma(\infty)=1.065$, $\rho_0(0)=0.1$; and (d) the same as (c) but $\rho_0(0)=10$.

$|\mathbf{w}|$ behavior $\overline{F}\simeq 1+\varphi\,\partial_\varphi \overline{F}+(\varphi^2/2)\partial^2_{\varphi^2}\overline{F}$ around $\varphi=0$ and find

$$\overline{F}-1\simeq(\varphi^2/8)(4\mathbf{w}^2-2d-(\mathbf{w}^2-d)^2)<0,$$
$$\mathbf{w}^2>d+2+\sqrt{2d+4}, \qquad (3.4')$$

still showing that the relaxation is from below. If the local density $\rho_0(t)=\rho_0(0)$ is conserved, the relaxation time is $T(2\sigma_2^d\,\rho_0(0))^{-1}$. Otherwise, in order to have an estimation, let us restrict either to spatial force or to velocity force plus source with $da_1+a_2=0$, where $\rho_0=\rho_0(0)\gamma^{-d}$, $\varphi=\varphi(0)\exp(-\sigma_2^{(d)}\int_0^t\rho_0\,dt')$. Let us assume $0<\gamma_{\inf}<\gamma<\gamma_{\sup}$, then the relaxation time T is such that $(\gamma_{\inf})^d<2T\sigma_2^{(d)}\rho_0(0)<(\gamma_{\sup})^d$.

B. Relaxation towards $f_{\mathrm{abs\,Max}}(|\mathbf{c}|)$ and study of F

The results of Sec. IV A agree with those of the homogeneous BKW solutions (without force). The relaxation from below means that no Tjon[16] effect can occur, and this was a drawback of the known homogeneous similarity solutions. On the contrary, when the effect exists there is at intermediate times, a population of high-velocity particles larger than the one present at initial time or at equilibrium. Here,

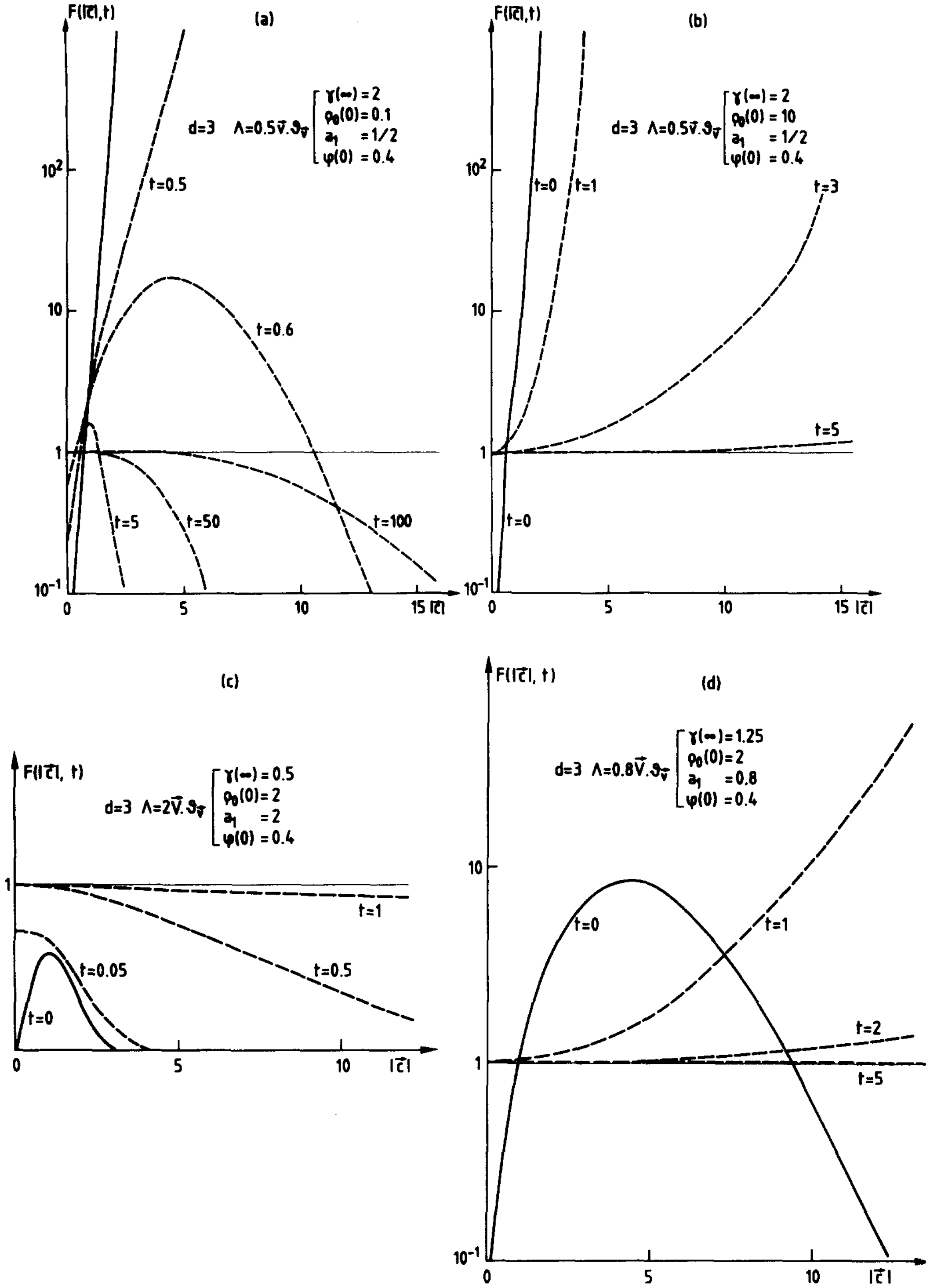

FIG. 3. Plot of the reduced $F(|\mathbf{c}|,t)$ against $|\mathbf{c}|$, given by (3.5) for $\Lambda = a_1(0)\mathbf{v}\cdot\partial_{\mathbf{v}}$, $d = 3$, $\sqrt{K} = \gamma(\infty)$, $\varphi(0) = 0.4$, $\mu_0 = 1$: (a) $a_1(0) = 0.5$, $\gamma(\infty) = 2$, $\rho_0(0) = 0.1$; (b) the same as (a) but $\rho_0(0) = 10$; (c) $a_1(0) = 2$, $\gamma(\infty) = 0.5$, $\rho_0(0) = 2$; and $a_1(0) = 0.8$, $\gamma(\infty) = 1.25$, $\rho_0(0) = 2$.

with the introduction of spatial forces (Sec. II) or velocity force plus source term [$da_1 + a_2 = 0$ (Sec. III)], we are able to construct inhomogeneous solutions relaxing towards $f_{\text{abs Max}}$ and it is natural to see whether this drawback can disappear.

Let us define $\bar{\gamma} = (\gamma/\gamma(\infty))^2 - 1 \to 0$ when $t \to \infty$, recall that $\gamma(0) = 1$, and rewrite the Gaussian appearing in F, given by (4.2), as

$$G(|\mathbf{c}|,t) = \exp\left(- \left[\frac{(\mathbf{c}\gamma(\infty))^2}{2K} \frac{(\bar{\gamma} + \varphi)}{1 - \varphi} \right] \right). \tag{4.4}$$

We have $\varphi > 0$, $1 - \varphi > 0$, $\varphi \to 0$, $|\bar{\gamma}| \to 0$, *but* $\bar{\gamma} + \varphi$ *can be either positive* ($G < 1$) or *negative* ($G > 1$). Introducing initial and asymptotic conditions on $\bar{\gamma} + \varphi$ we find for G, the dominant par of F, four possibilities.

(i) $\bar{\gamma} + \varphi > 0$, $\forall t$, $G < 1$, and the relaxation for $|\mathbf{c}|$ large is always from below.

(ii) $\bar{\gamma} + \varphi > 0$ at $t = 0$ becomes negative at large t. Then for large $|\mathbf{c}|$, $G(t = 0) < 1$ but $G \gg 1$ for large t. There exist intermediate times for which $G(t) > G(0)$, $G(t) > G(\infty)$, and the effect can exist.

(iii) $\bar{\gamma}+\varphi<0$ at $t=0$ becomes positive at large t.

(iv) $\bar{\gamma}+\varphi<0, \quad \forall t$.

These last two possibilities correspond to some antieffect because the large $|\mathbf{c}|$ populations become smaller at intermediate times than at $t=0$. In order to find the conditions under which $F\to 1^{\pm}$ (that we approximate by $G\to 1^{\pm}$), let us seek rough estimates of $\bar{\gamma}+\varphi$ at large t. From the definitions (Table II) of ρ_0, $\bar{\varphi}$ we find $\varphi \simeq \varphi(0)\exp[\sigma_2^d \rho_0(0)(\gamma(\infty))^{-d}t]$. If for large t, $\bar{\gamma}>0$, then $G\to 1^-$ while $G\to 1^+$ only if both $\bar{\gamma}$ becomes negative and decreases less than φ. In order to go on we consider explicit $\gamma(t)$ and if $\bar{\gamma}<0$, we give a criterion for the asymptotic sign of $\bar{\gamma}+\varphi$.

First, for the spatial $a_1(0)(\exp(-a_1 t)\mathbf{x}\cdot\partial_{\mathbf{v}}$ force we look at the Bessel solutions for γ, written down in (2.11) with $h=a_1(0)\exp(-a_1 t)$, a_1 being a constant. For the I_0 solutions, $a_1(0)>0$, we find $\bar{\gamma}=I_0^2-1>0$ and a relaxation from below. For the J_0 solutions, $a_1(0)<0$, $\bar{\gamma}=J_0^2-1$, and $\bar{\gamma}+\varphi$ can be either positive or negative, and in the following we restrict to these more interesting solutions. For large t we find $\bar{\gamma}\simeq 2a_1(0)\exp(-a_1 t)<0$ decreases like an exponential. We can compare $\bar{\gamma}$ with φ and look at the sign of the sum $\bar{\gamma}+\varphi$. When as for the J_0 solutions, $\bar{\gamma}$ becomes negative, we define a criterion

$$\text{crit}=\frac{\alpha_1(\gamma(\infty))^d}{\sigma_2^{(d)}\rho_0(0)}\to\begin{cases}\text{crit}\ll 1\to F \text{ (or } G)\to 1^+,\\ \text{crit}\gg 1\to F \text{ (or } G)\to 1^-,\end{cases} \tag{4.5}$$

which predicts whether the relaxation will be from above ($\bar{\gamma}+\varphi\to<0$) or from below ($\bar{\gamma}+\varphi\to>0$) and test in the three-dimensional case. The numerical calculations agree with this theoretical analysis. In Fig. 2 we choose $a_1(0)<0$ and so the Bessel solutions for γ. In Figs. 2(c) and 2(d) $\bar{\gamma}(0)+\varphi(0)<0$ but crit$\neq$30 and 0.3; in Figs. 2(c) and 2(d), $\bar{\gamma}(0)+\varphi(0)>0$ but crit$\neq$12 and 0.12. In Fig. 2(d) we observe for $t\gg 1$ a population of high $|\mathbf{c}|$ particles larger than at $t=0$ and $t=\infty$.

Second, for the velocity force plus source term $a_1\mathbf{v}\cdot\partial_{\mathbf{v}}$, $a_1>0$ being a constant, we look at the γ solutions written down in (3.5). We find both that $\bar{\gamma}(0)=\gamma(\infty)^{-2}-1$ and for large t, $\bar{\gamma}\simeq 2e^{-a_1 t}(\gamma(\infty)^{-1}-1)$ becomes negative (positive) for $\gamma(\infty)>1$ (<1). We define the same criterion as in (4.5) predicting the same results. For the numerical calculations, in Figs. 3(a) and 3(b) we have $\bar{\gamma}(0)+\varphi(0)<0$, $\gamma(\infty)=2$, but crit$\neq$40 and 0.4. In Fig. 3(c), $\gamma(\infty)=0.5$ predicts a relaxation from below (independently of the value of crit), while in Fig. 3(d), $\bar{\gamma}(0)+\varphi(0)>0$, $\gamma(\infty)=1.25$, and crit$\neq$0.78. In Fig. 3(d), at intermediate times, we still observe the excess of high $|\mathbf{c}|$ particles.

At the end we notice that the microscopic interaction $\sigma^{(d)}(\chi)$ is present in the criterion through the presence of a moment of the cross section and the initial distribution through $\rho_0(0)$, the local density at $t=0$.

V. CONCLUSION

The introduction of external forces gives the possibility of obtaining *exact inhomogeneous solutions with absolute Maxwellians as equilibrium states*. Furthermore, we find a rich variety of different relaxations towards equilibrium, for instance, large high-velocity populations at intermediate times. The relaxations does not depend only on the initial conditions but also on microscopic model of interaction and of course on the force. If we compare with the previously known homogeneous and inhomogeneous BKW solutions many of these properties are new for ($d>1$)-dimensional models although they were observed with the $d=1$ Kac model solutions.[13,14]

Concerning the *oscillating Maxwellians*, the results obtained in the companion paper[1] for the homogeneous Kac model $d=1$, confirm those presented here for the homogeneous and inhomogeneous $d>1$ models that preserve momentum conservations. These oscillating solutions, first obtained by Boltzmann from the linear part of the BE's exist also in the complete nonlinear BE. They constitute a new class of distributions, not explored in the usual study of the BE.

At the end we think it worthwhile to summarize the different stages of the results resulting from the following twofold original assumptions: (i) the distribution $f(\mathbf{v},\mathbf{x},t)$ depends on $\mathbf{v}$ through a variable $\boldsymbol{\eta}=\gamma(\mathbf{v}-\mathbf{v}_0)$ ($\gamma,\mathbf{v}_0$ unknown) and it is the product of a Gaussian in $\boldsymbol{\eta}$ by an even $\boldsymbol{\eta}$ polynomial with $\mathbf{x},t$ dependent coefficients; and (ii) asymptotically f reduces to a Gaussian in $\boldsymbol{\eta}$ (with $\mathbf{x},t$ dependent coefficients), which means that the collision term vanishes and f becomes a solution of the linear part of the BE. We deduce successively that γ is $\mathbf{x}$ independent, $\mathbf{v}_0=\langle\mathbf{v}\rangle$, the polynomial is of the first order, all the coefficients are only t dependent, and $\boldsymbol{\eta}^2$ is solution of the differential part of the linear operator of the BE. Consequently $\mathbf{x}$ appears only through $\langle\mathbf{v}\rangle$ and $f\equiv f(\boldsymbol{\eta},t)$ can be seen as a homogeneous distribution with velocity $\boldsymbol{\eta}$. Then we can either directly solve the "homogeneous relations" or use the Nikolskii[9] approach. Can we enlarge the class of exact solutions by weakening these assumptions?

Note added in proof: In this paper, for Maxwell molecules, the homogeneous distributions associated with the inhomogeneous ones are always BKW (or generalized) distributions. In a recent work[17] we study both other classes of associated distributions and other intermolecular potentials.

APPENDIX A

1. Equations for $Lf_{as}=0$, $d>2$ and $L=\partial_t+\mathbf{v}\cdot\partial_{\mathbf{x}}+(\mathbf{A}_0(t)+\mathbf{A}(x,t))\cdot\partial_{\mathbf{v}}$

We define

$$f_{as}=\nu(\mathbf{x},t)\exp(-\boldsymbol{\eta}^2/2),$$

$$\boldsymbol{\eta}=\gamma(t)(\mathbf{v}-\langle\mathbf{v}\rangle),$$

$\langle\mathbf{v}\rangle$ depends upon $\mathbf{x},t$, and we introduce the d components $A_{0i}, A_i, \langle\mathbf{v}\rangle_i, v_i, x_i, \eta_i$. Now $Lf_{as}=0$ gives

$$(\partial_t+\mathbf{v}\cdot\partial_{\mathbf{x}})\log\nu=L\boldsymbol{\eta}^2/2. \tag{A1}$$

The rhs contains powers $v_i v_j$, v_i^2, but not the lhs. Equating to zero the coefficients of like powers, we find

$$\partial_{x_i}\langle\mathbf{v}\rangle_i=\partial_t\log\gamma,\quad \partial_{x_i}\langle\mathbf{v}\rangle_j+\partial_{x_j}\langle\mathbf{v}\rangle_i=0, \tag{A2}$$

or, with the introduction of an antisymmetric tensor $\omega_{ij}(t)$: $\omega_{ij}+\omega_{ji}=0$, $\omega_{ii}=0$,

$$\langle\mathbf{v}\rangle_i=\alpha_i(t)+x_i\,\partial_t\log\gamma+\sum_{l\neq i}\omega_{il}(t)x_l. \tag{A2'}$$

For $d=3$ we have

$$\langle \mathbf{v}\rangle = \boldsymbol{\alpha} + \mathbf{x}\,\partial_t \log\gamma + \boldsymbol{\omega}\wedge\mathbf{x}, \quad \boldsymbol{\omega} = (\omega_{32},\omega_{13},\omega_{21}).$$

In $L\boldsymbol{\eta}^2$ remains only v_i powers (or η_i) and const., taking into account (A2′)

$$\frac{\gamma^{-1}L\boldsymbol{\eta}^2}{2} = \sum \eta_i \zeta_i(\mathbf{x},t),$$

$$\zeta_i = A_i + A_{0i} - (\partial_t + \gamma_t\gamma^{-1})\langle \mathbf{v}\rangle_i - \sum_j \langle \mathbf{v}_j\rangle \omega_{ij}. \quad \text{(A3)}$$

In (A1) the coefficients of η_i and const give finally

$$\gamma^{-2}\,\partial_{x_i}\log\nu = \zeta_i, \quad (\partial_t + \langle \mathbf{v}\rangle\cdot\partial_{\mathbf{x}})\log\nu = 0. \quad \text{(A1′)}$$

2. $L\boldsymbol{\eta}^2 = 0$ or $\zeta_i = 0$ in (A3)

First, from (A1′) we find for f_{as} that $\nu=$ const; second, we determine the compatible γ, $\langle \mathbf{v}\rangle$, and $\mathbf{A}$. From (A3) we find

$$\partial_{x_j}A_i - \partial_{x_i}A_j = 2\gamma^2 \frac{d}{dt}(\gamma^2\omega_{ij})$$

or

$$A_i = -\partial_{x_i}\psi(\mathbf{x},t) + \sum_l x_l \gamma^{-2}\frac{d}{dt}\omega_{il}\gamma^2,$$

showing that $\mathbf{A}(\mathbf{x},t)$ is the sum of a conservative force and a nonconservative one if $(d/dt)\omega_{ij}\gamma^2 \neq 0$. Writing $A_i = x_i A_i^{(1)}(t) + A_i^{(2)}(x_1,...,x_j,...,j\neq i)$, from (A2′) and (A3), we find

$$A_{0i} = (\gamma_t\gamma^{-1} + \partial_t)\alpha_i + \sum \alpha_j\omega_{ij}, \quad \text{(A4a)}$$

$$A_i^{(1)} + \sum_j \omega_{ij}^2 = \gamma_{tt}\gamma^{-1} = a_1(t), \quad \text{(A4b)}$$

$$A_i^{(2)} = \sum_j x_j\left(\frac{d}{dt}\omega_{ij} + \sum_l \omega_{il}\omega_{li} + 2\gamma_t\gamma^{-1}\omega_{ij}\right). \quad \text{(A4c)}$$

We restrict our study to particular classes of solutions (A4). If $\mathbf{A}(\mathbf{x},t)=0$, $\omega_{ij}=0$, then $\omega_{ij}=\gamma_{tt}=0$, $\langle \mathbf{v}\rangle = \boldsymbol{\alpha} + \text{const}\,\mathbf{x}$ ($\gamma = t$ in the inhomogeneous case) and $\boldsymbol{\alpha}\gamma = \boldsymbol{\alpha}(0) + \int_0^t \mathbf{A}_0(t')\gamma(t')dt'$. We look for spatial forces $\mathbf{A}(x,t)\neq 0$ and seek the spatial dependence of $\langle \mathbf{v}\rangle$ given by (A4b) and (A4c). We remark that $\boldsymbol{\alpha}$ and $\mathbf{A}_0(t)$ appear only in (A4a). For simplicity in the discussion we determine both ω_{ij} and γ from (A4b) and (A4c) while we consider $\boldsymbol{\alpha}$ as arbitrary and (A4a) is the equation that defines $\mathbf{A}_0(t)$.

(i) For $\omega_{ij}=0$, the case $\mathbf{A}(x,t)=0$ as be seen above; for $\mathbf{A}\neq 0$ we find $\mathbf{A} = a_1(t)\mathbf{x}$, $\gamma_{tt} = a_1(t)\gamma$ and $\langle \mathbf{v}\rangle = \mathbf{x}\,\partial_t\log\gamma + \boldsymbol{\alpha}(t)$.

(ii) $\gamma^2\omega_{ij} =$ const or (iii) $\gamma^2\omega_{ij}(t)$, from arbitrary $a(t)$ [γ being solution of the rhs of (A4b)] and ω_{ij}, then (A4b) and (A4c) determine the components of $A_i^{(1)}A_i^{(2)}$ of the spatial dependence of the force.

We consider the case without force $\mathbf{A}_0 = \mathbf{A} = 0$, but now with $\omega_{ij}\neq 0$. In (A4a) either we choose $\boldsymbol{\alpha}=0$ or the equation determines $\boldsymbol{\alpha}$ from known A_{0i}.

From (A4b) and (A4c) we find

$$\sum_l \omega_{il}\omega_{lj} = 0, \quad i\neq j, \quad \sum_j \omega_{ij}^2 = \omega^2(r) \quad \text{independent of } i, \quad \text{(A5)}$$

$$\omega_{ij} = \omega_{ij}(0)\gamma^{-2}, \quad \gamma_{tt} = \gamma^{-3}\omega^2(0), \quad \gamma = \sqrt{\omega^2(0)t^2+1}. \quad \text{(A6)}$$

For $d=2$, (A5) is an identity, for $d=3$ the two relations (A5) are incompatible while for $d=4$ they are satisfied for $\omega_{23} = \varphi\omega_{14}$, $\omega_{23} = -\varphi\omega_{13}$, $\omega_{34} = \varphi\omega_{12}$, $\varphi^2 = 1$. Relations (A6) give $\gamma = \sqrt{\lambda(t+t_0)^2 + \omega^2(0)\lambda^{-1}}$, we choose the constants $t_0 = 0$ and $\lambda = \omega(0)$ such that $\gamma(0)=1$.

We come back to the determination of $\boldsymbol{\alpha}$ from $\mathbf{A}_0$, ω_{ij}, γ, in (A4a). This problem becomes cumbersome in the general d-case. The linear first-order system for the d functions α_i gives a differential equation of order d for α_i. For the most simple $d=2$ case, it is, writing $D = \partial_t + \partial_t(\log\gamma)$,

$$(D^2 + \omega_{12}^2 - (\partial_t\log\omega_{12})D)\alpha_i$$
$$= DA_{0i} + \omega_{ji}A_{0j} - A_{0i}\partial_t\log\omega_{12}), \quad i=1,2,$$

with $\gamma^2(\alpha_1^2+\alpha_2^2) =$ const, if $A_{0j}=0$, while it is a complicated third-order differential equation for $d=3$. Of course, when $\omega_{ij}\equiv 0$, the equations decouple and

$$\gamma\boldsymbol{\alpha} = \boldsymbol{\alpha}(0) + \int_0^t \mathbf{A}_0(t')\gamma(t')dt', \quad \gamma(0)=1. \quad \text{(A7)}$$

3. Direct solution of $\partial_t\tilde f = \gamma^{-d}\,\text{Col}\,\tilde f$

We start with the ansatz

$$(2\pi\Delta)^{d/2}\tilde f = (\exp(-\boldsymbol{\eta}^2/2\Delta))(\alpha_0 + (\boldsymbol{\eta}^2/2)\alpha_2)$$

substitute in both sides of the BE and find

$$(2\pi\Delta)^{d/2}e^{\boldsymbol{\eta}^2/2\Delta}\Bigg\{\Bigg[\alpha_{0,t} - \frac{d}{2}\alpha_0\frac{\Delta_t}{\Delta} + \frac{\boldsymbol{\eta}^2}{2}\left(\alpha_{2t} + \alpha_0\frac{\Delta_t}{\Delta^2} - \frac{d\alpha_2}{2}\frac{\Delta_t}{\Delta}\right) + \frac{\boldsymbol{\eta}^4}{4}\alpha_2\frac{\Delta_t}{\Delta^2} - \Bigg[\gamma^{-d}\sigma_2^{(d)}\alpha_2^2\Bigg(\frac{\boldsymbol{\eta}^4}{4} - \frac{\boldsymbol{\eta}^2}{2}(d+2)\Delta + d(d+2)\frac{\Delta^2}{4}\Bigg)\Bigg]\Bigg\} = 0. \quad \text{(A8)}$$

Equating the coefficients of like powers of $\boldsymbol{\eta}^{2p}$ we get

$$[\boldsymbol{\eta}^4/4]: \Delta_t - \Delta^2\gamma^{-d}\sigma_2^{(d)}\alpha_2 = 0,$$

$$[\boldsymbol{\eta}^0]: \alpha_{0,t} - \frac{d}{2}\alpha_0\frac{\Delta_t}{\Delta} - \alpha_2\Delta_t\frac{d(d+2)}{4} = 0, \quad \text{(A9)}$$

$$[\boldsymbol{\eta}^2/2]: \alpha_{2t} + \alpha_{0t}\frac{\Delta_t}{\Delta^2} + \alpha_2\frac{\Delta_t}{\Delta}\frac{(d+4)}{2} = 0.$$

By linear combinations of the two last relations we get

$$\left[\frac{\boldsymbol{\eta}^2}{2}\right]\Delta\left(\frac{d}{2}+1\right) + \Delta[\boldsymbol{\eta}^0] = \partial_t\left(\alpha_0\Delta + \frac{(d+2)\alpha_2\Delta^2}{2}\right) = 0$$

and

$$[\boldsymbol{\eta}^0] + \frac{d}{2}\Delta\left[\frac{\boldsymbol{\eta}^2}{2}\right] = \partial_t\left(\alpha_0 + \frac{d}{2}\Delta\alpha_2\right),$$

which express the properties that $\int \boldsymbol{\eta}^i\tilde f\,d\boldsymbol{\eta} =$ const, $i=0$ and 2. The $[\boldsymbol{\eta}^4/4]$ and these two pseudoconservation laws for $\tilde f$ were the three relations used in Sec. II C for the determination of the exact solution.

4. Extrema of $x = \sin t + \lambda \sin qt$, λ and q reals, $q>0$

The number of extrema with different x values gives the number of different Maxwellians. For $\lambda=0$, we only have two extrema for $t\in[2N\pi, (2N+2)\pi]$, a countable set for $t\in[0,\infty]$, but only two of them are different. By continuity the number of extrema, for t on a fixed interval, does not change for $|\lambda|$ sufficiently small while the number of different extrema can abruptly change. If λ is varying and we look continuously at the deformation of the curve $x(t)$ such that new extrema can appear or disappear we note that necessarily at least $\dot{x} = \ddot{x} = 0$,

$$\dot{x} = \ddot{x} = 0 \rightarrow \sin^2 t = q^2(1-\lambda^2 q^2)(q^2-1)^{-1} \in [0.1],$$

or necessarily $q^{-2} < |\lambda| < q^{-1}$ if $q > 1$ (and the converse inequality if $q < 1$). Consequently if $q > 1$ and $|\lambda| < q^{-2}$ or $|\lambda| > q^{-1}$, x on a finite interval has the same number of extrema (and the converse result if $q < 1$). Of course this result does not hold for the number of different extrema, which can change suddenly.

Let us illustrate with a few examples for q integer > 1 for which it is sufficient to take $t \in [0,2\pi]$. We remark that for q even, $\cos qt$ is an even $\cos t$ polynomial; $\dot{x} = 0$ or $\lambda^{-1} \cos t + q \cos qt = 0$ gives the invariance $(\cos t, \lambda) \leftrightarrow (-\cos t, -\lambda)$. Consequently for q even it is sufficient to study $\lambda > 0$.

(i) For $q = 2$, for $|\lambda| \leqslant 0.5$, x has two extrema, four for $|\lambda| > 0.5$. All are different. For $\lambda = 0.5$ (-0.5) at $t = \pi$ (0) we find $\dot{x} = \ddot{x} = 0$, $\dddot{x} \neq 0$.

(ii) For $q = 3$, for $-\frac{1}{3} < \lambda < \frac{1}{3}$, x has two extrema and six for x outside this interval, however, only four have different x values, except three for $\lambda = 1$. For $\lambda = -\frac{1}{3}\,\frac{1}{3}$ at $t = \pi$ $(\pi/2)$ we find $\dot{x} = \ddot{x} = 0$, $\dddot{x} \neq 0$ $(=0)$, $\dddot{x} = 0$ $(\neq 0)$.

(iii) For $q = 4$, for $|\lambda| < 0.17$, x has two different extrema. For $\lambda \# 0.17$, $t \# 0.27\pi$, $\dot{x} = \ddot{x} = 0$, $\dddot{x} \neq 0$ and for $0.17 < |\lambda| < 1/4$ we have six extrema that are different. For $\lambda = \frac{1}{4}$, $t = \pi$, $\dot{x} = \ddot{x} = 0$, $\dddot{x} \neq 0$ and for $|\lambda| > \frac{1}{4}$ we have eight different extrema except for $\lambda \# 0.92$ with only seven.

Let us choose now $q = Q^{-1} < 1$, Q integer > 1, $x(t)$ is periodic with period $2\pi Q$. For $t \in [0,2\pi Q]$ and $\lambda = 0$, x has $2Q$ maxima but only two of them are different. Even if λ is very small and no new extrema appear, the number of different maxima will change suddenly when $\lambda \neq 0$. The results can be understood from the previous one if we write $x/\lambda = \sin \tau + \lambda^{-1} \sin \tau Q$ with $\tau = tQ^{-1}$. Here λ around 0 is equivalent to λ around infinity above.

(i) For $Q = 2$ and $t \in [0,4\pi]$, for $\lambda = 0$, x has four extrema but only two are different. For $|\lambda| < 2$, x has four extrema that are different and only two for $|\lambda| > 2$.

(ii) For $Q = 3$ and $t \in [0,6\pi]$, for $\lambda = 0$, x has six extrema and two different; for $0 < \lambda < 9$, x has six extrema and four different except three for $\lambda = 1$; for $\lambda > 9$, x has two extrema and they are different, and so on. It is an easy exercise to study other rational q values. What about q not rational?

We want to show that the extrema of x belong to a countable set of values. From $\dot{x} = 0$, we can either rewrite $\sin t$ or $\sin qt$, substitute into x and note that the extrema of x can be rewritten as either x_1 or x_2:

$$x_1 = \sin t + \eta_1 \sqrt{1 - \cos^2 t / q^2 \lambda^2},$$

$$x_2 = \lambda \sin qt + \eta_2 \sqrt{1 - q^2 \lambda^2 \cos^2 qt}, \quad \eta_i^2 = 1,$$

where x_1 is periodic with period $T_1 = 2\pi$ and x_2 with $T_2 = 2\pi/q$. The extrema of x belong to $x_1 \cap x_2$. Let us consider an interval $t \in [2N\pi, 2(N+1)\pi]$, $N = 0,1,\ldots$. For $t \in [2N\pi, (2N+1)\pi]$, $\sin t > 0$ and $\eta_2 = +1$, for $t \in [(2N+1)\pi, 2(N+1)\pi]$, $\sin t < 0$ and $\eta_2 = -$. From analyticity around the t axis, x_1 and x_2 cannot be identical and $x_1 \cap x_2$ has a finite number of common values for t belonging to an interval with N fixed. For $t \in [0, \infty]$ it follows that they have a countable set of common values that include the extrema of x.

APPENDIX B

1. Equations for $LF_{as} = 0$, $d > 2$, $L = \partial_t + \mathbf{v} \cdot \partial_{\mathbf{x}} + \mathbf{A}_0(t) \cdot \partial_{\mathbf{v}} + a_1(t) \partial_{\mathbf{v}} \cdot \mathbf{v} + a_2(t)$

We define the same f_{as} in Appendix A, $LF_{as}(\boldsymbol{\eta}^2, t) = 0$ leading to

$$(\partial_t + \mathbf{v} \cdot \partial_{\mathbf{x}} + a_1 + a_2 d)\, \nu = (L - a_1 d - a_2) \boldsymbol{\eta}^2/2,$$

$$\boldsymbol{\eta} = \gamma(\mathbf{v} - \langle \mathbf{v} \rangle), \quad \nu = \nu(\mathbf{x}, t), \tag{B1}$$

and we follow a study similar to Appendix A. At the rhs the powers $\mathbf{v}^2 v_i$ lead to $\gamma = \gamma(t)$ while v_i^2, $v_i\, v_j$ give the relations (A2) and (A2′) except that we add a_1 to $\gamma_t\, \gamma^{-1}$:

$$\langle \mathbf{v} \rangle_i = \alpha_i(t) + x_i(a_1 + \gamma_t\, \gamma^{-1}) + \sum_{l \neq i} \omega_{il}(t) x_l, \tag{B2}$$

ω_{ij} still being an antisymmetric tensor. In $(L - a_1 d - a_2)\boldsymbol{\eta}^2$ still remains only η_i terms, we find (A3) except that $\mathbf{A}$ and A_i are absent:

$$\frac{\gamma^{-1}(L - d a_1 - a_2)\boldsymbol{\eta}^2}{2} = \sum \eta_i\, \zeta_i(\mathbf{x}, t),$$

$$\zeta_i = A_{0i} - \left(\partial_t + \frac{\gamma_t}{\gamma}\right) \langle \mathbf{v} \rangle_i - \sum_j \langle \mathbf{v} \rangle_j\, \omega_{ij}. \tag{B3}$$

Finally (B1) is rewritten as

$$\gamma^{-2} \partial_{x_i} \log \nu = \zeta_i, \quad (\partial_t + \langle \mathbf{v} \rangle \cdot \partial_{\mathbf{x}} + a_1 d + a_2)\, \nu = 0. \tag{B1′}$$

2. $(L - da_1 - a_2)\boldsymbol{\eta}^2 = 0$ or $\zeta_i = 0$ in (B3)

First from (B1′), $\zeta_i = 0$, we find

$$\nu = \nu(t_0) \exp\left(- \int_{t_0}^{t} (d a_1(t') + a_2(t')) dt' \right),$$

second we determine the compatible γ, $\langle \mathbf{v} \rangle$, and $a_1(t)$. We substitute $\zeta_i = 0$ into (B2) and the coefficients of like x_i powers are zero:

$$A_{0i} = (\gamma_t\, \gamma^{-1} + \partial_t)\alpha_i - \sum \alpha_j\, \omega_{ij}, \tag{B4a}$$

$$(\partial_t + \gamma_t\, \gamma^{-1}) a_1 + \gamma_{tt}\, \gamma^{-1} + \sum_j \omega_{ij}^2 = 0, \tag{B4b}$$

$$0 = \sum_l \omega_{il}\, \omega_{lj}, \quad (\partial_t + a_1 + 2\gamma_t/\gamma)\, \omega_{ij} = 0, \quad j \neq i. \tag{B4c}$$

Two cases occur, depending whether the tensor is zero or not.

(i) For $\omega_{ij} = 0$, we find $\langle \mathbf{v} \rangle = \boldsymbol{\alpha} + \mu_0 \gamma^{-1} \mathbf{x}$ and $\gamma_t + a_1 \gamma = \mu_0$ (constant), $(\partial_t + \gamma_t\, \gamma^{-1})\boldsymbol{\alpha} = \mathbf{A}_0$, which are easily integrated.

(ii) For $\omega_{ij} \neq 0$, from (B4b) and (B4c) we determine both ω_{ij}, γ; for simplicity we choose either $\mathbf{A}_0(t) = \boldsymbol{\alpha}(t) = 0$ or $\alpha_i(t)$ arbitrary and look at (B4a) as the equations giving $A_{0i}(t)$. From (B4b) and (B4c) we find

$$\sum_l \omega_{il}\,\omega_{lj} = 0, \quad i \neq j, \quad \sum_j \omega_{ij}^2 = \omega^2, \quad i \text{ independent},$$

$$\omega_{ij} = \omega_{ij}(0)\exp\left(-\int_0^t a_1\,dt'\right)\gamma^{-2}, \tag{B5}$$

$$\partial_t(\gamma a_1) + \gamma_{tt} - \gamma^{-3}\omega^2(0)\exp\left(-2\int_0^t a_1\,dt'\right) = 0.$$

For $d = 2$, the two first relations are automatically satisfied, they are incompatible for $d = 3$ while they are satisfied for $d = 4$ if $\omega_{23} = \varphi\omega_{14}$, $\omega_{24} = -\varphi\omega_{13}$, $\omega_{34} = \varphi\omega_{12}$, $\varphi^2 = 1$. Finally the last relation is a nonlinear differential equation for γ that depends on the function $a_1(t)$.

[1] H. Cornille, Saclay preprint PhT 85-116, presented at the "2nd International Workshop MAFPD."

[2] L. Boltzmann, *Wissenschaptlich Abhandlungen*, edited by F. Hasenorl (Barth, Leipzig, 1909), Vol. II, p. 83.

[3] C. Cercignani, *Theory and Application of the Boltzmann Equation* (Scottish Academic, Edinburgh, 1975), pp. 142–149.

[4] V. Bobylev, Sov. Phys. Dokl. **20**, 823 (1976); Teor. Mat. Phys. **8**, 279 (1984).

[5] M. Krook and T. T. Wu, Phys. Rev. Lett. **16**, 1107 (1907).

[6] M. Ernst, Phys. Lett. A **61**, 320 (1979); Phys. Rep. **78**, 1 (1981).

[7] H. Cornille and A. Gervois, J. Stat. Phys. **23**, 167 (1980).

[8] R. C. Muncaster, Arch. Rat. Mech. Anal. **70**, 115 (1979).

[9] A. A. Nikolskii, Sov. Phys. Dokl. **8**, 633 (1964).

[10] G. TenTi and W. H. Hui, J. Math. Phys. **19**, 774 (1979).

[11] M. Kac, *Proceedings of the 3rd Berkeley Symposium on Mathematics, Statistics, and Probability*, Vol. 3 (Univ. California, Berkeley, 1954), p. 111.

[12] J. A. Tjon and T. T. Wu, Phys. Rev. A **19**, 885 (1979).

[13] H. Cornille, J. Phys. A Lett. **234**, 2355 (1984).

[14] H. Cornille, J. Math. Phys. **26**, 1203 (1985); J. Phys. A **18**, 1209 (1985).

[15] H. Cornille, J. Phys. A Lett. **18**, L839 (1985).

[16] J. A. Tjon, Phys. Lett. A **70**, 369 (1979).

[17] H. Cornille, Saclay P.H.T. /86/33.

Theorems pertaining to Fokker–Planck statistical equilibrium for multidimensional stochastic systems

Gerald Rosen
Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104

(Received 24 October 1985; accepted for publication 30 December 1985)

It is shown that a Fokker–Planck equation $\partial P/\partial t = -\Sigma_{i=1}^{n}\partial[Q_i(q)P]/\partial q_i + \frac{1}{2}\Sigma_{i,j=1}^{n}\sigma_{ij} \times\partial^2 P/\partial q_i\,\partial q_j$ with $n\geqslant 3$ may admit an asymptotic steady-state solution $P\rightarrow P_{\rm eq}(q)$ that is independent of t only if the necessary condition $\min\{2(1-2/n)[\int(Q\cdot\sigma^{-1}\cdot Q)^{n/2}\,d^n q]^{2/n}, [\int_\Omega |S|^{n/2}\,d^n q]^{2/n}\} > 2^{-1+2/n}\pi^{1+1/n}[\Gamma((n+1)/2)]^{-2/n}n^2[\det(\sigma_{ij})]^{1/n}$ is satisfied, where $S=S(q)\equiv\Sigma_{i=1}^{n}\partial Q_i(q)/\partial q_i$ and the second integral is over the q-space region $\Omega\equiv\{q\in R_n$ such that $S(q)<0\}$. In addition, it is demonstrated that the probability distribution $P=P(q,t)$ is localized in q-space as $t\rightarrow\infty$ if S is bounded from above by a negative constant.

I. INTRODUCTION

The Fokker–Planck equation is often employed to describe the time evolution of nonequilibrium systems in physics, chemistry, biology, and the engineering sciences.[1] Many nonequilibrium and steady-state equilibrium phenomena in classical statistical mechanics are amenable to quantitative description via modeling with the Fokker–Planck equation. Such phenomena include those involving Brownian motion,[2] chemical reactions,[3] and laser beam interactions.[4]

Consider a generic system of essentially coupled nonlinear Langevin equations

$$\frac{dq_i}{dt} = Q_i(q) + r_i(t), \quad i=1,\ldots,n, \tag{1}$$

where the $Q_i(q)$ are continuous, smoothly differentiable functions of the Cartesian rectilinear coordinates $q=(q_1(t),\ldots,q_n(t))$ and the $r_i(t)$ are Gaussian stochastic variables with zero mean values and δ-function covariances:

$$\langle r_i(t)\rangle = 0, \quad \langle r_i(t')r_j(t'')\rangle = \sigma_{ij}\delta(t'-t''),$$
$$(\sigma_{ij})\equiv(\sigma_{ji})\equiv(\text{constant positive-definite array}). \tag{2}$$

As observed many years ago, the probability density $P=P(q,t)$ for the q_i's at time t is governed by the associated Fokker–Planck equation[5]

$$\frac{\partial P}{\partial t} = -\sum_{i=1}^{n}\frac{\partial}{\partial q_i}[Q_i(q)P] + \frac{1}{2}\sum_{i,j=1}^{n}\sigma_{ij}\frac{\partial^2 P}{\partial q_i\,\partial q_j}. \tag{3}$$

A solution to Eq. (3) is required to be non-negative and normalizable in R_n [i.e., P must be a function in the space $L_+^1(R_n)$] for interpretation as a probability density: $0\leqslant P(q,t)$, $\int P(q,t)d^n q = 1$. Statistical equilibrium may obtain asymptotically with increasing t if the steady-state form of (3)

$$\sum_{i=1}^{n}\frac{\partial}{\partial q_i}([Q_i(q)P_{\rm eq}] = \frac{1}{2}\sum_{i,j=1}^{n}\sigma_{ij}\frac{\partial^2 P_{\rm eq}}{\partial q_i\,\partial q_j} \tag{4}$$

admits a non-negative solution $P_{\rm eq} = P_{\rm eq}(q)$ that is normalizable in R_n. From the general theory for linear elliptic partial differential equations with variable coefficients,[6] it follows that the existence of a non-negative normalizable $P_{\rm eq}$ hinges on the specific form and constant parameter values in the functions $Q_i(q)$ in (4).[7] For example, consider the linear expressions

$$Q_i(q) = k(q_i - a_i), \tag{5}$$

in which k and $a_1,\ldots,a_n$ are constant parameters. Subject to the (Green's function) initial condition

$$P(q,0) = \prod_{i=1}^{n}\delta(q_i - q_i'),$$
$$q_i'\equiv\text{const}, \quad \text{for } i=1,\ldots,n, \tag{6}$$

the solution to (3) with (5) is obtainable as a straightforward extension of the one-dimensional case.[8] One finds

$$P(q,t) = [k/\pi\bar\sigma(e^{2kt}-1)]^{n/2} \times\exp[-k(e^{2kt}-1)^{-1}\xi\cdot\sigma^{-1}\cdot\xi], \tag{7}$$

where

$$\bar\sigma\equiv[\det(\sigma_{ij})]^{1/n}, \tag{8}$$

$$\xi_i\equiv q_i - a_i - (q_i' - a_i)e^{kt}, \tag{9}$$

$$\sum_{j=1}^{n}(\sigma^{-1})_{ij}\sigma_{jl}\equiv\delta_{il}, \tag{10}$$

and the dots in $\xi\cdot\sigma^{-1}\cdot\xi$ denote n-tuple index contraction. Depending critically on the sign of k in (5) or equivalently on the sign of

$$S = S(q) = \sum_{i=1}^{n}\frac{\partial Q_i(q)}{\partial q_i}, \tag{11}$$

the solution (7) is such that

$$\lim_{t\rightarrow\infty}\left[\max_q P(q,t)\right] = 0, \quad \text{for } S=nk\geqslant 0,$$

$$\lim_{t\rightarrow\infty} P(q,t) = P_{\rm eq}(q) = (|k|/\pi\bar\sigma)^{n/2}\exp[-|k|(q-a)\cdot\sigma^{-1}\cdot(q-a)],$$
$$\text{for } S=nk<0, \tag{12}$$

and hence steady-state equilibrium is attained for large t if and only if $S<0$. That the magnitudes $|Q_i(q)|$ must be sufficiently large and S defined by (11) must be sufficiently negative for existence of a non-negative normalizable P_{eq} satisfying (4) is shown for general $Q_i(q)$ and $n \geqslant 3$ by the theorem proved in the following section.

II. NECESSARY CONDITION FOR EXISTENCE OF A STEADY-STATE SOLUTION

Theorem 1: A necessary condition for existence of a non-negative normalizable solution to the steady-state Fokker–Planck equation (4) is[9]

$$\min\left\{(1-2n^{-1})^2\left[\int (Q\cdot\sigma^{-1}\cdot Q)^{n/2}\,d^nq\right]^{2/n},\right.$$
$$\left.\frac{1}{2}(1-2n^{-1})\left[\int_\Omega |S|^{n/2}\,d^nq\right]^{2/n}\right\} > c_n\bar{\sigma}, \quad (13)$$

where the first integral is over all R_n, the second integral is over $\Omega \equiv \{q\in R_n$ such that $S(q)<0\}$, and the constants c_n, $\bar{\sigma}$ are defined by (A2) and (8).

Proof: Multiplication of (4) by $P_{eq}^{-2/n}$ and integration by parts over R_n yields

$$2\int P_{eq}^{-2/n}\left(Q\cdot\frac{\partial P_{eq}}{\partial q}\right)d^nq = I, \quad (14)$$

where the integral I on the right side is defined and bounded from below by the Appendix relation (A7). By expressing $Q\cdot\partial P_{eq}/\partial q = Q\cdot\sigma^{-1/2}\cdot\sigma^{1/2}\cdot\partial P_{eq}/\partial q$ and applying the Schwarz inequality to the left side of (14), one obtains

$$\int P_{eq}^{-2/n}\left(Q\cdot\frac{\partial P_{eq}}{\partial q}\right)d^nq$$
$$\leqslant\left[\int (Q\cdot\sigma^{-1}\cdot Q)P_{eq}^{1-2/n}\,d^nq\right]^{1/2} I^{1/2}$$
$$\leqslant\left[\int (Q\cdot\sigma^{-1}\cdot Q)^{n/2}\,d^nq\right]^{1/n} I^{1/2}, \quad (15)$$

in which use is made of the Appendix definition (A7) and inequality (A8) with $F = -Q\cdot\sigma^{-1}\cdot Q$, $\Lambda = R_n$. Hence, by combining (15) and (14), dividing both sides of the resulting inequality by $I^{1/2}$, and employing the inequality (A7), one finds the first part of (13), viz.

$$\left[\int (Q\cdot\sigma^{-1}\cdot Q)^{n/2}\,d^nq\right]^{2/n} > n^2(n-2)^{-2}c_n\bar{\sigma}. \quad (16)$$

On the other hand, by performing an additional integration by parts on the left side of (14), introducing the definition (11), and making use of (A8), one gets

$$\int P_{eq}^{-2/n}\left(Q\cdot\frac{\partial P_{eq}}{\partial q}\right)d^nq$$
$$= -(1-2n^{-1})^{-1}\int S P_{eq}^{1-2/n}\,d^nq$$
$$\leqslant (1-2n^{-1})^{-1}\left(\int_\Omega |S|^{n/2}\,d^nq\right)^{2/n}, \quad (17)$$

and therefore the second part of (13),

$$\left[\int_\Omega |S|^{n/2}\,d^nq\right]^{2/n} > 2n(n-2)^{-1}c_n\bar{\sigma}, \quad (18)$$

follows by combining (14), (17), and (A7). ■

Remark 1: In the case of the linear expressions (5), the integral on the left side of (16) is divergent for $k\neq 0$, and therefore (13) reduces to the necessary condition (18). The latter inequality is satisfied *a fortiori* for $S = nk < 0$ but it is not satisfied for $S \geqslant 0$ (with Ω the null set); thus condition (13) or (18) is actually necessary and sufficient for steady-state statistical equilibrium in the case of (5).

Remark 2: Although general as a necessary condition, in no sense is (13) sufficient to guarantee a non-negative normalizable solution to (4) for arbitrarily prescribed $Q_i(q)$. As an example, consider $Q_i(q) = kaq_i \times(\alpha^2 + |q|^{2n})^{-1/2}$, with k and α constant parameters; although (13) is satisfied if k is negative and the dimensionless grouping $(|k|\alpha^{2/n}\bar{\sigma}^{-1})$ is sufficiently large, it is easily shown by asymptotic analysis for large $|q|$ that (4) does not possess an admissible solution $P_{eq}(q)$ for such $Q_i(q)$. It is noteworthy that Theorem 2 below does *not* apply to this set of $Q_i(q)$, since the associated quantity (11), $S(q) = kn\alpha^3(\alpha^2 + |q|^{2n})^{-3/2}$, tends to zero as $|q|\to\infty$.

In order to satisfy the necessary condition (13) for steady-state statistical equilibrium, the $Q_i(q)$ must be such that both integrals in (13) are sufficiently large in magnitude. From (A2) it follows that the prefactor constant on the right side of (13) has the values

$$c_3 = 5.4779,\quad c_5 = 14.8119,$$
$$c_4 = 10.2604,\quad c_6 = 19.2595, \quad (19)$$

while the asymptotic Stirling formula for the gamma function implies that

$$c_n > 4.2699(n-2),\quad \text{for all } n\geqslant 3,$$
$$\lim_{n\to\infty}(c_n/n) = 4.2699. \quad (20)$$

In view of (20), the condition (13) is likely to be more stringent for large n. If (13) is *not* satisfied by $Q_i(q)$ in (3) and (4), then time-independent statistical equilibrium is unattainable for the stochastic dynamical system.

III. SUBSIDIARY RESULTS

Further general insight regarding the time evolution prescribed by (3) is achieved by considering the functional

$$\mathscr{F} = \mathscr{F}(t) \equiv \int P^2\,d^nq. \quad (21)$$

The existence of (21) as a finite quantity for all $t>0$ is usually guaranteed by the dynamical equation that follows from (21) and (3),

$$\frac{d\mathscr{F}}{dt} = -\int SP^2\,d^nq - \mathscr{G}, \quad (22)$$

where integration by parts has been performed and the quantities defined by (11) and (A3) appear. For example, if $S = S(q)$ is bounded *from below* by a negative constant $-b$,

$$S(q) \geqslant -b,\quad \text{for all } q, \quad (23)$$

then (22), (23), and (A6) imply the differential inequality

$$\frac{d\mathscr{F}}{dt} < (b\mathscr{F} - c_n\bar{\sigma}\mathscr{F}^{1+2/n}), \quad (24)$$

which yields the integral upper bound

$$\mathscr{F}(t)\leqslant[(\mathscr{F}(0)^{-2/n}-b^{-1}c_n\bar{\sigma})e^{-2bt/n}+b^{-1}c_n\bar{\sigma}]^{-n/2}, \tag{25}$$

for all $t\geqslant 0$. Hence, if $\mathscr{F}(0)$ is finite [i.e., $P(q,0)$ is contained in the function space $L^2(R_n)\cup L^1_+(R_n)$], (3) and (23) imply that $\mathscr{F}(t)$ is finite and bounded according to (25) for all $t\geqslant 0$ [i.e., $P(q,t)$ is contained in $L^2(R_n)\cup L^1_+(R_n)$]. Both $\mathscr{F}$ and $\mathscr{G}$ defined by (21) and (A3) approach finite positive constant values with increasing t if the solution to (3) tends to a steady-state equilibrium, $P\rightarrow P_{\rm eq}(q)$. In particular for the solution (7), the induced time dependence of $\mathscr{F}$ and $\mathscr{G}$ for large t is shown by explicit evaluation of the integrals (21) and (A3),

$$\mathscr{F}\doteq\begin{cases}(k/2\pi\bar{\sigma})^{n/2}e^{-nkt}, & \text{for } k>0,\\ (4\pi\bar{\sigma}t)^{-n/2}, & \text{for } k=0,\\ (|k|/2\pi\bar{\sigma})^{n/2}(1+\frac{1}{2}ne^{-2|k|t}), & \text{for } k<0;\end{cases} \tag{26}$$

$$\mathscr{G}\doteq\begin{cases}(k/2\pi\bar{\sigma})^{n/2}kne^{-(n+2)kt}, & \text{for } k>0,\\ (4\pi\bar{\sigma}t)^{-n/2}(n/2t), & \text{for } k=0,\\ (|k|/2\pi\bar{\sigma})^{n/2}n|k|[1+(\frac{1}{2}n+1)e^{-2|k|t}], & \\ \quad\text{for } k<0,\end{cases} \tag{27}$$

with $\mathscr{F}$ and $\mathscr{G}$ approaching finite positive values only if $k<0$.

A measure of the q-space extension or breadth of a general probability distribution $P(q,t)$ is given the quantity

$$\lambda=\lambda(t)\equiv(\bar{\sigma}\mathscr{F}/\mathscr{G})^{1/2}. \tag{28}$$

In the case of the solution (7) the characteristic q-space extension (28) has the asymptotic forms for large t indicated by (26) and (27), viz.

$$\lambda\doteq\begin{cases}(\bar{\sigma}/nk)^{1/2}e^{kt}, & \text{for } k>0,\\ (2\bar{\sigma}t/n)^{1/2}, & \text{for } k=0,\\ (\bar{\sigma}/n|k|)^{1/2}(1-\frac{1}{4}e^{-2|k|t}), & \text{for } k<0.\end{cases} \tag{29}$$

By definition, a probability distribution $P(q,t)$ is "localized in q-space for large t" if λ defined by (28) is bounded from above by a finite positive constant as t tends to infinity: $\lim_{t\rightarrow\infty}\lambda(t)<\infty$. Clearly, localization in q-space for large t is necessary (but not sufficient) for steady-state statistical equilibrium. Conversely, if the quantity (28) increases without bound as t increases, a steady-state equilibrium is precluded. Thus, for example, in the case of the linear expressions (5) the associated P are localized for large t if and only if $S=nk$ is negative, as shown by (29). More generally for arbitrary $Q_i(q)$ one has the following theorem.

Theorem 2: A non-negative normalizable solution to the Fokker–Planck equation (3) is localized in q-space for large t [in the sense that $\lim_{t\rightarrow\infty}\lambda(t)<\infty$] if S is bounded *from above* by a negative constant:

$$S(q)\leqslant -a, \quad \text{for all } q. \tag{30}$$

Proof: In cases for which the $Q_i(q)$ and definition (11) admit a bound of the form (30), the dynamical equation (22) yields the differential inequality

$$\frac{d\mathscr{F}}{dt}\geqslant a\mathscr{F}-\bar{\sigma}\mathscr{F}\lambda^{-2}, \tag{31}$$

where $\mathscr{G}$ has been eliminated by employing the definition (28). The integral of (31),

$$\mathscr{F}(t)\geqslant\mathscr{F}(0)\exp\left(at-\bar{\sigma}\int_0^t\lambda(t')^{-2}\,dt'\right) \tag{32}$$

can be combined with the inequality derived from (28) and (A6),

$$\lambda(t)^{-2}>c_n\mathscr{F}(t)^{2/n}, \tag{33}$$

to produce

$$\lambda(t)^{-2}>c_n\mathscr{F}(0)^{2/n}\times\exp\left[2n^{-1}\left(at-\bar{\sigma}\int_0^t\lambda(t')^{-2}\,dt'\right)\right]. \tag{34}$$

By observing that (34) is expressed equivalently as

$$\frac{1}{2}n\bar{\sigma}^{-1}\frac{d}{dt}\left[\exp\left(2n^{-1}\bar{\sigma}\int_0^t\lambda(t')^{-2}\,dt'\right)\right]>c_n\mathscr{F}(0)^{2/n}e^{2at/n}, \tag{35}$$

with the integral

$$2n^{-1}\bar{\sigma}\int_0^t\lambda(t')^{-2}\,dt'\geqslant\ln\left[1+c_n\bar{\sigma}\mathscr{F}(0)^{2/n}a^{-1}(e^{2at/n}-1)\right], \tag{36}$$

one obtains the asymptotic bound

$$\lim_{t\rightarrow\infty}\left[\frac{1}{t}\int_0^t\lambda(t')^{-2}\,dt'\right]\geqslant(a/\bar{\sigma}), \tag{37}$$

which obviously requires

$$\lim_{t\rightarrow\infty}\lambda(t)\leqslant(\bar{\sigma}/a)^{1/2}. \tag{38}$$

IV. CONCLUDING REMARKS

Limiting the characteristic q-space extension of a general probability distribution $P(q,t)$ for large t if (30) is satisfied, the constant bound on the right side of (38) is consonant with the steady-state equilibrium form shown in (12), an example with $a=n|k|$. More generally, since the $Q_i(q)$ in (3) may foster limit-cycle or other essentially time-dependent changes in $P(q,t)$ for arbitrarily large values of t, it is clear that sufficient conditions on the $Q_i(q)$ for attainment of steady-state equilibrium must involve more than just the quantity (11). Nevertheless, Theorem 2 shows that a negative constant upper bound on S of the form (30) is sufficient to guarantee a $P(q,t)$ that is localized in q-space for large t and thus a candidate for steady-state equilibrium. With broad applicability to stochastic dynamical phenomena in physics, chemistry, and biology, the theorems reported here demonstrate that basic properties of the $Q_i(q)$ and the associated quantity (11) act to either preclude or promote the attainment of steady-state statistical equilibrium for Fokker–Planck stochastic systems with three or more q_i's.

APPENDIX: LOWER BOUND ON THE INTEGRAL I

Consider the Sobolev inequality[10]

$$\sum_{i=1}^{n}\int\left(\frac{\partial f}{\partial x_i}\right)^2 d^nx\geqslant c_n\left[\int|f|^{2n/(n-2)}\,d^nx\right]^{(n-2)/n}, \tag{A1}$$

$$c_n \equiv 2^{-2+2/n} \pi^{1+1/n}(n^2-2n)\left[\Gamma\left(\frac{n+1}{2}\right)\right]^{-2/n}, \tag{A2}$$

valid for all $n > 3$ with $f = f(x)$ any real continuous function of $x = (x_1,...,x_n)$ such that its first derivatives $\partial f/\partial x_i$ are at least piecewise continuous and the integral over R_n on the right side of (A1) is a finite quantity. By putting $f = P(q)$ with $q_i \equiv \Sigma_{k=1}^n (\sigma^{1/2})_{ik} x_k$ in (A1) and defining $(\sigma^{1/2})_{ik}$ as a real symmetric array that satisfies

$$\sum_{k=1}^n (\sigma^{1/2})_{ik} (\sigma^{1/2})_{jk} = \sigma_{ij},$$

one obtains

$$\mathscr{G} = \mathscr{G}(t) \equiv \sum_{i,j=1}^n \sigma_{ij} \int \frac{\partial P}{\partial q_i} \frac{\partial P}{\partial q_j} d^n q > c_n \bar{\sigma} \left[\int P^{2n/(n-2)} d^n q\right]^{(n-2)/n}, \tag{A3}$$

where $\bar{\sigma} = (\det \sigma_{ij})^{1/n}$ is the geometric mean of the eigenvalues of the positive-definite dispersion matrix (σ_{ij}). Strict inequality is indicated in (A3) because the special case of equality in (A1) is ruled out[10] for P in the function space $L^1_+(R_n)$. In view of the Hölder inequality

$$\mathscr{F} = \mathscr{F}(t) \equiv \int P^2 d^n q < \left(\int P d^n q\right)^{4/(n+2)} \left(\int P^{2n/(n-2)} d^n q\right)^{(n-2)/(n+2)} \tag{A4}$$

and the normalization condition

$$\int P d^n q = 1, \tag{A5}$$

(A3) implies that

$$\mathscr{G} > c_n \bar{\sigma} \mathscr{F}^{1+2/n}. \tag{A6}$$

Also observe that by putting $f = P_{\text{eq}}^{(n-2)/2n}$ in (A1) and making use of (A5), one obtains

$$I \equiv \sum_{i,j=1}^n \sigma_{ij} \int P_{\text{eq}}^{-(1+2/n)} \frac{\partial P_{\text{eq}}}{\partial q_i} \frac{\partial P_{\text{eq}}}{\partial q_j} d^n q > 4n^2(n-2)^{-2} c_n \bar{\sigma}. \tag{A7}$$

Finally observe that for any aribitrary function $F(q)$ that is non-positive through the region $\Lambda \equiv \{q \in R_n$ such that $F(q) \leqslant 0\}$, one has

$$-\int F P_{\text{eq}}^{1-2/n} d^n q \leqslant \left(\int_\Lambda |F|^{n/2} d^n q\right)^{2/n} \left(\int_\Lambda P_{\text{eq}} d^n q\right)^{(n-2)/n} \leqslant \left(\int_\Lambda |F|^{n/2} d^n q\right)^{2/n}, \tag{A8}$$

by virtue of the indicated Hölder inequality and the normalization condition (A5).

[1] H. Haken, Rev. Mod. Phys. **47**, 67 (1975); N. G. van Kampen, *Stochastic Processes in Chemistry and Physics* (North-Holland, Amsterdam, 1981); H. Risken, *The Fokker–Planck Equation* (Springer, New York, 1984).

[2] N. G. van Kampen, J. Stat. Phys **17**, 71 (1977); M. Suzuki, Adv. Chem. Phys. **46**, 195 (1981); M. Wehner and W. G. Wolfer, Phys. Rev. A **27**, 2663 (1983); **28**, 3003 (1983); P. S. Hubbard, Phys. Rev. A **19**, 907 (1979).

[3] E. Marechal and M. Moreau, Mol. Phys. **51**, 133 (1984); J. A. Montgomery, Jr., D. Chandler, and B. Berne, J. Chem. Phys. **70**, 4056 (1979); R. S. Larson and M. Kostin, J. Chem. Phys. **69**, 4821 (1978); **72**, 1392 (1980); B. Carmeli and A. Nitzan, J. Chem. Phys. **80**, 3596 (1984).

[4] C. M. Bowden, M. Ciften, and H. R. Roble, *Optical Bistability* (Plenum, New York, 1981); R. Bonifacio, *Dissipative Systems in Quantum Optics* (Springer, Berlin, 1982); R. Graham and A. Schenzle, Phys. Rev. A **23**, 1302 (1981); H. Brand, A. Schenzle, and G. Schroder, *ibid.* **25**, 2324 (1982).

[5] H. A. Kramers, Physica **7**, 284 (1940); M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945); N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, New York, 1981); Z. Schuss, *Theory and Applications of Stochastic Differential Equations* (Wiley–Interscience, New York, 1980).

[6] For example, L. Hörmander, *Linear Partial Differential Operators* (Academic, New York, 1964), pp. 242–274.

[7] Stochastic models for condensed-phase chemical reactions are of considerable current interest [M. Schell and R. Kapral, Chem. Phys. Lett. **81**, 83 (1981); J. T. Hynes, R. Kapral, and G. M. Torrie, J. Chem. Phys. **72**, 177 (1980); M. Sitarski, Int. J. Chem. Kinetics **13**, 125 (1981); K. L. Ngai and F.-S. Liu, Phys. Rev. B **24**, 1049 (1981)] and have been formulated in terms of the Langevin velocity equation and the associated six-dimensional Fokker–Planck equation with spatial inhomogeneity (independent variables $\mathbf{v}, \mathbf{x}, t$). Here, as with other Fokker–Planck equations of contemporary importance,[1–4] one must deal with nonlinear $Q_i(q)$ and $n > 3$. The attainment of steady-state equilibrium for such systems, as represented by an asymptotically time-independent solution to the Fokker–Planck equation (3), is by no means automatically guaranteed. Indeed, the existence of perpetuating time-dependent reactions, such as discovered by A. M. Zhabotinski [Biofiz. **2**, 306 (1964); Russ. J. Phys. Chem. **42**, 1649 (1968)] underscores the importance attached to the general question of existence of steady-state solutions to this class of Fokker–Planck equations.

[8] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **30**, 823 (1930).

[9] This necessary condition for steady-state equilibrium is sharper than the author's earlier result [G. Rosen, Phys. Rev. A **30**, 3359 (1984)].

[10] E. H. Lieb, Ann. Math. **118**, 349 (1983); P. L. Lions, C. R. Acad. Sci. **296**, 645 (1983); G. Talenti, Ann. Mat. Pura Appl. **110**, 353 (1976); T. Aubin, C. R. Acad. Sci. **280**, 279 (1975); G. Rosen, SIAM J. Appl. Math. **21**, 30 (1971).

On a property of a classical solution of the nonlinear mass transport equation $u_t = u_{xx}/1 + u_x^2$

Akihiko Kitada
Liberal Arts, Nippon College of Physical Education, Fukazawa, Setagaya-ku, Tokyo 158, Japan

(Received 30 July 1985; accepted for publication 21 January 1986)

A mechanism of smoothing due to evaporation–condensation of the roughly perturbed surface of solid is described by Mullins [W. W. Mullins, J. Appl. Phys. **28**, 333 (1957); **30**, 77 (1959)] in terms of the Cauchy problem (P) in R^1 (real line) of a nonlinear parabolic equation for $u(x,t)$ representing the evolution of the profile of the surface: $u_t = u_{xx}/1 + u_x^2$, $(x,t)\in R^1\times(0,\infty)$; $u(x,0) = \alpha(x), x\in R^1$ [model (P)]. In the present paper, it is demonstrated that each peak height of the initial surface $\alpha(x)$ in Mullins' model (P) does not increase with time.

I. INTRODUCTION

Mullins[1] reported that a mechanism of surface smoothing in a solid due to evaporation–condensation could be formulated as the Cauchy problem (P) in R^1 (real line) of a nonlinear mass transport equation describing the evolution of the profile $u(x,t)$ of the surface of solid:

$$\left.\begin{array}{l} u_t = u_{xx}/1 + u_x^2, \quad (x,t)\in R^1\times(0,\infty) \\ u(x,0) = \alpha(x) \quad [\alpha\in\mathscr{B}^1(R^1)], \quad x\in R^1 \end{array}\right\} (\mathrm{P}).$$

[Here $\mathscr{B}^1(R^1) = \{f; f\in C^1(R^1), |f(x), f'(x)| < \infty\}$. In the present paper, $f'(x)$ denotes the first derivative of $f(x)$, and $C^1(R^1)$ is a set of all continuously differentiable functions defined on R^1.]

Mullins verified the physical validities of the model (P), i.e., the smoothing properties of solutions of (P), by making the problem (P) analytically soluble in the classical sense,[2] through the linearization of the nonlinear equation in (P) by the small slope approximation (ssa) characterized by the relation $|u_x(x,t)| \ll 1$ for any point $(x,t)\in R^1\times(0,\infty)$. But, necessarily, this linearization based on the ssa has restricted the extent of application of this model. In fact, the unique solution of (P) under the ssa is described as

$$u(x,t) = \int_{R^1} H(x-y,t)\alpha(y)dy,$$

$$H(p,q) = (1/2\sqrt{\pi q})\exp(-p^2/4q).$$

The operation, permitted by Lebesgue's convergence theorem,

$$\lim_{t\downarrow 0} u_x(x,t) = \lim_{t\downarrow 0} \frac{1}{\sqrt{\pi}} \int_{R^1} e^{-k^2}\alpha'(x - 2\sqrt{t}\,k)dk = \alpha'(x)$$

(compact convergence) demands that the slope of the peak in the initial surface also should be small enough, namely, $|\alpha'(x)| \ll 1$.

In the present paper, without employing this linearization, by demonstrating that each peak height of the initial surface $\alpha(x)$ does not increase with time, we shall indirectly show that the nonlinear model (P) does not contradict the smoothing properties of the surface of solid. (In the same way, we can also demonstrate that each valley of the initial surface does not increase in depth with time. Both height of the peak and depth of the valley are measured from the origin of the vertical axis.)

II. A PROPERTY OF A SOLUTION OF A CAUCHY PROBLEM FOR $u_t = u_{xx}/1 + u_x^2$

Consider a Cauchy problem (P*) for a more general nonlinear parabolic equation:

$$\left.\begin{array}{l} u_t = F(u_x, u_{xx}) \quad [F\in C^1(R^2)], \\ \qquad (x,t)\in R^1\times(0,\infty) \\ u(x,0) = \alpha(x), \quad x\in R^1 \end{array}\right\} (\mathrm{P}^*),$$

where R^2 denotes the real plane.

Let x_0 be a point at which $\alpha(x)$ attains the strict relative maximum (local maximum), i.e., there exists a peak whose top is located at x_0; and let the curve C in $R^1\times[0,t_f]$ be a part of a trajectory drawn by the migration with time of the peak top initially located at x_0. Suppose C is characterized by a differentiable function $g(t)$ on the interval $[0,t_f]$ as indicated in (1):

$$C = \{(x,t); x = g(t)\ (x_0 = g(0)),\ t\in[0,t_f]\}. \tag{1}$$

From the definition of the set C, at any point in C, the following relations hold:

$$u_x(x,t) = 0, \quad u_{xx}(x,t) < 0. \tag{2}$$

Theorem: Let the conditions

$$\frac{\partial F(p,q)}{\partial q} > 0, \quad F(0,0) = 0, \tag{3}$$

hold for the right-hand side $F(p,q)$ of the nonlinear equation in (P*). Suppose, for the classical solution[2] $u(x,t)$ of (P*), there exists a set C characterized by (1) and (2). Then,

$$u(x,t) < \alpha(x_0), \quad (x,t)\in C\setminus\{(x_0,0)\}, \tag{4}$$

where $A\setminus B = \{x; x\in A, x\notin B\}$.

Before demonstrating this theorem, we prepare a lemma for a linear parabolic equation.

Lemma: Let $u(x,t)$ be a classical solution of the following linear problem (LP):

$$\left.\begin{array}{l} u_t - a(x,t)u_{xx} + b(x,t)u_x = f(x,t), \\ \qquad [a(x,t) > 0, \quad |f(x,t)| < \infty], \\ \qquad (x,t)\in R^1\times(0,\infty) \\ u(x,0) = \alpha(x), \quad x\in R^1 \end{array}\right\} (\mathrm{LP}).$$

Suppose, for $u(x,t)$, there exists a set C with a form of (1) and, on the set C, $u(x,t)$ satisfies the relations (2). Then the estimate (5) is valid for any point $(x^*,t^*)\in C\setminus\{(x_0,0)\}$,

 0022-2488/86/051391-02$02.50

$$u(x^*,t^*)\leqslant\max\{\sup_{C\setminus\{(x_0,0)\}}[f(x,t)e^{\lambda(t^*-t)}/\lambda],$$
$$\alpha(x_0)e^{\lambda t^*}\},\qquad (5)$$

where λ is an arbitrary positive number.

Proof of Lemma: Set $v(x,t)=u(x,t)e^{-\lambda t}$. On the set $C\setminus\{(x_0,0)\}$, $v(x,t)$ satisfies the linear equation

$$v_t-a(x,t)v_{xx}+b(x,t)v_x+\lambda v=f(x,t)e^{-\lambda t}.$$

Since $v(x,t)$ is continuous on the compact set C in R^2, $v(x,t)$ attains its maximum on C at some point (ξ,η) in C. The point (ξ,η) lies either in $C\setminus\{(x_0,0)\}$ or in the singleton set $\{(x_0,0)\}$.

First, let the point (ξ,η) exist in the set $C\setminus\{(x_0,0)\}$. From the relations (1) and (2), we obtain

$$v_x(\xi,\eta)g'(\eta)+v_t(\xi,\eta)\geqslant 0.$$

Since $v_x(\xi,\eta)=u_x(\xi,\eta)e^{-\lambda\eta}=0$, we get

$$v_t(\xi,\eta)\geqslant 0.$$

Namely, at the point (ξ,η), it must hold that

$$v_x(\xi,\eta)=0,\quad v_{xx}(\xi,\eta)<0,\quad v_t(\xi,\eta)\geqslant 0.$$

Therefore, for any $(\tilde{x},\tilde{t})\in C\setminus\{(x_0,0)\}$,

$$v(\tilde{x},\tilde{t})\leqslant v(\xi,\eta)\leqslant\sup_{C\setminus\{(x_0,0)\}}[f(x,t)e^{-\lambda t}/\lambda].$$

Taking (x^*,t^*) as $(\tilde{x},\tilde{t})$ and replacing $v(x^*,t^*)$ by $u(x^*,t^*)$, we can obtain an estimate

$$u(x^*,t^*)\leqslant\sup_{C\setminus\{(x_0,0)\}}[f(x,t)e^{\lambda(t^*-t)}/\lambda].$$

If the maximum on C is attained at the point $(x_0,0)$, the estimate

$$u(x^*,t^*)\leqslant\alpha(x_0)e^{\lambda t^*}$$

is valid.

These two estimates are unified to (5). □

Proof of Theorem: Let $u(x,t)$ and $v(x,t)$ be classical solutions of the problems (P*) and (P**), respectively. Then

$$\left.\begin{array}{ll}v_t=F(v_x,v_{xx}), & (x,t)\in R^1\times(0,\infty)\\ v(x,0)=\alpha^*(x), & x\in R^1\end{array}\right\}\ (\mathrm{P}^{**}).$$

Taking the differences between two nonlinear equations, we can get the following linear homogeneous equation for $w(x,t)=u(x,t)-v(x,t)$:

$$u_t-F(u_x,u_{xx})-v_t+F(v_x,v_{xx})=w_t-\int_0^1\frac{dF(hu_x+(1-h)v_x,hu_{xx}+(1-h)v_{xx})}{dh}dh$$
$$=w_t-\left\{w_{xx}\int_0^1\frac{\partial F(hu_x+(1-h)v_x,hu_{xx}+(1-h)v_{xx})}{\partial q}dh\right.$$
$$\left.+w_x\int_0^1\frac{\partial F(hu_x+(1-h)v_x,hu_{xx}+(1-h)v_{xx})}{\partial p}dh\right\}=0.$$

If we set $\alpha^*\equiv 0$, it is easy to see from the second relation in (3) that $v(x,t)\equiv 0$ is a solution of (P**), that is, $w(x,t)$ identically equals $u(x,t)$. Then $u(x,t)$, the solution of (P*), must satisfy a homogeneous linear equation

$$u_t-u_{xx}\int_0^1\frac{\partial F(hu_x,hu_{xx})}{\partial q}dh$$
$$-u_x\int_0^1\frac{\partial F(hu_x,hu_{xx})}{\partial p}dh=0.$$

Applying the estimate (5) in the lemma to this function $u(x,t)$, we can easily obtain, for any point $(x^*,t^*)\in C\setminus\{(x_0,0)\}$,

$$u(x^*,t^*)\leqslant\max\{0,\alpha(x_0)e^{\lambda t^*}\}.$$

Since we may take $\alpha(x_0)$ positive[3] and λ is arbitrary, it holds that

$$u(x^*,t^*)\leqslant\alpha(x_0).$$

Thus, we can obtain the estimate (4). □

Since $F\in C^1(R^2)$, $\partial F/\partial q=1/1+p^2>0$, and $F(0,0)=0$, the conditions required in (P*) and (3) are well satisfied for the right-hand side of the nonlinear mass transport equation $u_t=F(u_x,u_{xx})=u_{xx}/1+u_x^2$. Therefore, if the solution $u(x,t)$ of the Cauchy problem (P) has such a set as C characterized by the relations (1) and (2), the desired result

$$u(x,t)\leqslant\alpha(x_0),\quad (x,t)\in C\setminus\{(x_0,0)\},$$

is obtained.

APPENDIX: THE ESTIMATE WITH $\alpha(x_0)<0$

If $u(x,t)$ is a solution of the problem (P*), $\tilde{u}(x,t)=u(x,t)+C$ (C is an arbitrary constant) is a solution of the following Cauchy problem ($\tilde{\mathrm{P}}$*):

$$\left.\begin{array}{ll}\tilde{u}_t=F(\tilde{u}_x,\tilde{u}_{xx}), & (x,t)\in R^1\times(0,\infty)\\ \tilde{u}(x,0)=\alpha(x)+C, & x\in R^1\end{array}\right\}\ (\tilde{\mathrm{P}}^*).$$

In the same manner as in the text, we can obtain the estimate

$$u(x^*,t^*)+C\leqslant\tilde{u}(x^*,t^*)\leqslant\max\{0,(\alpha(x_0)+C)e^{\lambda t^*}\},$$

where C is a positive constant such that $|\alpha(x_0)|<C$. Hence,

$$u(x^*,t^*)\leqslant\alpha(x_0)e^{\lambda t^*}+C(e^{\lambda t^*}-1).$$

[1] W. W. Mullins, J. Appl. Phys. **28**, 333 (1957); **30**, 77 (1959).

[2] O. A. Ladyzenskaja, V. A. Solonikov, and N. N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type* (Am. Math. Soc., Providence, RI, 1968), p. 12.

[3] Even if $\alpha(x_0)<0$, the same results are obtained. See the Appendix.

Differential geometry in the large and compactification of higher-dimensional gravity

I. J. Muzinich[a)]
Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853

(Received 26 September 1985; accepted for publication 8 January 1986)

Some well-known results from differential geometry are applied to some of the major issues of compactification of higher-dimensional gravity. The results apply both to the theories generally known as Kaluza–Klein and the recently more promising super string theories. These results are primarily due to Yano [K. Yano, *Integral Formulas in Differential Geometry* (Marcel Dekker, New York, 1970); *Differential Geometry on Complex and Almost Complex Manifolds* (Macmillian, New York, 1965)] and have profound implications for the Kaluza–Klein scenario with respect to the cosmological constant problem and the massless sector of the theory. While the results are well known in the mathematical literature, the present author has only seen a fragmentary account presented by a few physicists. The necessary introduction to complex manifolds is also provided including Kähler manifolds and their possible relevance to the problem of compactification. The Ricci tensor provides the central role in the discussion of metric isometries, holomorphy, and holonomy. The incumbent role of Calabi–Yau manifolds with Ricci flat curvature and SU(n) holonomy, which have been recently conjectured in regard to super string compactification, is also mentioned.

I. INTRODUCTION

Compactification in the Kaluza–Klein framework[1] and in the more recent super string theories has been and continues to be a formidable obstruction in obtaining physically acceptable theories of gravity and gauge theories in four dimensions. Usually such theories are formulated in greater than four dimensions, and have an aesthetic and necessary motivation in higher dimensions. Next, a particular ground state $M_4 \otimes C$ has to be singled out, where M_4 is the usual Minkowski space and C is a compact manifold.

In the Kaluza–Klein scenario,[1] C is a manifold whose metric contains isometries that are physically the transformations of a compact non-Abelian Lie group on gauge fields that lift M_4 into the principal bundle on $M_4 \otimes C$. These symmetries are those of the weak, strong, and electromagnetic forces in nature. This stands in marked contrast to the super string theory where the gauge symmetries are supported by gauge fields defined on ten-dimensional Minkowski space M_{10}. The particular gauge groups $E_8 \otimes E_8$ or SO(32) are motivated for different reasons, namely freedom from anomalies and ultraviolet finiteness. This has been examined in the zero slope limit or field theory limit recently by Green and Schwarz.[2] It is apparent that the six-manifold C need not support any additional compact symmetries since all symmetries are manifestly taken into account by the Lie algebra valued gauge fields. The selection of the compact manifold is purely a dynamic question related to the true quantum ground state of the super string theory. It has been conjectured recently that an appropriate manifold will be relevant to the understanding of the number of fermion generations.[3] The number of generations in this scenario is one half the Euler characteristic of the manifold C.

In this paper, I would like to formalize some general notions on the use of results in global differential geometry[4] on the inevitability of Ricci positive manifolds for the compact manifold C in the Kaluza–Klein strategy. The definition of a Ricci positive manifold is given simply: Let V_i be any vector field on C, then $R_{ij}V^iV^j = R(V,V) > 0$ for $V \neq 0$ on a Ricci positive manifold with Ricci curvature $R(X,Y)$. In the context of Einstein manifolds, compatibility with the equations of motion introduces the problem of the cosmological constant and the incumbent lack of harmonic fermions and vector fields. If some of these considerations were adequately noted, many of the papers on the Kaluza–Klein strategy would have been precluded. While these results are generally well known to a few, I have found only passing remarks in the physics literature.[5]

The other remarks that will be focused upon are directed toward Ricci flat compact manifolds, which have no infinitesimal continuous symmetries (Killing vectors) unless they are trivial such as the multitorus. Generally such manifolds are complex structures and fall into two classes if they are Kähler manifolds: (i) multitori, complex parallelizable, and (ii) Calibi–Yau manifolds.

In the latter case, Yau proves[6] that if the first Chern class (two-form) $\Omega = \sqrt{-1}R_{i\bar{j}}\, dz^i \wedge d\bar{z}^j$, $(z^i, \bar{z}^i)$ complex coordinates vanishes as a cohomology class, then there exists a metric $g_{i\bar{j}}$ such that Ω vanishes as a differential form. Again the Ricci tensor is the fundamental object in classifying the symmetries, holonomy, and holomorphic vector fields on such manifolds. The classic results of Yano and Bochner[4] are relevant to this exposition, which is hopefully made clear and accessible to most readers.

Comments are also directed to the strategy of Candelas *et al.*[3] on the question of the Euler characteristic on such manifolds. This problem may have fundamental importance to the question of the family problem (the number of generations of light $\ll m_{\text{Planck}}$ chiral fermions).

The organization of this paper is as follows. In Sec. II

a) Address after July 1985: Institute for Theoretical Physics, University of California, Santa Barbara, California 93106. On leave from Brookhaven National Laboratory, Upton, New York 11973.

 0022-2488/86/051393-05$02.50

some notations and conventions are fixed, and the integral formula of differential geometry due to Yano is derived. These results are then applied to the Kaluza–Klein strategy explicitly. Section III deals with the analogous results on complex manifolds. Elementary definitions and concepts are introduced. The reader can skip Sec. III on first reading since it is mathematical in nature. Next, certain general results on compact complex structures are made and summarily applied to the super string compactification. Finally, we discuss the physical relevance, if any, of all of these considerations.

II. THE YANO INTEGRAL FORMULA

In classical differential geometry, integral formulas have been useful since the time of Gauss and many global results have been obtained. About 40 years ago, Bochner, Chern, Hodge, and others introduced these very powerful results into Riemannian geometry.[4] The purpose of this section is to introduce these techniques into the physics literature and apply them substantially to the problem of compactification of the Kaluza–Klein theories. A complete treatment of integral formulas can be found in the book by Yano,[4] as well as in his general introduction to Riemannian geometry. In the entirety of this section we consider a compact orientable manifold C without boundary. The orientability of C is added to insure that integration over C can be considered as an integration over differential forms.

First, we recall the Green's theorem. In a compact orientable manifold C without boundary, we have for a vector field X_i,

$$\int_C \delta X = \int_C \nabla_i X^i \, dV = 0, \qquad (2.1)$$

where $dV = \sqrt{g}\, d\xi^1\, d\xi^2 \cdots d\xi^n$; g is the Riemannian metric on C, and $\xi^1, \xi^2, \ldots, \xi^n$ is a coordinatization of C. The ∇_i is the covariant derivative $(\nabla_i X)^j = \partial_i X^j + \Gamma^j_{ik} X^k$. We can realize Eq. (2.1) in terms of differential forms: If ω is a p form $p < n$ on C (n = dimensionality of C), then

$$\begin{aligned} d\omega &= [1/(p+1)!](\partial_i \omega_{i_1 \cdots i_p} - \partial_{i_1} \omega_{i i_2 \cdots i_p} \\ &\quad - \partial_{i_p} \omega_{i_1 i_2 \cdots i}) d\xi^i d\xi^{i_1} \cdots d\xi^{i_p}, \\ \delta\omega &= {*d*\omega} \\ &= [1/(p-1)!](g^{ji} \nabla_j \omega_{i i_2 \cdots i_p}) d\xi^{i_2} \cdots d\xi^{i_p}. \end{aligned} \qquad (2.2)$$

The * operation is defined[4] in terms of the antisymmetric symbol ϵ as

$$*d\xi^{i_1} \cdots d\xi^{i_p} = \frac{(g)^{1/2}}{(n-p)!} \epsilon^{i_1 \cdots i_p}{}_{i_{p+1} \cdots i_n} d\xi^{i_{p+1}} \cdots d\xi^{i_n},$$

and the Green's theorem for a manifold without boundary $\partial C = 0$ is simply the statement

$$\int_C d*X = \int_{\partial C} *X = 0, \qquad (2.3)$$

which is Eq. (2.1) and X is a vector field.

Next we use the Green's theorem for the particular vector field

$$\begin{aligned} W^i &= (X^j \nabla_j X)^i - (\nabla_j X^j X)^i, \\ \nabla_i W^i &= (\nabla_i X)^j (\nabla_j X)^i + X^j (\nabla_i \nabla_j X)^i \\ &\quad - (\nabla_i \nabla_j X)^j X^i - ((\nabla_j X)^j)^2 \\ &= R_{ij} X^j X_i + (\nabla^j X)^i (\nabla_i X)^j - (\nabla_j X^j)^2. \end{aligned} \qquad (2.4)$$

In Eq. (2.4) the Ricci identity

$$(\nabla_i \nabla_j X)^i - (\nabla_j \nabla_i X)^i = R_{ji} X^i$$

has been used. From Eq. (2.1) we obtain the central result

$$\int_C [R(X,X) + (\nabla^j X)^i (\nabla_i X)^j - (\nabla_j X^j)^2] = 0. \qquad (2.5)$$

We now discuss various consequences of Yano's formula Eq. (2.5). First, let X be a Killing vector field; the manifold has infinitesimal isometries. Killing's equation is derived from the fact that the Lie derivative[4] of the metric g is zero:

$$\mathscr{L}_X g = X \nabla g(Y,Z) = 0, \qquad (2.6)$$

or in terms of the more familiar coordinate basis

$$X^k \nabla_k g_{ij} + g_{kj} \nabla_i X^k + g_{ik} \nabla_j X^k = 0. \qquad (2.7)$$

From the fact that $\nabla g = 0$, Killing's equation

$$\nabla_i X_j + \nabla_j X_i = 0 \qquad (2.8)$$

emerges.

Then Yano's formula Eq. (2.5) for a Killing vector is simply

$$\langle X, RX \rangle - \langle \nabla X, \nabla X \rangle = 0, \qquad (2.9)$$

where an obvious scalar product notation is used for L_2 integration over C.

From Eq. (2.9) we can derive some obvious but powerful consequences. First, if $R(X,X)$ and hence $\langle X,RX \rangle < 0$ (Ricci negative); positivity of $\|\nabla X\|$ implies that there are no nonzero Killing vectors or infinitesimal continuous isometries. If $R(X,X) = 0$, the Ricci flat case, then for a compact manifold of Euclidean signature, $\|\nabla X\| = 0$ implies that $\nabla X = 0$ identically. It follows that if there exists a complete set of covariantly constant Killing vectors $\frac{1}{2}n(n+1)$ for M_n, then the Riemann tensor vanishes.[4] The manifold is either Euclidean space (noncompact) or the multitorus $S^1 \otimes S^1 \cdots S^1$ (compact). Therefore, nontrivial Ricci flat manifolds with nonvanishing Riemann curvature $R(X,Y,U,V)$, Killing vectors, and the related continuous symmetries do not exist. In conclusion, compact manifolds that admit nontrivial isometries, those of a compact non-Abelian Lie group, are Ricci positive, $R(X,X) > 0$. Examples consist of Lie group manifolds, and homogeneous spaces such as spheres.

Unfortunately, for the standard Kaluza–Klein scenario with the usual Einstein manifolds $R(X,Y) = \lambda g(X,Y)$ the cosmological constant problem is a *fait accompli*. The cosmological constant is 0 (gauge coupling $\times m^2_{\rm Planck}$). Especially since we are looking for solutions to the Einstein equations with cosmological constant λ,

$$R_{AB} - \tfrac{1}{2} g_{AB} R = \lambda g_{AB}, \qquad (2.10)$$

with a $M_4 \otimes C$ ground state. Since $R_{AB} = R_{\mu\nu} \oplus R_{ij}$, where $R_{\mu\nu}$ and R_{ij} are the Ricci tensors on M_4 and C, respectively, and $R_{\mu\nu} = 0$ for M_4, $R_{ij}X^iX^j > 0$ implies $R > 0$ and $\lambda > 0$; $AB = (1,...,4+n)$, $\mu\nu = (1,...,4)$, $i = (4+1,...,4+n)$. In addition to the cosmological constant problem, there is the related problem of massless chiral fermions, as has been identified in the literature.[1,3] The argument presented here is similar in spirit to that of Duff[1] but perhaps more detailed in its presentation.

Another feature of Yano's integral formula worth noting is the absence of harmonic vectors on Ricci positive manifolds, for harmonic vector fields $\nabla_i X_j - \nabla_j X_i = 0$, $\nabla_i X^i = 0$. Yano's integral formula then takes the form

$$\int_C R(X,X) + \nabla_i X_j \nabla^i X^j = \langle X,RX\rangle + \langle \nabla X,\nabla X\rangle. \tag{2.11}$$

However, for a Ricci positive manifold with Euclidean signature it is obvious that

$$\langle X,RX\rangle + \langle \nabla X,\nabla X\rangle = \langle X,RX\rangle + \langle X,-\Delta X\rangle \geqslant 0. \tag{2.12}$$

Hence, the only possible harmonic vector is $X = 0$. This is the analogous statement for harmonic vectors that Lichnerowicz proved for harmonic spinors.[1]

Finally the above statements do not preculde the existence of Killing spinors. The analogous integrability theorem for Killing spinors, spinors satisfying

$$\nabla_a \eta^\pm = \pm C\gamma_a \eta^\pm, \tag{2.13}$$

is that the manifold be the particular Einstein manifold[6] with C a constant:

$$R_{ij} = 4C^2(n-1)\, g_{ij}. \tag{2.14}$$

The Killing spinors are intimately related to the existence of super symmetries and the particular case $\nabla\eta = 0$ implies Ricci flatness, which is conjectured to be relevant to the compactification of the super string theory. We will turn to these questions in subsequent sections.

III. COMPLEX MANIFOLDS

In Sec. II, the Ricci tensor occupies a special position in the discussion of the isometries of real Riemannian manifolds with the usual notion of differentiability. This section will be a brief introduction to complex manifolds[4,7] such that the analogous results can be established in Sec. IV.

A complex manifold of complex dimensions n is a $2n$-dimensional topological manifold endowed with a complex analytic structure. The notion of holomorphy replaces that of differentiability. A separable Hausdorff space is said to have a complex analytic structure if the following properties hold: (i) each point of M has a neighborhood homeomorphic with an open subset of C_n (the space of n complex variables); and (ii) for any pair of open sets U_1 and U_2 with nonempty intersection, the map $u_1 u_2^{-1}$: $U_1 \cap U_2 \to C_n$ is defined by holomorphic functions.

The coordinates on a complex manifold can be taken locally as

$$z^k = x^k + iy^k, \quad \bar{z}^k = x^k - iy^k, \tag{3.1}$$

where x^k and y^k are real coordinates $k = (1,...,n)$. Differential forms are of bidegree (q,r), where the form of degree p consists of q, dz^k and r, $d\bar{z}^k$, with $p = q + r$. Let

$$\omega^p = \sum \omega_{i_1 \cdots i_q \bar{j}_1 \cdots \bar{j}_r} dz^k \wedge \cdots \wedge dz^{i_q}\, d\bar{z}^{\bar{j}_1} \wedge \cdots \wedge d\bar{z}^{\bar{j}_r} \tag{3.2a}$$

be such a form; then the exterior derivative in analogy to Eq. (2.2) is

$$d:\ \omega^p \to \underbrace{\omega^{(q+1,r)}}_{d'} + \underbrace{\omega^{(q,r+1)}}_{d''}. \tag{3.2b}$$

The operations d' and d'' are defined in terms of $\partial/\partial z^i$ and $\partial/\partial \bar{z}^i$.

Examples of complex manifolds are (1) the space of n complex variables C_n, (2) the Riemann sphere S^2, and (3) complex projective space CP_n, the space of complex lines through the origin of C_{n+1}.

The complex structure is an instance of a more general type of structure, which is defined by the complexification of the tangent space. If a differentiable manifold contains a complex structure in the tangent space at every point, then such a manifold is said to have an almost complex structure. It is obvious that such a differentiable manifold must be of real dimension $2n$. The tangent space consists of a complex vector space V^c over the field of complex numbers, which is the complexification of the vector space V over the field of real numbers. Vectors $v \in V^c$ have the unique representation

$$v = w_1 + \overline{w}_2,$$
$$w_1 \in W^c, \quad w_2 \in W^c, \quad \text{where} \quad V = W^c \oplus \overline{W^c}. \tag{3.3}$$

An almost complex structure is defined by the tensor of type (1,1) with the property

$$Jv = iw - i\,\overline{w}, \tag{3.4}$$

for every real vector $v = w + \overline{w}$. The operator J has the properties of linearity and $J^2 = -1$. The vector space V^c can be decomposed $V^c = V^{1,0} + V^{0,1}$ with vectors of bidegree (1,0) and (0,1). And, as mentioned before, J is of bidegree (1,1). It is conventional to let the indices run as follows: $1,...2n$ with $k,j = (1,...,n)$, $\bar{k},\bar{j} = (n,...2n)$ with complex basis $(e_j, e_{\bar{j}})$. With respect to the real basis $f_j, f_{\bar{j}}$,

$$f_j = (1/\sqrt{2})(e_j + e_{\bar{j}}), \quad f_{\bar{j}} = (i/\sqrt{2})(e_j - e_{\bar{j}}), \tag{3.5}$$

the tensor J is given by

$$J = \left(\begin{array}{c|c} 0 & I_n \\ \hline -I_n & 0 \end{array}\right), \tag{3.6}$$

where I_n is the identity with respect to V_n. Conversely with respect to the basis $(e_j, e_{\bar{j}}) J_j^{\ k} = i\delta_j^k$ and

$$J_{\bar{j}}^{\ \bar{k}} = -i\delta_{\bar{j}}^{\bar{k}}, \quad J_j^{\ \bar{k}} = J_{\bar{j}}^{\ k} = 0.$$

The metric on V^c has the properties of a Hermitian structure

$$g(Jv_1, Jv_2) = g(v_1, v_2), \tag{3.7}$$

with mixed nonvanishing components $g_{k\bar{j}}$, symmetry $g_{k\bar{j}} = g_{\bar{j}k}$, and Hermiticity $\overline{g_{k\bar{j}}} = g_{j\bar{k}}$.

In analogy with the real case the concepts of frame and coframe can be introduced

$$X_a = \xi_a^{\ k} \frac{\partial}{\partial z^k}, \quad X_{\bar{a}} = \bar{\xi}_a^{\ k} \frac{\partial}{\partial \bar{z}^k},$$
$$\alpha^a = \xi_j^{\ a}\, dz^j, \quad \overline{\alpha}^a = \xi_j^{\ a}\, d\bar{z}^j, \tag{3.8}$$

with the properties $\delta_{ab} = \xi^{\,k}_{\,a}\bar{\xi}^{\,l}_{\,b} g_{k\bar{l}}$. The indices a and b are vielbein indices transforming under $U(n)$ and i,j, etc. are tensor indices transforming under general coordinate transformations $\xi^{(i)}_{k} = (\partial z^j/\partial z'^k)\xi^{(i)}_{k}$, etc.

Next, structural and metricity conditions can be used to define connection and curvature forms, respectively. The differential of the metric tensor is evaluated

$$dg_{i\bar{j}} = \omega^k_i g_{k\bar{j}} + \bar{\omega}^k_j g_{i\bar{k}} = \frac{\partial g_{i\bar{j}}}{\partial z^m} dz^m + \frac{\partial g_{i\bar{j}}}{\partial \bar{z}^m} d\bar{z}^m. \qquad (3.9)$$

The connection one-forms ω^k_i and $\bar{\omega}^k_j$ are unmixed and can be taken to be

$$\omega^i_j = \Gamma^i_{jk}\, dz^k, \quad \bar{\omega}^i_j = \overline{\Gamma^i_{jk}}\, d\bar{z}^k, \qquad (3.10a)$$

with

$$\Gamma^i_{ik} = g^{i\bar{l}} \frac{\partial}{\partial z_k} g_{\bar{l}j} \quad \text{and} \quad \Gamma^{\bar{i}}_{\bar{j}\bar{k}} = \overline{\Gamma^i_{jk}}. \qquad (3.10b)$$

The equations of structure read

$$\theta^i = d\alpha^i + \omega^i_j \alpha^j. \qquad (3.11)$$

Elementary computation gives for the torsion form θ^i the results for the torsion tensor

$$T^i_{jk} = \Gamma^i_{jk} - \Gamma^i_{kj} = g^{i\bar{l}}\left(\frac{\partial}{\partial z^k} g_{\bar{l}j} - \frac{\partial}{\partial z^k} g_{\bar{l}k}\right),$$

$$T^{\bar{i}}_{\bar{j}\bar{k}} = g^{\bar{i}l}\left(\frac{\partial}{\partial \bar{z}^k} g_{l\bar{j}} - \frac{\partial}{\partial \bar{z}^k} g_{l\bar{k}}\right). \qquad (3.12)$$

The curvature form in local coordinates $(z^i, \bar{z}^i)$ is denoted by Ω^j_i and is given locally by

$$\Omega^j_i = d\omega^j_i + \omega^j_r \wedge \omega^r_i. \qquad (3.13)$$

Using the decomposition

$$\Omega^j_i = \tfrac{1}{2}(R^{\,j}_{\,ilm} dz^l \wedge dz^m + R^{\,j}_{\,il\bar{m}} dz^l \wedge d\bar{z}^m),$$

where $R^{\,j}_{\,ilm} + R^{\,j}_{\,iml} = 0$, we obtain from Eqs. (3.10a) and (3.10b) the result

$$R^{\,j}_{\,ilm} = \frac{\partial \Gamma^j_{il}}{\partial z^m} - \frac{\partial \Gamma^j_{im}}{\partial z^l} + \Gamma^r_{il}\Gamma^j_{rm} - \Gamma^r_{im}\Gamma^j_{rl} = 0.$$

Remark: This is obvious since Γ is a pure gauge $\Gamma = g^{-1}(\partial/\partial z)\, g$. Therefore, the only nonvanishing components of the curvature form are $R^{\,j}_{\,il\bar{m}}$, $R^{\,j}_{\,i\bar{l}m}$, $R^{\,\bar{j}}_{\,\bar{i}\bar{l}m}$, and $R^{\,j}_{\,i\bar{l}\bar{m}}$. Furthermore,

$$R^{\,j}_{\,il\bar{m}} = \frac{\partial \Gamma^j_{il}}{\partial \bar{z}^m} = \frac{\partial g^{kj}}{\partial \bar{z}^m}\frac{\partial g_{i\bar{k}}}{\partial z^l} + g^{j\bar{k}}\frac{\partial^2 g}{\partial z^l\, \partial \bar{z}^m}. \qquad (3.14)$$

Obviously, if the connection is holomorphic $(\partial/\partial\bar{z})\Gamma = 0$, the manifold is flat [the conditions

$$\frac{\partial}{\partial \bar{z}}(\ ) = \left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)(\ ) = 0$$

are equivalent to the Cauchy Riemann conditions].

In addition, Kähler geometry can be introduced at this point. A Hermitian metric for which the torsion form vanishes, Eq. (3.12), leads to the conditions

$$\frac{\partial g_{\bar{i}j}}{\partial z^k} = \frac{\partial g_{\bar{i}k}}{\partial z^j}, \quad \frac{\partial g_{i\bar{j}}}{\partial \bar{z}^k} = \frac{\partial g_{i\bar{k}}}{\partial \bar{z}^j}. \qquad (3.15)$$

A complex manifold with this particular metric is called a Kähler manifold. The system of equations (3.15) is completely integrable and there exists a real valued function K for which $g_{i\bar{j}} = \partial^2 K/\partial z^i\, \partial\bar{z}^j$, and from Eq. (3.15) it is obvious that the Kähler form $\sqrt{-1}\, g_{i\bar{j}}\, dz^i \wedge d\bar{z}^j$ is closed.

The Ricci curvature $R_{l\bar{m}}$ is defined by

$$R^{\,l}_{\,il\bar{m}} = R_{l\bar{m}} = \frac{\partial^2 \log\det(|g|)}{\partial z^l \partial \bar{z}^m}. \qquad (3.16)$$

The (1,1) form

$$\psi = \tfrac{1}{2}\pi\sqrt{-1}R_{k\bar{l}}\, dz^k \wedge d\bar{z}^l \qquad (3.17)$$

is the first Chern class that, as we shall see later, is an obstruction for SU(n) holonomy as opposed to U(n) holonomy.

The holonomy group is now defined.[8] Let (C, g) be a connected $2n$-dimensional Riemannian manifold; for $x \in C$, let H be the holonomy group at x for Levi-Cività connection. The $H \subset O(C)$, where $O(C)$ is isomorphic to $O(2n)$ the transformation of the tangent space at x. The main results of a number of people[8] is that (C, g) is the Riemannian structure underlying a Kähler structure if there exists an imbedding $H \subset U(C) \subset O(C)$. Basically J [Eq. (3.4)] has to be defined and U(n) is imbedded into O(C). The element, multiplication by i, has to be reached smoothly in the tangent space; $J \in O(2n)$, $JJ^{\text{tr}} = 1$, $J = -J^{\text{tr}}$, $J^2 + 1 = 0$. The elements of the holonomy group for Levi-Cività connection ω are defined with path ordering P:

$$u = P\exp\int \omega_k\, dx^k. \qquad (3.18)$$

IV. COMPACTIFICATION IN UNIFIED THEORIES

In the Kaluza–Klein scenario, general results on the compactification to manifolds with isometries, are by now generally known. Unfortunately, most attempts, with a remote exception, of parallelizable manifolds[9] (endowed with torsion) lead to embarrassment even if quantum mechanics[5] is taken into account. The situation for the super string theories is in one sense simpler and in a different sense more complicated. It is simpler, since the manifold is apparently not required to carry any continuous isometries. The symmetries are carried by the explicit gauge fields of the anomaly-free SO(32) or $E_8 \otimes E_8$ gauge group.[2] The compactificaton is more complicated since it is not really known what class of manifolds is promising and leads to acceptable phenomenology.

In a recent work, Candelas *et al.*[3] have advanced a strategy to determine the manifold $M_4 \otimes C$ in the compactification of the ten-dimensional super string theory. They base their considerations on the fact that M_4 should be eventually Minkowski space, with an unbroken $N = 1$ supersymmetry, and the fermion spectrum should be realistic. Arguments are presented that the spaces C should be Calabi–Yau manifolds with vanishing first Chern class (Ricci-flat and Kähler) with SU(3) holonomy group. With the properties of Ricci flat and Kähler, many such six-manifolds are known to exist as subspaces of CP_4 (see Ref. 10). I am not going to comment on the wisdom or viability of such a strategy but merely address some general questions on the application of complex manifolds to the question of compactification. Most of this material is available in the mathematical literature; in particular, the monograph on curvature and homology by Goldberg[7] is useful.

First, if C is a compact Kähler manifold of constant holomorphic curvature k

$$R_{i\bar{j}k\bar{l}} = (k/2)(g_{i\bar{j}}g_{k\bar{l}} + g_{i\bar{l}}g_{k\bar{j}}), \tag{4.1}$$

then its universal covering manifold M is classified as follows: $k>0$, complex projective space; $k=0$, C_n the space of n complex variables, or the interior of the unit sphere $k<0$. Furthermore, it is known that if C is parabolic, $k=0$, it can be represented as the quotient space C_n/D, where D is a discrete group of motions in C_n. The complex torus is then a covering space of C. If C is simply connected, a necessary condition for zero curvature is complex parallelizability. Complex parallelizability means the existence of a complete set of n globally defined linearly independent holomorphic vector fields. On such a manifold there exists a Hermitian metric of zero curvature.

Proof: As in Eq. (3.8), let $z_a = \xi^k_a(\partial/\partial z^k)$, $\alpha_a = \xi^a_j\, dz^j$, etc.; holomorphy means simply $(\partial/\partial\bar{z}^j)\xi^k_a = 0$ and $\partial\xi^a_n/\partial\bar{z}^j = 0$. The connection is

$$\Gamma^i_{jk} = g^{i\bar{l}}\frac{\partial g_{j\bar{l}}}{\partial\bar{z}^k} = \sum_a \xi^i_a \frac{\partial\xi^a_i}{\partial\bar{z}^k},$$

and using Eq. (3.14), $R^{\,j}_{i\bar{l}m} = 0$. Holomorphy is simply $d''X_a = 0$. Another theorem stated without proof is that a compact complex parallelizable manifold is Kählerian if and only if it is a complex multitorus.[7]

The effect of positive Ricci curvature on holomorphy is similar to the obstruction for harmonicity proved in Sec. III. This is formalized in the following theorem: If the Ricci curvature is strictly negative, there are no holomorphic contravariant tensor fields of bidegree $(p,0)$; otherwise, a tensor field of this type must be a parallel tensor field. In particular, for negative Ricci curvature there are no holomorphic vector fields. Such manifolds are unacceptable to super string compactification in any case for the same reasons as Ricci negative is unattractive for the Kaluza–Klein theory.

However, there are nontrivial Hermitian metrics for which the Ricci tensor vanishes together with the first Chern class. Quite generally, there are inequalities on the Ricci tensor and the characteristic classes of such manifolds. These are the famous Calabi–Yau manifolds first conjectured to exist by Calabi[11] and a decisive proof presented twenty years later by Yau.[12] These manifolds do not have the infinitesimal symmetries (parallel fields) of a complex Lie group and are a start for a discussion of compactification. The lack of infinitesimal symmetries is not a detriment since the explicit gauge fields of the super string theories support all of the symmetries of nature together with chiral fermions in ten dimensions. The chirality remains upon compactification.[3] Whether this class of Calabi–Yau manifolds continues to be promising and provides acceptable phenomenology remains to be seen.

V. CONCLUSION

The effects of the Ricci curvature on the question of compactification of higher-dimensional gravity was clearly demonstrated. The relevance to the super string theory in four dimensions is obscure for many reasons. First, a covariant formalism is not developed to an extent to substantially address the question of the ground state. Second, the manifold does not seem to be very well constrained; it appears that there are a large number of possibilities. Questions of the breaking of supersymmetry and the motivation of such complex manifolds in the first place has been only mildly compelling. However, one aspect of the super string strategy is clear: the possibility of an anomaly-free finite theory of quantum gravity is very attractive.

In this paper certain mathematical aspects of compactification were presented to sharpen some of the issues involved.

ACKNOWLEDGMENTS

I wish to thank the members of the Theory Group at the Laboratory of Nuclear Studies at Cornell University during the academic year 1984–1985. I have enjoyed a number of discussions with the Cornell Group during this period of time.

This research was supported in part by the National Science Foundation under Grant No. PHY82-09011 at Cornell and Grant No. PHY82-17853 at ITP, supplemented by funds from the National Aeronautics and Space Administration.

[1]M. J. Duff, in *Supersymmetry and Supergravity 1984*, in *Proceedings of the Trieste Spring School*, Pasadena, 1984, edited by B. de Wit, P. Fayet, and P. van Nieuwenhuizen, and references contained therein.

[2]M. Green and J. Schwarz, California Institute of Technology preprints, 1984; Phys. Lett. B **149**, 117 (1984); D. J. Gross, J. A. Harvey, E. Martinec, and R. Rohm, Phys. Rev. Lett. **54**, 502 (1985).

[3]P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, Institute for Theoretical Physics, University of California preprint No. NSF-ITP-84-170, 1985.

[4]K. Yano, *Integral Formulas in Differential Geometry* (Dekker, New York, 1970); *Differential Geometry on Complex and Almost Complex Manifolds* (Macmillian, New York, 1965).

[5]See Ref. 1. This remark is also attributed to V. Moncrief, in T. Applequist and A. Chodos, Phys. Rev. D **28**, 772 (1983).

[6]P. van Nieuwenhuizen, in Ref. 1.

[7]S. I. Goldberg, *Curvature and Homology* (Dover, New York, 1962).

[8]E. Calabi, *Chern Symposium* (Springer, New York, 1979).

[9]I. J. Muzinich, J. Math. Phys. **26**, 1942 (1985); C. Destri, C. A. Orzalesi, and P. Rossi, Ann. Phys. (NY) **147**, 321 (1983); Y. S. Wu and A. Zee, J. Math. Phys. **25**, 2696 (1984).

[10]S. T. Yau, private communication to E. Witten.

[11]E. Calabi, Proc. Inst. Congr. Math. Amsterdam **2**, 206 (1954); *A Symposium in Honor of S. Lifschitz* (Princeton U. P., Princeton, NJ, 1955).

[12]S. T. Yau, Proc. Natl. Acad. Sci. USA **74**, 1798 (1977).

Separation of Dirac and Kähler equations in spherically symmetric space-times

M. Panahi and R. W. Tucker
University of Lancaster, Department of Physics, Lancaster, United Kingdom

(Received 23 September 1985; accepted for publication 3 January 1986)

The Dirac and Kähler equations are separated in a class of spherically symmetric space-times by using new techniques of Clifford analysis. New operators, which raise and lower the "spin weight" of spin-weighted functions by $\frac{1}{2}$, are also introduced. The separated solutions enable us to analyze the distinction between spinor solutions to the Dirac separation equation and Kähler solutions.

I. INTRODUCTION

The vector space isomorphism between the exterior and Clifford algebras has been observed and exploited[1,2] by many physicists since the work of Chevalley.[3] When the Clifford algebra is associated with a metric on a (pseudo-) Riemannian manifold, one may construct a useful calculus for the study of physical field theories involving gravitation.

The formulation of spinors on a manifold in terms of certain local sections of its exterior bundle, equipped with a Clifford structure, has been investigated recently.[4,5] Regarding spinors in this way affords a new insight into spinor covariance and enables one to use powerful methods of Clifford analysis to study wave equations. The Clifford bundle approach also can be used to study the relation between the Kähler equation for a general inhomogeneous differential form[6] and the local Dirac equation for a spinor on a manifold.

Since the original paper by Kähler,[7] in which the Kähler equation was solved in flat space (in a polar chart), there has been little effort in examining solutions of the Dirac equation in terms of inhomogeneous forms. Furthermore there exists in the literature a certain confusion about the conditions under which these equations may be identified.

In this paper we develop a method for separating both the Dirac and Kähler equations in space-times M with a metric that possesses Killing vectors generating the SO(3) rotation group. Although separable solutions of the Dirac equation in spherically symmetric space-times are well known from the pioneering work of Schrödinger and Chandrasekhar,[8] the methods we develop here will be in terms of a Clifford calculus of differential forms rather than the component oriented formalism developed by Newman and Penrose.[9] These methods will also be used to study the Kähler equation and to achieve a new insight into the relation between spinor harmonics and the angular harmonics used to separate the Kähler equation.

II. THE WAVE EQUATIONS

Let $\{f_i\}$, $i=1,2,3,4$, be a set of linearly independent local sections of the Clifford bundle (the exterior bundle over space-time equipped with a Clifford structure) such that $f_i = f_i P$ for some primitive idempotent P in $C_{3,1}(\mathbf{R})\times\mathbf{C}$, which represents a minimal left ideal $I_L(P)$. The juxtaposition of elements in the Clifford bundle will denote their Clifford product. If α and β are one-forms, then $\alpha\beta$ is related to the exterior product $\alpha\wedge\beta$ by the relation

$$\alpha\beta = \alpha\wedge\beta + g(\alpha,\beta), \tag{1}$$

where g is the space-time metric tensor field on one-forms. Linearity and associativity uniquely extend the above relation to general inhomogeneous forms. A local algebraic spinor ψ is defined as a local section of the Clifford bundle such that $\psi = \psi P$. In the above algebraic spinor basis it may be written as

$$\psi = \sum_i \psi_i f_i, \tag{2}$$

where $\{\psi_i\}$ may be regarded as a set of four complex functions on space-time when the section is pulled back to the space-time base manifold M. The space of all local algebraic spinors $\psi = \psi P$ will be denoted by $\Gamma(I_L(P))$.

The local Dirac equation for a particle of mass M is written as[3]

$$\not{S}\psi = M\psi, \tag{3}$$

where $\not{S} = e^a S_{X_a}$ in terms of any naturally dual local bases $\{e^a\}$, $\{X_a\}$, $e^a(X_b) = \delta^a_b$ for $\Gamma(T^*M)$ and $\Gamma(TM)$, respectively. The spinor covariant derivative S_X: $\Gamma(I_L(P))\mapsto\Gamma(I_L(P))$ can be related to the Levi-Cività connection ∇_X on arbitrary Clifford sections by

$$S_X\psi = \nabla_X\psi + \psi\Sigma_X, \tag{4}$$

where

$$\Sigma_X \equiv \tfrac{1}{4}\nabla_X e^a\wedge e_a, \quad X\in\Gamma(TM), \tag{5}$$

and $\{e^a\}$ is here an orthonormal basis for $\Gamma(T^*M)$. From the metric compatibility of ∇_X and relation (1) it follows that

$$\nabla_X(\rho\Phi) = \nabla_X\rho\Phi + \rho\nabla_X\Phi, \tag{6}$$

and, furthermore, that

$$S_X(\Phi\psi) = \nabla_X\Phi\psi + \Phi S_X\psi, \tag{7}$$

where ρ and Φ are any sections of the Clifford bundle.

The Kähler equation[4] for a particle of mass M by contrast is expressed in terms of a general inhomogeneous differential form Φ as

$$e^a\nabla_{X_a}\Phi = M\Phi, \tag{8}$$

and in general will not admit solutions lying in any minimal left ideal of the space-time Clifford bundle.

 0022-2488/86/051398-06$02.50

TABLE I. Connection one-forms $\omega^a{}_b$ for an orthonormal basis.

$a\backslash b$	0	1	2	3
0	0	$\frac{H_0' e^{-\lambda}}{H_0 H_1} e^0 + \frac{\dot\lambda}{H_0} e^1$	$\frac{\dot\lambda}{H_0} e^2$	$\frac{\dot\lambda}{H_0} e^3$
1		0	$-\frac{H_2' e^{-\lambda}}{H_1 H_2} e^2$	$-\frac{H_2' e^{-\lambda}}{H_1 H_2} e^3$
2			0	$-\frac{e^{-\lambda}\cot\theta}{H_2} e^3$
3				0

III. SEPARATION OF THE DIRAC EQUATION IN SPHERICALLY SYMMETRIC SPACE-TIMES

A class of spherically symmetric metrics on space-time that includes the Minkowski, Schwarzschild, Reissner–Nordstrøm, and Robertson–Walker metrics is conveniently represented in terms of local coordinates (t,r,θ,ϕ) and the orthonormal coframe:

$$\begin{aligned} e^0 &= H_0(r)dt, \quad 0 \leqslant t < \infty, \\ e^1 &= e^{\lambda(t)} H_1(r) dr, \\ e^2 &= e^{\lambda(t)} H_2(r) d\theta, \quad 0 < \theta \leqslant \pi/2, \\ e^3 &= e^{\lambda(t)} H_2(r) \sin\theta\, d\phi, \quad 0 \leqslant \phi < 2\pi, \end{aligned} \tag{9}$$

where H_0, H_1, and H_2 are real functions of r and λ is a real function of time. We shall leave open the domain of definition of the coordinate r to accommodate different topologies, These metrics are of Lorentzian signature $g(e^a,e^b) = \eta^{ab} = \mathrm{diag}(-,+,+,+)$. The Levi-Cività orthonormal connection one-forms $\omega^a{}_b$ defined by $\nabla_{X_a} e^b = -\omega^b{}_c(X_a) e^c$, are given in Table I, and

$$\omega_{ab} \equiv \eta_{ac}\omega^c{}_b = -\omega_{ba}.$$

All the covariant derivatives in this paper can be evaluated with the aid of Table II for $\nabla_{X_a} e^b$, where

$$\dot\lambda \equiv \frac{d}{dt}\lambda, \quad H' \equiv \frac{d}{dr} H. \tag{10}$$

If the spinor basis is constructed in terms of a Clifford polynomial of orthonormal one-forms with constant complex coefficients then $\not{D}\psi$ may be expressed as

$$\not{D}\psi = e^a(X_a\psi) + \Sigma\psi, \tag{11}$$

where the orthonormal derivative X_a differentiates the components of ψ in the above basis and $\Sigma \equiv e^a \Sigma_{X_a}$. For the above metric,

$$\Sigma = L_0 e^0 + L_1 e^1 + L_2 e^2, \tag{12}$$

with

$$\begin{aligned} L_0 &\equiv 3\dot\lambda/2H_0, \\ L_1 &\equiv \frac{e^{-\lambda}}{H_1}\left(\frac{H_0'}{2H_0} + \frac{H_2'}{H_2}\right), \\ L_2 &\equiv (e^{-\lambda}\cot\theta)/2H_2, \end{aligned} \tag{13}$$

and

$$\begin{aligned} X_0 &= (1/H_0)\partial_t, \quad X_1 = (e^{-\lambda}/H_1)\partial_r, \\ X_2 &= (e^{-\lambda}/H_2)\partial_\theta, \quad X_3 = (e^{-\lambda}/H_2 \sin\theta)\partial_\phi. \end{aligned} \tag{14}$$

Our basic approach to the separability of the Dirac equation (3) is to express the angular dependence of its solutions in terms of a spinor χ on an S^2 whose metric $\underset{(2)}{g}$ is that induced from g:

$$\begin{aligned} \underset{(2)}{g} &= e^2 \otimes e^2 + e^3 \otimes e^3 \\ &= e^{2\lambda} H_2^2 [d\theta \otimes d\theta + \sin^2\theta\, d\phi \otimes d\phi]. \end{aligned} \tag{15}$$

Here $\underset{(2)}{g}$ is conformal to the standard metric on a sphere. A basis for spinors on the upper hemisphere is

$$\tilde f_1 = \tfrac{1}{2}(1 + ie^2e^3), \quad \tilde f_2 = e^2 \tilde f_1, \tag{16}$$

and we may write

$$\chi = \chi_1 \tilde f_1 + \chi_2 \tilde f_2. \tag{17}$$

Denoting the induced spinor operator S^2 by $\mathscr{S}$, i.e.,

$$\mathscr{S}\chi = \sum_{a=2}^{3} e^a(X_a\chi) + (e^{\lambda}\cot\theta/2H_2)e^2\chi, \tag{18}$$

we require that

$$\mathscr{S}\chi = i(ke^{-\lambda}/H_2)\chi, \quad k \text{ a real constant}, \tag{19}$$

which in terms of the spinor components χ_1 and χ_2 gives

$$\partial_\theta\chi_1 - \frac{i}{\sin\theta}\partial_\phi\chi_1 + \frac{\cot\theta}{2}\chi_1 = ik\chi_2, \tag{20}$$

TABLE II. Levi-Cività covariant derivatives $\nabla_{X_a} e^b$.

	e^0	e^1	e^2	e^3
∇_{X_0}	$-\frac{H_0' e^{-\lambda}}{H_0 H_1} e^1$	$-\frac{H_0' e^{-\lambda}}{H_0 H_1} e^0$	0	0
∇_{X_1}	$-\frac{\dot\lambda}{H_0} e^1$	$-\frac{\dot\lambda}{H_0} e^0$	0	0
∇_{X_2}	$-\frac{\dot\lambda}{H_0} e^2$	$\frac{H_2' e^{-\lambda}}{H_1 H_2} e^2$	$-\frac{\dot\lambda}{H_0} e^0 - \frac{H_2' e^{-\lambda}}{H_1 H_2} e^1$	0
∇_{X_3}	$-\frac{\dot\lambda}{H_0} e^3$	$\frac{H_2' e^{-\lambda}}{H_1 H_2} e^3$	$\frac{e^{-\lambda}\cot\theta}{H_2} e^3$	$-\frac{\dot\lambda}{H_0} e^0 - \frac{H_2' e^{-\lambda}}{H_1 H_2} e^1 - \frac{e^{-\lambda}\cot\theta}{H_2} e^2$

$$\partial_\theta \chi_2 + \frac{i}{\sin\theta}\partial_\phi \chi_2 + \frac{\cot\theta}{2}\chi_2 = ik\chi_1. \quad (21)$$

Thus χ_1 and χ_2 are "spin-$\frac{1}{2}$ weighted harmonics."[9]

A spinor on space-time may be taken to lie in the ideal generated by the primitive idempotent $P = (1 + ie^2e^3)T$, where $T = \frac{1}{2}(1 + ie^0)$. A suitable spinor basis may be chosen as

$$f_1 = P_1, \quad f_2 = e^1P, \quad f_3 = e^2P, \quad f_4 = e^1e^2P. \quad (22)$$

If we define $\chi^+ \equiv \chi_1 \tilde{f}_1$ and $\chi^- \equiv \chi_2 \tilde{f}_2$, then a separable spinor solution of the form

$$\psi = [w^+\chi^+ + w^-\chi^-] \quad (23)$$

will be sought, where

$$w^\pm \equiv f^\pm(r,t) + g^\pm(r,t)e^1, \quad (24)$$

and $f^\pm$ and $g^\pm$ are some complex functions. It is straightforward to calculate $\not{\partial}\psi - M\psi$ and, using the fact that $\mathscr{I}\chi^\pm = i(ke^{-\lambda}/H_2)\chi^\mp$, we find consistency if $w^+ = -ie^1w^-$. This reduces the Dirac equation to the following pair of coupled partial differential equations:

$$\partial_r g^- + J_+ g^- + e^\lambda H_1\left[\frac{i}{H_0}\partial_t f^- - Mf^- + \frac{3}{2}\frac{i\dot{\lambda}}{H_0}f^-\right] = 0, \quad (25)$$

$$\partial_r f^- + J_- f^- - e^\lambda H_1\left[\frac{i}{H_0}\partial_t g^- + Mg^- + \frac{3}{2}\frac{i\dot{\lambda}}{H_0}g^-\right] = 0, \quad (26)$$

where

$$J_\pm \equiv \frac{H_0'}{2H_0} + \frac{H_2'}{H_2} \pm \frac{kH_1}{H_2}. \quad (27)$$

The above equations simplify to

$$\partial_r G + J_+ G + H_1 e^\lambda[(i/H_0)\partial_t F - MF] = 0, \quad (28)$$

$$\partial_r F + J_- F - H_1 e^\lambda[(i/H_0)\partial_t G + MG] = 0, \quad (29)$$

if we introduce

$$F(r,t) \equiv f^-(r,t)e^{(3/2)\lambda}, \quad (30)$$

$$G(r,t) \equiv g^-(r,t)e^{(3/2)\lambda}. \quad (31)$$

To effect a further separation, we write

$$F(r,t) = \tilde{F}(r)\alpha(t), \quad (32)$$

$$G(r,t) = \tilde{G}(r)\beta(t), \quad (33)$$

to obtain

$$\frac{\partial_r \tilde{G}}{H_1} + \frac{J_+\tilde{G}}{H_1} + \tilde{F}\left[\frac{ie^\lambda}{H_0}\frac{\dot{\alpha}}{\beta} - Me^\lambda\frac{\alpha}{\beta}\right] = 0, \quad (34)$$

$$\frac{\partial_r \tilde{F}}{H_1} + \frac{J_-\tilde{F}}{H_1} - \tilde{G}\left[\frac{ie^\lambda}{H_0}\frac{\dot{\beta}}{\alpha} + Me^\lambda\frac{\beta}{\alpha}\right] = 0. \quad (35)$$

These equations can be decoupled if the metric function H_0 is any constant (say 1) and we obtain, for separation constants c_1 and c_2,

$$\partial_r \tilde{G} + J_+\tilde{G} + H_1\tilde{F}c_1 = 0, \quad (36)$$

$$\partial_r \tilde{F} + J_-\tilde{F} - H_1\tilde{G}c_2 = 0, \quad (37)$$

$$i\dot{\alpha} - M\alpha = c_1 e^{-\lambda}\beta, \quad (38)$$

$$i\dot{\beta} + M\beta = c_2 e^{-\lambda}\alpha. \quad (39)$$

Thus $\tilde{F}(r)$, $\tilde{G}(r)$, $\alpha(t)$, and $\beta(t)$ also satisfy the ordinary differential equations

$$\partial_r^2\tilde{G} + (J_+ + J_- - (H_1'/H_1))\partial_r\tilde{G} + [c_1c_2H_1^2 + J_+J_- + \partial_r J_+ - (J_+H_1'/H_1)]\tilde{G} = 0, \quad (40)$$

$$\partial_r^2\tilde{F} + (J_+ + J_- - (H_1'/H_1))\partial_r\tilde{F} + [c_1c_2H_1^2 + J_+J_- + \partial_r J_- - (J_-H_1'/H_1)]\tilde{F} = 0, \quad (41)$$

$$\ddot{\alpha} + \dot{\lambda}\dot{\alpha} + [c_1c_2e^{-2\lambda} + M^2 + iM\dot{\lambda}]\alpha = 0, \quad (42)$$

$$\ddot{\beta} + \dot{\lambda}\dot{\beta} + [c_1c_2e^{-2\lambda} + M^2 - iM\dot{\lambda}]\beta = 0. \quad (43)$$

Note that in flat space-time, $\tilde{F}$ and $\tilde{G}$ are expressible in terms of spherical Bessel functions with $\alpha = \beta = e^{-i\omega t}$ and $c_1 = \omega - M$, $c_2 = \omega + M$. The Minkowski space spinor in a polar spinor basis may be represented in the form

$$\psi = e^{i\omega t}[\tilde{F}(r) + \tilde{G}(r)dr]\mathscr{S}_k^m T, \quad (44)$$

where

$$\mathscr{S}_k^m \equiv \chi^- - i\,dr\,\chi^+ \quad (45)$$

is a "solid spinor harmonic."

Returning to the general case with $H_0 \neq 1$, but seeking solutions of the form

$$G(r,t) = \tilde{G}(r)e^{-i\omega t}, \quad (46)$$

$$F(r,t) = \tilde{F}(r)e^{-i\omega t}, \quad \omega \text{ a real constant}, \quad (47)$$

we obtain consistency only if λ is any constant (say 0), in which case

$$\partial_r\tilde{G} + J_+\tilde{G} + H_1[\omega/H_0 - M]\tilde{F} = 0, \quad \partial_r\tilde{F} + J_-\tilde{F} - H_1[\omega/H_0 + M]\tilde{G} = 0. \quad (48)$$

Thus

$$\partial_r^2\tilde{G} + (J_+ + J_- - (a'/a))\tilde{G} + (J_+J_- - J'_+ - (a'/a)J_+ - ab)\tilde{G} = 0,$$

$$\partial_r^2\tilde{F} + (J_+ + J_- - (b'/b))\tilde{F} + (J_+J_- - J'_- - (b'/b)J_- - ab)\tilde{F} = 0, \quad (49)$$

where

$$a \equiv H_1[M - \omega/H_0], \quad b \equiv H_1[M + \omega/H_0]. \quad (49')$$

This case includes Schwarzschild and Reissner–Nordstrøm backgrounds.

IV. SEPARATION OF THE KÄHLER EQUATION IN SPHERICALLY SYMMETRIC SPACE-TIMES

To apply the same methodology to separate solutions of $\not{d}\Phi = M\Phi$, where $\not{d}\Phi = e^a\nabla_{X_a}\Phi$, we must first find the appropriate basis of angular solutions analogous to the spinor harmonics discussed above. First note that if Y is any zero-form independent of r and t on a space-time with metric (9) in the (r,t,θ,ϕ)-chart, then

$$d\!\!\!/^2 Y = \frac{e^{-2\lambda}}{H_2^2}\left[\frac{1}{\sin\theta}\partial_\theta(\sin\theta\,\partial_\theta Y) + \frac{1}{\sin^2\theta}\partial_\phi^2 Y\right]. \quad (50)$$

Here and in the following it is useful to note the relations

$$e^a u = \eta u e^a + 2 i^a u, \quad (51)$$

$$d\!\!\!/(uv) = d\!\!\!/ u v + \eta u\, d\!\!\!/ v + 2 i^a u \nabla_{X_a} v, \quad (52)$$

for any inhomogeneous forms u and v. In these formulas, if $u = \Sigma_{p=0}^4 u_p$, where u_p is a p-form, $u_p \in \Gamma(\Lambda^p(M))$, then

$$\eta u \equiv \sum_{p=0}^{4}(-1)^p u_p, \quad (53)$$

$$i^a u \equiv g^{ab}\, i_{X_b} u. \quad (54)$$

It follows from (50) that $d\!\!\!/^2 Y_k^m = c(r,t) Y_k^m$, where Y_k^m is a standard spherical harmonic and

$$c(r,t) \equiv -k(k+1) e^{-2\lambda}/H_2^2, \quad k = 0,1,2,\ldots. \quad (55)$$

In the metric (9) we readily observe that

$$\begin{aligned} d\!\!\!/ e^0 &= \sigma(r,t) e^0 e^1 - 3\dot\lambda/H_0, \\ d\!\!\!/ e^1 &= (\dot\lambda/H_0) e^0 e^1 + \rho(r,t), \\ d\!\!\!/(e^0 e^1) &= \gamma(r,t) e^0 - (2\dot\lambda/H_0) e^1, \end{aligned} \quad (56)$$

where

$$\begin{aligned} \sigma(r,t) &\equiv -H_0' e^{-\lambda}/H_0 H_1 \\ \gamma(r,t) &\equiv -2H_2' e^{-\lambda}/H_1 H_2, \\ \rho(r,t) &\equiv -(\sigma+\gamma). \end{aligned} \quad (57)$$

Thus if for each k, u and w are inhomogeneous forms of the type

$$\begin{aligned} u &= f(r,t) + g(r,t) e^1 + p(r,t) e^0 + q(r,t) e^0 e^1, \\ w &= F(r,t) + G(r,t) e^1 + P(r,t) e^0 + Q(r,t) e^0 e^1, \end{aligned} \quad (58)$$

where f, g, p, q, F, G, P, and Q are complex functions, then the form

$$\Phi = u\, dY_k^m + w Y_k^m \quad (59)$$

is a Kähler solution provided

$$\begin{aligned} d\!\!\!/ u + \frac{2\dot\lambda}{H_0} i_{X_0} u - \frac{2H_2' e^{-\lambda}}{H_1 H_2} i_{X_1} u + \eta w &= M u, \\ d\!\!\!/ w + c\eta u &= M w, \end{aligned} \quad (60)$$

and we have reduced the problem to coupled equations for u and w. [In deriving (60) we note that since Y_k^m is a function only θ and ϕ, we have $i_{\widetilde{dY}}\, w = 0$, where $\widetilde{dY}$ is the metric dual of dY, $\widetilde{dY} = g(dY,\cdot)$. This implies that $dY\, w = \eta w\, dY$.] We shall not present a detailed analysis of these equations here, but discuss only solutions in the static metric where $e^\lambda = 1$ and

$$\begin{aligned} u &= \tau(t)[f(r) + e^1 g(r)], \\ w &= \tau(t)[F(r) + e^1 G(r)]. \end{aligned} \quad (61)$$

From (60), static solutions $[\tau(t) = 1]$ exist satisfying

$$\begin{aligned} &(1/H_1)\partial_r f(r) - G(r) = M g(r), \\ &(1/H_1)\partial_r g(r) - \sigma g(r) + F(r) = M f(r), \\ &(1/H_1)\partial_r F(r) - c(r) g(r) = M G(r), \\ &(1/H_1)\partial_r G(r) + \rho G(r) + c(r) f(r) = M f(r). \end{aligned} \quad (62)$$

These equations readily decouple, although it is simpler to write them

$$\begin{aligned} &\partial_r((1/H_1)\partial_r f(r)) + \rho\partial_r f(r) \\ &\quad + (c - M^2) H_1 f(r) = -M\gamma H_1 g(r), \\ &\partial_r((1/H_1)\partial_r g(r)) - \partial_r(\sigma g(r)) \\ &\quad + (c - M^2) H_1 g(r) = 0, \\ &\partial_r\left(\frac{1}{H_1 c}\partial_r G(r)\right) + \partial_r\left(\frac{\rho G(r)}{c}\right) \\ &\quad + (c - M^2)\frac{H_1}{c} G(r) = -M\gamma H_1 \frac{F(r)}{c}, \\ &\partial_r\left(\frac{1}{H_1 c}\partial_r F(r)\right) - \frac{\sigma}{c}\partial_r F(r) + (c - M^2)\frac{H_1}{c} F(r) = 0. \end{aligned} \quad (63)$$

Of particular interest is the Minkowski space, $H_0 = H_1 = 1$, $H_2 = r$ $(0 \leqslant r < \infty)$, with $\tau(t) = e^{i\omega t}$. The solution to Kähler's equation can then be written as

$$\Phi = e^{i\omega t}[A(r) + B(r) e^1] S_k^m T, \quad (64)$$

where

$$T \equiv \tfrac{1}{2}(1 + i e^0), \quad (65)$$

$$S_k^m = k e^1 Y_k^m + r\, dY_k^m = r^{1-k} d(r^k Y_k^m), \quad (66)$$

and A and B satisfy

$$\partial_r^2 A + \frac{2}{r}\partial_r A + \left[\omega^2 - M^2 + \frac{(-k)(-k+1)}{r^2}\right] A = 0, \quad (67)$$

$$\partial_r^2 B + \frac{2}{r}\partial_r B + \left[\omega^2 - M^2 + \frac{k(k+1)}{r^2}\right] B = 0.$$

These equations coincide with the equations defining the r dependence of the components of a Dirac spinor satisfying the Dirac equation in flat space. Indeed, although Φ above does not lie in any Minkowski space minimal ideal, it is trivial to modify it to obtain a solution that does. Let (t, x^1, x^2, x^3) be the standard inertial coordinates for Minkowski space in terms of which the metric is

$$g = -dt \otimes dt + \sum_{i=1}^{3} dx^i \otimes dx^i, \quad (68)$$

and

$$\begin{aligned} x^1 &= r\sin\theta\cos\phi, \\ x^2 &= r\sin\theta\sin\phi, \\ x^3 &= r\cos\theta. \end{aligned} \quad (69)$$

Since $\mathscr{Y} \equiv \tfrac{1}{2}(1 + i\, dx^1 dx^2)$ is ∇-parallel in Minkowski space, $\nabla_X \mathscr{Y} = 0$, $X \in \Gamma(TM)$, then $\hat\Phi = \Phi\mathscr{Y}$ is a spinorial Kähler solution $\hat\Phi = \hat\Phi T\mathscr{Y}$. The components of this spinor are in a polar Minkowski chart but are defined with respect to a parallel spinor basis generated by the primitive idempotent $T\mathscr{Y}$. The element $S \equiv \exp(-\tfrac{1}{2}\phi e^2 e^3)\exp(-\tfrac{1}{2}\theta e^1 e^2)$ induces an orthogonal transformation that relates the coframe $\{e^1, e^2, e^3\}$ to $\{dx^1, dx^2, dx^3\}$:

$$S\, dx^1 S^{-1} = e^2, \quad S\, dx^2 S^{-1} = e^3, \quad S\, dx^3 S^{-1} = e^1, \quad (70)$$

and transforms the primitive $\mathscr{Y} T$ to

$$S\mathscr{Y} T S^{-1} = \tfrac{1}{4}(1 + i e^2 e^3)(1 + i e^0). \quad (71)$$

The transformed spinor $\psi = \hat{\Phi} S^{-1}$ can be written

$$\psi = e^{i\omega t}[A(r) + B(r)e^1]\{S_k^m(1 + i\,dx^1\,dx^2)S^{-1}\}T. \quad (72)$$

It is not difficult to check that the term in the curly bracket is proportional to the solid spinor harmonic introduced in (45). Both ψ above and $\hat{\Phi}$ solve the Dirac equation $\not{S}\xi = M\xi$; the former when $\not{S}$ is defined with respect to a primitive $\mathscr{Y}T$, the latter when $\not{S}$ is defined with respect to $S\mathscr{Y}TS^{-1}$. It is, however, important to note that it is only the algebraic spinor $\hat{\Phi}$ that solves the Kähler equation, since this equation lacks the spinor covariance of the Dirac equation.

V. DISCUSSION

We have shown how, in a class of spherically symmetric space-times, to find a class of solutions of the Dirac equation $\not{S}\psi = M\psi$ of the form

$$\psi = \sum_{m,k} a_{m,k}\psi_k^m, \quad (73)$$

where

$$\psi_k^m = w_k \mathscr{S}_k^m T. \quad (73')$$

Similar techniques have enabled us to find a class of solutions of the Kähler equation $\not{d}\Phi = M\Phi$,

$$\Phi = \sum_{m,k} b_{m,k}\Phi_k^m, \quad (74)$$

where

$$\Phi = (1/r)u_k S_k^m. \quad (74')$$

In the spinor case, the spinor solid harmonics may be related to standard spherical harmonics $Y_k^m(\theta,\phi)$ satisfying

$$\frac{1}{\sin\theta}\partial_\theta[\sin\theta\,\partial_\theta Y_k^m] + \frac{1}{\sin\theta}\partial_\phi^2 Y_k^m + k(k+1) = 0,$$
$$k = 0,1,2,\ldots, \quad -k<m<k, \quad (75)$$

by the formula

$$\mathscr{S}_k^m = \overset{m}{\chi}{}_k^- - ie^1\overset{m}{\chi}{}_k^+, \quad (76)$$

where

$$\overset{m}{\chi}{}_k^+ = ie^{-i\phi/2}(kY_k^m F_1 - e^\lambda H_2\,dY_k^m\,F_2),$$
$$\overset{m}{\chi}{}_k^- = e^{-i\phi/2}(kY_k^m F_2 + e^\lambda H_2\,dY_k^m\,F_1), \quad (77)$$

and

$$F_1 \equiv \cos(\theta/2)\tilde{f}_1, \quad F_2 \equiv \sin(\theta/2)\tilde{f}_2. \quad (78)$$

The solid angular forms S_k^m that appear in (74) have similar representations:

$$S_k^m = ru_k^{-1}w_k Y_k^m + r\,dY_k^m, \quad (79)$$

and coincide with (66) in Minkowski space-time. Using (76) and (77) the components $(\chi_1)_k^m$, $(\chi_2)_k^m$ of $\overset{m}{\chi}{}_k^+$ and $\overset{m}{\chi}{}_k^-$ in the basis $\{\tilde{f}_1, \tilde{f}_2\}$ may be written in terms of standard spin-weighted spherical harmonics[7] of spin weight $-\frac{1}{2}$ and $\frac{1}{2}$:

$$\overset{m}{\chi}{}_k^+ = (\chi_1)_k^m\tilde{f}_1, \quad \overset{m}{\chi}{}_k^- = (\chi_2)_k^m\tilde{f}_2, \quad (80)$$

$$i(\chi_1)_k^m \equiv {}_{-1/2}Y_{j,m_j} = C_{j,m_j}\Omega_0' Y_{j+1/2}^{m_j+1/2},$$
$$(\chi_2)_k^m \equiv {}_{1/2}Y_{j,m_j} = C_{j,m_j}\Omega_0 Y_{j+1/2}^{m_j+1/2}, \quad (81)$$
$$j = \tfrac{1}{2},\tfrac{3}{2},\ldots, \quad m_j = -j-1, -j,\ldots, j,$$

where

$$\Omega_0 \equiv e^{-i\phi/2}\left\{\cos\frac{\theta}{2}\left(\partial_\theta - \frac{i}{\sin\theta}\partial_\phi\right) + \left(j+\frac{1}{2}\right)\sin\frac{\theta}{2}\right\},$$
$$(82)$$
$$\Omega_0' \equiv e^{-i\phi/2}\left\{\sin\frac{\theta}{2}\left(\partial_\theta + \frac{i}{\sin\theta}\partial_\phi\right) - \left(j+\frac{1}{2}\right)\cos\frac{\theta}{2}\right\}.$$

The constants

$$C_{j,m_j} \equiv -1/\sqrt{(j+1)(j+1+m_j)} \quad (83)$$

insure the normalization on S^2

$$\int_{S^2}({}_sY_{j,m_j})({}_sY^*_{j',m_j'})d\Omega = \delta_{jj'}\delta_{m_jm_j'};$$
$$s = \pm\tfrac{1}{2}. \quad (84)$$

One may regard Ω_0 and Ω_0' as operators that act on a function of spin weight zero to produce a function of spin weight $\frac{1}{2}$ higher and lower, respectively. If we introduce the anti-involution ξ by $\xi(\alpha\beta) = \xi\beta\xi\alpha$, where

$$\xi\left(\sum_p \alpha_p\right) = \sum_p(-1)^{[p/2]}\alpha_p,$$

for any Clifford sections α, β, and $\alpha_p \in \Gamma(\Lambda^p(M))$, then the adjoint spinor $\bar{\psi}$ is defined to be $\bar{\psi} = b^{-1}\xi\psi^*$ for some element b that satisfies $\xi P^* = bPb^{-1}$. For the choice of projector in Sec. III we have $b = e^1$. Let $S_3(U)$ denote the three-form part of a general inhomogeneous form U, then for any solution ψ to Eq. (3) the real current three-form

$$J = (4/i)S_3(\psi\bar{\psi}) \quad (85)$$

is closed, $dJ = 0$. Hence if Σ denotes a spacelike hypersurface on which e^0 has no component and J vanishes on $\partial\Sigma$, then writing $\psi = \Sigma_{i=1}^4\psi_i f_i$ yields

$$\int_\Sigma J = \sum_{i=1}^4\int_\Sigma(\psi_i\psi_i^*)e^1\wedge e^2\wedge e^3, \quad (86)$$

which is independent of Σ. This result may be used to construct a normalizable basis of spinor solutions on space-time. The solutions ψ_k^m (73) and Φ_k^m (74) share an important property. They are simultaneous eigenforms of appropriate Lie derivatives associated with the rotational Killing vector fields

$$K_1 = -\sin\theta\,\partial_\theta - \cot\theta\cos\phi\,\partial_\phi,$$
$$K_2 = \cos\phi\,\partial_\theta - \cot\theta\sin\phi\,\partial_\phi, \quad (87)$$
$$K_3 = \partial_\phi.$$

For the Kähler case (74)

$$L^2\Phi_k^m = -k(k+1)\Phi_k^m,$$
$$\mathscr{L}_{K_3}\Phi_k^m = im\Phi_k^m, \quad (88)$$

where $L^2 = \Sigma_{i=1}^3\mathscr{L}_{K_i}\mathscr{L}_{K_i}$ and $\mathscr{L}_K$ is the standard Lie derivative on forms with respect to the local Killing vector field K.

For a space-time spinor ψ we define

$$\mathscr{L}_K\psi = S_K\psi + \tfrac{1}{4}d\tilde{K}\,\psi, \quad (89)$$

where $\tilde{K}$ is the metric dual of any Killing vector field K.

The definition (89) ensures that for any two Killing vectors K and K'

$$[\mathscr{L}_K, \mathscr{L}_{K'}] = \mathscr{L}_{[K,K']}. \tag{90}$$

In terms of Killing Lie derivatives on spinors we find

$$Ł^2 \psi_k^m = -(k+\tfrac{1}{2})(k+\tfrac{3}{2})\psi_k^m,$$

$$\mathscr{L}_{K_3}\psi_k^m = i(m+\tfrac{1}{2})\psi_k^m,$$

where

$$Ł^2 \equiv \sum_{i=1}^{3} \mathscr{L}_{K_i}\mathscr{L}_{K_i}. \tag{91}$$

It is generally recognized[11] that the existence of a coordinate system that renders the Laplace–Beltrami equation for a p-form

$$\not{d}^2\alpha = M\alpha, \quad \alpha \in \Lambda^p(M) \tag{92}$$

separable, depends on its relation to the Killing symmetries of the underlying metric. It is one of the main results of this paper that techniques similar to those separating the second-order tensor equation (92) can be used to analyze the first-order Kähler equation $\not{d}\Phi = M\Phi$ as well as the spinor equation $\not{S}\psi = M\psi$. In both cases we find that the angular dependence of solutions can be split off as simultaneous eigenforms of L^2 and $\mathscr{L}_{K_3}$ or $Ł^2$ and $\mathscr{L}_{K_3}$. This approach may be contrasted with the method in Ref. 12. The solutions introduced in this paper have been used to illustrate the differences and similarities between solutions of the Dirac and Kähler equations in spherically symmetric space-times. They also emphasize the manner in which spinor wave equations for spinor sections of a Clifford bundle may be analyzed using techniques of differential geometry.

[1]N. Salingaros and M. Dresden, Phys. Rev. Lett. **43**, 1 (1979); P. Becher and H. Joos, Z. Phys. C. **15**, 343 (1982); J. M. Rabin, Nucl. Phys. B **210**, 315 (1982); T. Banks, Y. Dothan, and D. Horn, Phys. Lett. B **117**, 413 (1982); F. Gliozzi, Nucl. Phys. B **204**, 419 (1982); N. Salingaros, J. Math. Phys. **22**, 226, 1919 (1982); **23**, 1 (1982); P. Basarab-Horwath, and R. W. Tucker, C. R. Acad. Sci. Ser. A **20**, 299 (1984); "A quantisation for Kähler fields in static space-time," Lancaster University preprint, 1985; I. M. Benn and R. W. Tucker, Phys. Lett. B **130**, 177 (1983); **125**, 47 (1983); J. Phys. A **16**, 4147 (1983); Phys. Lett. B **132**, 325 (1984); Nuovo Cimento A **88**, 273 (1984); J. Phys. A: Math. Gen. **16**, 4123 (1983); T. Dereli, M. Onder, and R. W. Tucker, Class. Quantum Gravit. **1**, L67 (1984); A. Al-Saad and I. M. Benn, "Ideal preserving Lorentzian connections," Lancaster University Preprint (1985); I. M. Benn, B. Dolan, and R. W. Tucker, Phys. Lett. B **105**, 100 (1985); I. M. Benn and R. W. Tucker, Phys. Lett. B **119**, 348 (1982); "Pure spinors and real Clifford algebras," Lancaster University preprint (1984); I. M. Benn, M. Panahi, and R. W. Tucker, Class. Quantum Gravit. **2**, L71, L109 (1985); P. Budinich, and K. Bugajska, J. Math. Phys. **26**, 588 (1985); P. Budinich and L. Dabrowski, Lett. Math. Phys. **10**, 7 (1985).

[2]R. Ablamowicz, Z. Oziewicz, and J. Rzewski, J. Math. Phys. **23**, 231 (1982); R. Ablamowicz and N. Salingaros, Lett. Math. Phys. **9**, 149 (1985).

[3]C. Chevalley, *The Algebraic Theory of Spinors* (Columbia U.P., New York, 1954).

[4]I. M. Benn and R. W. Tucker, Commun. Math. Phys. **98**, 53 (1985).

[5]I. M. Benn and R. W. Tucker, "Clifford analysis of wave equations," in *Proceedings of the Colloquium on Differential Geometry* (Bolyai Math. Soc., Hajduszoboszlo, Hungary, 1984).

[6]W. Graf, Ann. Inst. H. Poincaré **XXIX**, 85 (1978). I. M. Benn and R. W. Tucker, Commun. Math. Phys. **89**, 341 (1983).

[7]E. Kähler, Rend. Mat. **21**, 425 (1962).

[8]E. Schrödinger, Proc. R. Ir. Acad. Sect. A **XLVI**, 25 (1940); S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., London, 1983).

[9]R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge U.P., Cambridge, England, 1984); also references therein.

[10]S. K. Chakrabati, Proc. R. Soc. London **391**, 27 (1984).

[11]W. Miller, "Symmetry and separation of variables," in *Encyclopedia of Mathematics and its Applications*, Vol. 4 (Addison-Wesley, Reading, MA, 1977).

[12]B. R. Iyer and C. V. Vishveshwara, J. Math. Phys. **26**, 1034 (1985).

Symmetries of the massive Thirring model

H. M. M. Ten Eikelder
Department of Mathematics and Computing Science, Eindhoven University of Technology, P. O. Box 513, Eindhoven, The Netherlands

(Received 26 September 1985; accepted for publication 20 November 1985)

For a Hamiltonian system every non-Hamiltonian symmetry gives rise to a recursion operator for symmetries. Using this method two recursion operators for symmetries of the massive Thirring model are constructed. The structure of the Lie algebra of symmetries generated by these operators is given.

I. INTRODUCTION

The existence of infinite series of symmetries is a very special property of a dynamical or Hamiltonian system. These series are often constructed by using a recursion operator for symmetries (also called Lénard operator, or strong symmetry or squared eigenfunctions operator). In Sec. II of this paper we make some general remarks on symmetries and tensor symmetries of a dynamical system. In particular, we show that for a Hamiltonian system every non-Hamiltonian symmetry gives rise to a recursion operator for symmetries. This method is applied to the massive Thirring model in Sec. III. Using two symmetries found by Kersten[1] and Kersten and Martini,[2,3] we construct two recursion operators for symmetries of the massive Thirring model. These operators turn out to be each others' inverses. With these recursion operators we generate two infinite series of symmetries. One of these series corresponds to an infinite series of constants of the motion in involution. The other series consists of non-Hamiltonian symmetries. The corresponding Lie algebra of symmetries is also described. In Secs. II and III we use the framework of differential geometry. In Appendix A we show how the, at first instance finite-dimensional, differential geometry can be introduced on the topological vector space in which the Thirring model is studied. Some long expressions are given in Appendix B. Similar results as given in this paper for the massive Thirring model can be obtained for several other equations, see Ten Eikelder.[4]

We now make some remarks on the notation and terminology. A tensor field with contravariant order p and covariant order q will be called a (p,q) tensor field. The set of vector fields [$=(1,0)$ tensor fields] and the set of one-forms [$=(0,1)$ tensor fields] on a manifold $\mathscr{M}$ will be denoted by $\mathscr{X}(\mathscr{M})$ [resp. $\mathscr{X}^*(\mathscr{M})$]. The contraction between a one-form α and a vector field A will be written as $\langle \alpha, A\rangle$. The Lie derivative in the direction of a vector field A will be denoted as $\mathscr{L}_A$. Applied to a vector field B this Lie derivative equals the Lie bracket $[A,B]$, i.e., $\mathscr{L}_A B = [A,B]$. Further we use the operators $\partial = \partial/\partial x$ and ∂^{-1}, defined by

$$(\partial^{-1} f)(x) = \int_{-\infty}^{x} f(y)dy - \frac{1}{2}\int_{-\infty}^{\infty} f(y)dy.$$

Then ∂ and ∂^{-1} are both skew symmetric with respect to the L_2 inner product. These operators are assumed to act on everything that follows them, except when otherwise indicated.

II. TENSOR SYMMETRIES OF A DYNAMICAL SYSTEM

In this section we make some general remarks on symmetries of dynamical and Hamiltonian systems. Let X be a vector field on a manifold $\mathscr{M}$. With X the following dynamical system is associated:

$$\dot u(t) = X(u(t)) \quad \left(\dot u(t) = \frac{d}{dt}u(t)\right). \tag{2.1}$$

A, possibly t-parametrized, tensor field Ξ on $\mathscr{M}$, which satisfies

$$\dot\Xi + \mathscr{L}_X \Xi = 0 \quad \left(\dot\Xi = \frac{\partial}{\partial t}\Xi, \quad t\in\mathbb{R}\right) \tag{2.2}$$

on $\mathscr{M}$, will be called a tensor symmetry of (2.1). It follows from Leibniz' rule that the tensor product of two tensor symmetries is again a tensor symmetry. Also every possible contraction in a tensor symmetry (or contracted multiplication of two tensor symmetries) yields again a tensor symmetry. If Ξ is a completely skew-symmetric $(0,p)$ tensor field (i.e., a differential p-form), then a new tensor symmetry can be constructed by exterior differentiation.

A tensor symmetry of type (0,0) (i.e., a function) is called a constant of the motion or first integral. A tensor symmetry of type (1,0) (i.e., a vector field on $\mathscr{M}$) will be called a symmetry. Finally a tensor symmetry of type (1,1) will be called a recursion operator for symmetries.

Let Z be a symmetry and Ξ be an arbitrary tensor symmetry. Then

$$\begin{aligned}\mathscr{L}_X\mathscr{L}_Z\Xi + \frac{\partial}{\partial t}\mathscr{L}_Z\Xi &= \mathscr{L}_Z\mathscr{L}_X\Xi + \mathscr{L}_{[X,Z]}\Xi + \mathscr{L}_{\dot Z}\Xi + \mathscr{L}_Z\dot\Xi\\ &= \mathscr{L}_Z(\mathscr{L}_X\Xi + \dot\Xi) + \mathscr{L}_{[X,Z]+\dot Z}\Xi = 0.\end{aligned}$$

So the Lie derivative of a tensor symmetry in the direction of a symmetry yields again a tensor symmetry (of the same type as the original one).

Suppose Λ is a (1,1) tensor field and Φ and Ψ are skew symmetric (0,2) [resp. (2,0)] tensor fields. With these tensor fields the following linear mappings are associated:

$$\hat\Lambda:\mathscr{X}(\mathscr{M})\to\mathscr{X}(\mathscr{M}),$$
$$\hat\Phi:\mathscr{X}(\mathscr{M})\to\mathscr{X}^*(\mathscr{M}),$$
$$\hat\Psi:\mathscr{X}^*(\mathscr{M})\to\mathscr{X}(\mathscr{M}).$$

To simplify the notation we shall drop the hat and identify the tensor fields with the corresponding mappings (see also Appendix A). This also enables us to speak of the Lie deriva-

 0022-2488/86/051404-07$02.50

tive of such a mapping. In particular a two-form [= skew-symmetric (0,2) tensor field] Ω is identified with a skew-symmetric mapping $\Omega:\mathscr{X}(\mathscr{M})\to\mathscr{X}^*(\mathscr{M})$. If the two-form is nondegenerate this mapping has an inverse $\Omega^{\leftarrow}:\mathscr{X}^*(\mathscr{M})\to\mathscr{X}(\mathscr{M})$.

Now suppose that X is a Hamiltonian vector field, i.e., there exist a Hamiltonian H and a symplectic form Ω on $\mathscr{M}$ such that

$$X=\Omega^{\leftarrow}\, dH. \tag{2.3}$$

The closedness of Ω implies that $\mathscr{L}_X\Omega = d(\Omega X) = d\,dH = 0$. Since $\dot{\Omega}=0$ this means that Ω is a tensor symmetry of type (0,2). From $\Omega^{\leftarrow}\Omega = I$ we obtain that $\Omega^{\leftarrow}$ is a tensor symmetry of type (2,0). Suppose that F is a constant of the motion. Then the one-form dF is a tensor symmetry of type (0,1) and $Y=\Omega^{\leftarrow}\, dF$ is a tensor symmetry of type (1,0), i.e., a symmetry. So every constant of the motion gives rise to a symmetry. Note that all symmetries obtained in this way are (possibly t-parametrized) Hamiltonian vector fields on $\mathscr{M}$.

Let Z be a symmetry. Then $\mathscr{L}_Z\Omega$ is a tensor symmetry of type (0,2). The contracted multiplication of the tensor symmetries $\Omega^{\leftarrow}$ and $\mathscr{L}_Z\Omega$ [in terms of mappings: the composition of $\mathscr{L}_Z\Omega$: $\mathscr{X}(\mathscr{M})\to\mathscr{X}^*(\mathscr{M})$ and $\Omega^{\leftarrow}$: $\mathscr{X}^*(\mathscr{M})\to\mathscr{X}(\mathscr{M})$] is a tensor symmetry of type (1,1). So for every symmetry Z,

$$\Lambda=\Omega^{\leftarrow}\mathscr{L}_Z\Omega \tag{2.4}$$

is a recursion operator for symmetries. Since Ω is closed we have $\mathscr{L}_Z\Omega = d(\Omega Z)$. So if Z is a Hamiltonian vector field we obtain by (2.4) the trivial recursion operator $\Lambda = 0$. Only in the case where Z is a non-Hamiltonian symmetry (i.e., a symmetry with ΩZ not closed), we obtain by (2.4) a nonvanishing recursion operator for symmetries. So every non-Hamiltonian symmetry of a Hamiltonian system gives rise to a recursion operator for symmetries.

If a system has a recursion operator for symmetries Λ, an infinite series of symmetries can be constructed by repeated application of this recusion operator to some symmetry. An important concept for understanding the algebra of symmetries generated in this way is the Nijenhuis tensor of Λ (see Nijenhuis[5] and Schouten[6]). With every (1,1) tensor field Λ is associated a (1,2) tensor field N_Λ, called the Nijenhuis tensor field of Λ, such that for all vector fields A,

$$\mathscr{L}_{\Lambda A}\Lambda - \Lambda\mathscr{L}_A\Lambda = N_\Lambda A. \tag{2.5}$$

The right-hand side of this expression is the contracted multiplication of the (1,2) tensor field N_Λ and the vector field A. This results again in a (1,1) tensor field. The importance of recursion operators for symmetries with a vanishing Nijenhuis tensor field has already been noticed by Magri,[7] Fuchssteiner,[8] Fuchssteiner and Fokas,[9] and Gel'fand and Dorfman.[10] It is easily seen how this property can be used. Let A and B be vector fields such that $\mathscr{L}_A\Lambda = a\Lambda$ and $\mathscr{L}_B\Lambda = b\Lambda$ for $a,b\in\mathbb{R}$. Define $A_k=\Lambda^kA$ and $B_k=\Lambda^kB$ for $k=0,1,2,\dots$. Then

$$\begin{aligned}[A_k,B_l] &= \mathscr{L}_{A_k}(\Lambda^lB) = (\mathscr{L}_{A_k}\Lambda^l)B + \Lambda^l\mathscr{L}_{A_k}B\\ &= (\mathscr{L}_{A_k}\Lambda^l)B - \Lambda^l\mathscr{L}_B(\Lambda^kA)\\ &= (\mathscr{L}_{A_k}\Lambda^l)B - \Lambda^l(\mathscr{L}_B\Lambda^k)A\\ &\quad + \Lambda^{k+l}[A,B]\\ &= (\mathscr{L}_{A_k}\Lambda^l)B - k\Lambda^{k+l}bA + \Lambda^{k+l}[A,B].\end{aligned} \tag{2.6}$$

If the Nijenhuis tensor field of Λ vanishes we have

$$\mathscr{L}_{A_k}\Lambda = \Lambda^k\mathscr{L}_A\Lambda = a\Lambda^{k+1}. \tag{2.7}$$

Substitution in (2.6) finally results in

$$[A_k,B_l] = laB_{k+l} - kbA_{k+l} + \Lambda^{k+l}[A,B]. \tag{2.8}$$

If Λ is invertible we can also define A_k and B_k, for $k=-1,-2,-3,\dots$. Using

$$\mathscr{L}_C\Lambda^{-1} = -\Lambda^{-1}(\mathscr{L}_C\Lambda)\Lambda^{-1}$$

for every vector field C it is easily shown that in this case (2.7) and (2.8) also hold for negative integers k and l.

III. RECURSION OPERATORS FOR SYMMETRIES OF THE MASSIVE THIRRING MODEL

The massive Thirring model is the following system of partial differential equations for the functions $u_1(x,t)$, $u_2(x,t)$, $v_1(x,t)$, and $v_2(x,t)$:

$$\begin{aligned} u_{1t} &= u_{1x} + mv_2 - R_2v_1,\\ u_{2t} &= -u_{2x} + mv_1 - R_1v_2,\\ v_{1t} &= v_{1x} - mu_2 + R_2u_1,\\ v_{2t} &= -v_{2x} - mu_1 + R_1u_2,\\ &-\infty < x < \infty,\quad t>0,\end{aligned} \tag{3.1}$$

where $R_1 = u_1^2 + v_1^2$ and $R_2 = u_2^2 + v_2^2$. We assume that u_1, u_2, v_1, and v_2 are smooth and, together with their x-derivatives, decay sufficiently fast for $|x|\to\infty$. We shall study (3.1) in some reflexive topological vector space $\mathscr{W}$, which is the Cartesian product of function spaces for u_1, u_2, v_1, and v_2. $\mathscr{W}$ and $\mathscr{W}^*$ are constructed in such a way that their duality map $\langle\cdot,\cdot\rangle$ is just the L_2 inner product. In terms of $u=(u_1,u_2,v_1,v_2)\in\mathscr{W}$ we can write (3.1) as

$$\dot{u} = X(u)\quad \left(\dot{u}=\frac{d}{dt}u\right). \tag{3.2}$$

The nonlinear mapping X can be considered as a vector field on $\mathscr{W}$. In this section we shall continue to use the differential geometrical language of Sec. II. For a definition of the various differential geometrical objects in this infinite-dimensional case see Appendix A.

Define the function (functional) H on $\mathscr{W}$ by

$$H=\int_{-\infty}^{\infty}\left(v_1u_{1x} - v_2u_{2x} + mR - \frac{1}{2}R_1R_2\right)dx,$$

where $R=u_1u_2+v_1v_2$. Moreover let the symplectic form Ω be (represented by the linear mapping $\Omega:\mathscr{W}\to\mathscr{W}^*$)

$$\Omega=\begin{pmatrix}0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{pmatrix}.$$

Then it is easily verified that

$$X=\Omega^{\leftarrow}\,dH, \tag{3.3}$$

i.e., the massive Thirring model is a Hamiltonian system with Hamiltonian H and symplectic form Ω.

Symmetries for the massive Thirring model have recently been studied by Kersten[1] and Kersten and Martini.[2,3] Amongst others they give the following symmetries:

$$X_0=\begin{pmatrix} v_1\\ v_2\\ -u_1\\ -u_2\end{pmatrix},$$

$$Z_0=\begin{pmatrix} -xu_{1x}-mv_2x-\tfrac12u_1+v_1R_2x\\ xu_{2x}-mv_1x+\tfrac12u_2+v_2R_1x\\ -xv_{1x}+mu_2x-\tfrac12v_1-u_1R_2x\\ xv_{2x}+mu_1x+\tfrac12v_2-u_2R_1x\end{pmatrix}-t\begin{pmatrix}u_{1x}\\ u_{2x}\\ v_{1x}\\ v_{2x}\end{pmatrix}, \tag{3.4}$$

$$Z_1=(1/m)p_2X_0-\tfrac12mx(X_2+X_0)+\tfrac12mt(X_2-X_0)+\frac1m\begin{pmatrix}\tfrac12mu_2\\ \tfrac32mu_1+3v_{2x}-\tfrac32R_1u_2-\tfrac12R_2u_2\\ \tfrac12mv_2\\ \tfrac32mv_1-3u_{2x}-\tfrac32R_1v_2-\tfrac12R_2v_2\end{pmatrix}, \tag{3.5}$$

$$\widehat{Z}_{-1}=(1/m)p_1X_0+\tfrac12mx(X_{-2}+X_0)+\tfrac12mt(X_{-2}-X_0)+\frac1m\begin{pmatrix}\tfrac32mu_2-3v_{1x}-\tfrac32R_2u_1-\tfrac12R_1u_1\\ \tfrac12mu_1\\ \tfrac32mv_2+3u_{1x}-\tfrac32R_2v_1-\tfrac12R_1v_1\\ \tfrac12mv_1\end{pmatrix}, \tag{3.6}$$

with $p_2=(\partial^{-1}(u_2v_{2x}-u_{2x}v_2-\tfrac12R_1R_2+mR))$ and $p_1=(\partial^{-1}(u_1v_{1x}-u_{1x}v_1+\tfrac12R_1R_2-mR))$. The expressions for X_2 and X_{-2} equal $4m^{-2}Y_5$ (resp. $-4m^{-2}Y_6$) in Refs. 1 — 3. These four symmetries have been found by Kersten as symmetries of a prolonged exterior differential system that describes the massive Thirring model. However, a straightforward computation shows that X_0,Z_0,Z_1, and $\widehat{Z}_{-1}$ are also symmetries of the type considered in this paper.

A simple computation shows that $X_0=\Omega^{\leftarrow}\,dF_0$ and $Z_0=\Omega^{\leftarrow}\,dG$, where the constants of the motion F_0 and G are given by

$$F_0=\frac12\int_{-\infty}^{\infty}(R_1+R_2)dx, \tag{3.7}$$

$$G=\int_{-\infty}^{\infty}\Big(x\Big(v_2u_{2x}-v_1u_{1x}-mR+\frac12R_1R_2\Big)-\frac12u_1v_1+\frac12u_2v_2-t(v_1u_{1x}+v_2u_{2x})\Big)dx. \tag{3.8}$$

More interesting results are obtained from Z_1 and $\widehat{Z}_{-1}$. These symmetries are non-Hamiltonian vector fields, so we can construct recursion operators for symmetries Λ_1 and Λ_{-1} by

$$\Lambda_1=\Omega^{\leftarrow}\mathscr{L}_{Z_1}\Omega,\quad \Lambda_{-1}=\Omega^{\leftarrow}\mathscr{L}_{\widehat{Z}_{-1}}\Omega. \tag{3.9}$$

The rather lengthy expressions for Λ_1 and Λ_{-1} found in this way are given in Appendix B. A tedious computation shows that

$$\Lambda_1\Lambda_{-1}=I, \tag{3.10}$$

where I is the identity (1,1) tensor field on $\mathscr{W}$ (i.e., the identity mapping $\mathscr{W}\to\mathscr{W}$). So the two recursion operators for symmetries Λ_1 and Λ_{-1} are each others' inverses. Application of Λ_1 and Λ_{-1} to Z_0 results in

$$\Lambda_1Z_0=Z_1,\quad \Lambda_{-1}Z_0=-\widehat{Z}_{-1}. \tag{3.11}$$

We now define two infinite series of symmetries X_k and Z_k by

$$X_k=\Lambda_1^kX_0,\quad Z_k=\Lambda_1^kZ_0,\quad k=0,\pm1,\pm2,\pm3,\dots . \tag{3.12}$$

By considering the highest derivatives with respect to x in X_k and Z_k and the structure of Λ_1 and Λ_{-1} it is easily seen that none of these symmetries vanishes. It follows from (3.11) and (3.12) that $\widehat{Z}_{-1}=-Z_{-1}$. The expressions for the symmetries X_1 and X_{-1} are given in Appendix B. From these expressions we see that the vector field X, which is trivially a symmetry, is given by

$$X=\tfrac12m(X_1+X_{-1}). \tag{3.13}$$

Because X_0 and Λ_1 do not depend explicitly on t (i.e., $\dot X_0=0$ and $\dot\Lambda_1=0$), the same holds for all symmetries X_k. Similarly we see that all symmetries X_k do not depend explicitly on x. The time derivative of Z_0 is given by

$$\dot Z_0=-\begin{pmatrix}u_{1x}\\ u_{2x}\\ v_{1x}\\ v_{2x}\end{pmatrix}=\frac12m(X_1-X_{-1}).$$

So the time derivatives of the symmetries Z_k are given by

$$\dot Z_k=\tfrac12m(X_{k+1}-X_{k-1}).$$

The symmetries X_0, X_1, X_{-1}, X_2, X_{-2}, X_3, X_{-3}, Z_0, Z_1, and Z_{-1} have already been given by Kersten.[1-3] In his notation they are called Y_4, $-2m^{-1}Y_1$, $2m^{-1}Y_2$, $4m^{-2}Y_5$, $-4m^{-2}Y_6$, $8m^{-3}Y_7$, $-8m^{-3}Y_8$, $-Y_3$, $m^{-1}Z_1$, and $-m^{-1}Z_2$.

After these elementary properties of the symmetries X_k and Z_k we now turn to the structure of the corresponding Lie algebra. Straightforward but long computations show that

$$[Z_0,Z_1]=Z_1,\quad [X_0,Z_1]=0,\quad [X_0,Z_0]=0.$$

Hence

$$\begin{aligned}\mathscr{L}_{Z_0}\Lambda_1&=\mathscr{L}_{Z_0}(\Omega^{\leftarrow}\mathscr{L}_{Z_1}\Omega)=\Omega^{\leftarrow}\mathscr{L}_{Z_0}\mathscr{L}_{Z_1}\Omega\\ &=\Omega^{\leftarrow}(\mathscr{L}_{[Z_0,Z_1]}\Omega+\mathscr{L}_{Z_1}\mathscr{L}_{Z_0}\Omega)\\ &=\Omega^{\leftarrow}\mathscr{L}_{Z_1}\Omega=\Lambda_1,\end{aligned} \tag{3.14}$$

where we used that Z_0 is a Hamiltonian vector field (i.e., $\mathscr{L}_{Z_0}\Omega=0$ and $\mathscr{L}_{Z_0}\Omega^{\leftarrow}=0$) and the formula $\mathscr{L}_{[A,B]}=\mathscr{L}_A\mathscr{L}_B-\mathscr{L}_B\mathscr{L}_A$ for all vector fields A and B. Similarly

$$\mathscr{L}_{X_0}\Lambda_1 = \mathscr{L}_{X_0}(\Omega^{\leftarrow}\mathscr{L}_{Z_1}\Omega) = \Omega^{\leftarrow}\mathscr{L}_{X_0}\mathscr{L}_{Z_1}\Omega$$
$$= \Omega^{\leftarrow}(\mathscr{L}_{[X_0,Z_1]}\Omega + \mathscr{L}_{Z_1}\mathscr{L}_{X_0}\Omega) = 0. \qquad (3.15)$$

The structure of the Lie algebra of symmetries generated by the X_k and Z_k can be found now from (2.8) if the Nijenhuis tensor field of Λ_1 vanishes. A gigantic computation shows that this is indeed the case. From (2.8) we now obtain that

$$[X_k,X_l] = 0,$$
$$[Z_k,Z_l] = (l-k)Z_{k+l}, \qquad (3.16)$$
$$[Z_k,X_l] = lX_{k+l},$$
$$k,l = 0, \pm 1, \pm 2, \pm 3,\ldots .$$

Also (2.7) yields

$$\mathscr{L}_{X_k}\Lambda_1 = 0, \quad \mathscr{L}_{Z_k}\Lambda_1 = \Lambda_1^{k+1},$$
$$k = 0, \pm 1, \pm 2, \pm 3,\ldots . \qquad (3.17)$$

The recursion operators for symmetries Λ_1 and Λ_{-1} have been found by substitution of Z_1 and $\widehat{Z}_{-1}$ $(= -Z_{-1})$ in (2.4). There are several other ways to construct recursion operators for symmetries. For instance, as explained in Sec. II, the Lie derivative of a recursion operator for symmetries in the direction of a symmetry yields again a recursion operator for symmetries. From (3.17) we see that in this way we only obtain powers of Λ_1. Another possible method is to use higher derivatives of Ω, i.e., to construct recursion operators of the form

$$\Omega^{\leftarrow}\mathscr{L}^p_{Z_1}\Omega, \quad \Omega^{\leftarrow}\mathscr{L}^p_{Z_{-1}}\Omega, \quad p = 1,2,3,\ldots . \qquad (3.18)$$

It is easily shown that this method also yields only powers of Λ_1. From (3.9) and $\widehat{Z}_{-1} = -Z_{-1}$ we obtain

$$\mathscr{L}_{Z_1}\Omega = \Omega\Lambda_1, \quad \mathscr{L}_{Z_{-1}}\Omega = -\Omega\Lambda_{-1}.$$

Using Leibniz' rule and (3.17) for $k = 1$ it is now easily shown by induction that

$$\mathscr{L}^p_{Z_1}\Omega = p!\Omega\Lambda_1^p,$$
$$\mathscr{L}^p_{Z_{-1}}\Omega = (-1)^p p!\Omega\Lambda^p_{-1} = (-1)^p p!\Omega\Lambda_1^{-p}. \qquad (3.19)$$

So the recursion operators for symmetries constructed by (3.18) are also powers of Λ_1.

The Lie derivative commutes with exterior differentiation. So we obtain from (3.19) the nontrivial conclusion that the two-forms $\Omega\Lambda_1^p$ $(p = 0, \pm 1, \pm 2,\ldots)$ are all closed. This result is in fact a special property of recursion operators with a vanishing Nijenhuis tensor field, see, for instance, Fuchssteiner and Fokas.[9] Using the closedness of $\Omega\Lambda_1^p$ it is easily shown that the symmetries X_k are Hamiltonian vector fields while the symmetries Z_k $(k \neq 0)$ are non-Hamiltonian vector fields. This follows because the closedness of $\Omega\Lambda_1^p$ implies that

$$d(\Omega X_k) = d(\Omega\Lambda_1^k X_0) = \mathscr{L}_{X_0}(\Omega\Lambda_1^k) = 0,$$
$$d(\Omega Z_k) = d(\Omega\Lambda_1^k Z_0) = \mathscr{L}_{Z_0}(\Omega\Lambda_1^k) \qquad (3.20)$$
$$= k\Omega\Lambda_1^k \neq 0, \quad \text{for } k \neq 0,$$

where we used (3.14) and (3.15) and that X_0 and Z_0 are Hamiltonian vector fields. The non-Hamiltonian symmetries Z_k $(k \neq 0)$ again give rise to recursion operators for symmetries. From (3.20) and the closedness of Ω we obtain

$$\Omega^{\leftarrow}\mathscr{L}_{Z_k}\Omega = \Omega^{\leftarrow}d(\Omega Z_k) = k\Lambda_1^k.$$

So also in this way we obtain only powers of Λ_1.

On the linear space $\mathscr{W}$ the closed one-forms ΩX_k are exact, so there exist constants of the motion F_k such that

$$X_k = \Omega^{\leftarrow} dF_k, \quad k = 0, \pm 1, \pm 2,\ldots .$$

The explicit form of F_1, F_{-1}, F_2 and F_{-2} is given in Appendix B. From (3.13) we obtain $H = \frac{1}{2}m(F_1 + F_{-1})$. The proof that the constants of the motion F_k are in involution is standard. Using the skew symmetry of $\Omega^{\leftarrow}$ and of $\mathscr{L}_{Z_1}\Omega$ we obtain for the Poisson bracket

$$\{F_k,F_l\} \equiv \langle dF_k, \Omega^{\leftarrow} dF_l\rangle$$
$$= \langle \Omega(\Omega^{\leftarrow}\mathscr{L}_{Z_1}\Omega)^k X_0, (\Omega^{\leftarrow}\mathscr{L}_{Z_1}\Omega)^l X_0\rangle = 0.$$

Thus we have constructed an infinite series of constants of the motion in involution for the massive Thirring model.

An infinite set of Hamiltonian forms of the massive Thirring model is now easily obtained. Some elementary manipulations lead to

$$X = (\Omega\Lambda_1^k)^{-1} d(\tfrac{1}{2}m(F_{k+1} + F_{k-1})),$$
$$k = 0, \pm 1, \pm 2,\ldots .$$

So we can consider X as the Hamiltonian vector field with Hamiltonian $\frac{1}{2}m(F_{k+1} + F_{k-1})$ and symplectic form $\Omega\Lambda_1^k$, for $k = 0, \pm 1, \pm 2,\ldots$. Note that the original Hamiltonian form of the Thirring model (3.3) is obtained for $k = 0$.

Finally we give a very simple recursion formula for the constants of the motion F_k. The Hamiltonian vector field corresponding to $\mathscr{L}_{Z_1}F_k$ is given by

$$\Omega^{\leftarrow}d\mathscr{L}_{Z_1}F_k = \Omega^{\leftarrow}\mathscr{L}_{Z_1}dF_k = \Omega^{\leftarrow}\mathscr{L}_{Z_1}(\Omega X_k)$$
$$= \Lambda_1 X_k + [Z_1,X_k] = (1+k)X_{k+1}$$
$$= (1+k)\Omega^{\leftarrow} dF_{k+1}.$$

This yields the recursion formula

$$F_{k+1} = \frac{1}{k+1}\mathscr{L}_{Z_1}F_k \equiv \frac{1}{k+1}\langle dF_k, Z_1\rangle, \quad k \neq -1.$$

In a similar way we obtain

$$F_{k-1} = \frac{1}{k-1}\mathscr{L}_{Z_{-1}}F_k \equiv \frac{1}{k-1}\langle dF_k, Z_{-1}\rangle, \quad k \neq 1.$$

In terms of the operator implementation of the differential geometry (see Appendix A) these two expressions read

$$F_{k+1} = \frac{1}{k+1}\int_{-\infty}^{\infty}\left(\frac{\delta F_k}{\delta u_1}Z_1^1 + \frac{\delta F_k}{\delta u_2}Z_1^2 + \frac{\delta F_k}{\delta v_1}Z_1^3 + \frac{\delta F_k}{\delta v_2}Z_1^4\right)dx,$$

$$F_{k-1} = \frac{1}{k-1}\int_{-\infty}^{\infty}\left(\frac{\delta F_k}{\delta u_1}Z^1_{-1} + \frac{\delta F_k}{\delta u_2}Z^2_{-1} + \frac{\delta F_k}{\delta v_1}Z^3_{-1} + \frac{\delta F_k}{\delta v_2}Z^4_{-1}\right)dx,$$

where $Z_1^1, Z_1^2, Z_1^3, Z_1^4$ and $Z^1_{-1}, Z^2_{-1}, Z^3_{-1}, Z^4_{-1}$ are the four components of the symmetries Z_1 (resp. Z_{-1}).

ACKNOWLEDGMENT

I thank Ms. M. Van Heijst for her assistance with several extremely long computations. In particular the computation of the Nijenhuis tensor field of Λ_1 could never have been completed without her help. I also thank Professor J. de Graaf for stimulating this research.

APPENDIX A: DIFFERENTIAL GEOMETRY ON A TOPOLOGICAL VECTOR SPACE

In the preceding section we worked completely in the setting of differential geometry. The used differential geometrical methods have a sound foundation on finite-dimensional manifolds. However, the massive Thirring model is considered on an infinite-dimensional topological vector space $\mathscr{M} = \mathscr{W}$. In this Appendix we shortly describe how the necessary differential geometry can be introduced on the topological vector space $\mathscr{W}$. A more comprehensive treatment is given in Ten Eikelder.[4] We assume that $\mathscr{W}$ is reflexive. The duality map between $\mathscr{W}$ and $\mathscr{W}^*$ will be denoted by $\langle\cdot,\cdot\rangle$. Since $\mathscr{W}$ is a linear space, we can make the following identifications for its tangent bundle and cotangent bundle:

$$\mathscr{T}\mathscr{W} = \mathscr{W}\times\mathscr{W}, \quad \mathscr{T}^*\mathscr{W} = \mathscr{W}\times\mathscr{W}^*.$$

Using these identifications it is easy to introduce (objects similar to) vector fields, differential forms, and tensor fields on $\mathscr{W}$. A vector field A on $\mathscr{W}$ is a mapping

$$A:\ \mathscr{W}\rightarrow\mathscr{W}\times\mathscr{W}:\ u\mapsto(u,\widetilde{A}(u)),$$

where $\widetilde{A}:\mathscr{W}\rightarrow\mathscr{W}$ is a possible nonlinear mapping. So we can identify the vector field A with the mapping $\widetilde{A}$. Therefore $\widetilde{A}$ also will be called a vector field. To simplify the notation we shall drop the tilde and write A instead of $\widetilde{A}$. In a similar way we can introduce one-forms and tensor fields of higher order. This results in the following "conversion table":

$A\in\mathscr{X}(\mathscr{W})$, vector field	$A:\mathscr{W}\rightarrow\mathscr{W}$	
$\alpha\in\mathscr{X}^*(\mathscr{W})$, one-form	$\alpha:\mathscr{W}\rightarrow\mathscr{W}^*$	
Φ (0,2) tensor field	$\Phi:\mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W}^*)$,	(A1)
Ψ (2,0) tensor field	$\Psi:\mathscr{W}\rightarrow L(\mathscr{W}^*,\mathscr{W})$	
Λ (1,1) tensor field	$\Lambda:\mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W})$	

where $L(\mathscr{W}_1,\mathscr{W}_2)$ denotes the linear continuous mappings from $\mathscr{W}_1$ to $\mathscr{W}_2$. For instance, the contracted multiplication between a (0,2) tensor field Φ and a (1,1) tensor field Λ yields a (0,2) tensor field represented by the mapping $\Phi\Lambda$: $\mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W}^*)$: $u\mapsto\Phi(u)\Lambda(u)$. In a similar way we can introduce higher-order tensor fields on $\mathscr{W}$. For instance, a (0,3) tensor field Ξ on $\mathscr{W}$ can be represented by a mapping Ξ: $\mathscr{W}\rightarrow L(\mathscr{W},L(\mathscr{W},L(\mathscr{W},\mathbb{R})))$.

Next we introduce Lie derivatives and (for differential forms) exterior derivatives. First some remarks on differential calculus in a topological vector space. Suppose $\mathscr{W}_1$ is some topological vector space and f is a (nonlinear) mapping f: $\mathscr{W}\rightarrow\mathscr{W}_1$. Then f is called Gateaux differentiable in $u\in\mathscr{W}$ if there exists a mapping $f'(u)\in L(\mathscr{W},\mathscr{W}_1)$ such that for all $v\in\mathscr{W}$

$$\lim_{\epsilon\rightarrow 0}(1/\epsilon)(f(u+\epsilon v)-f(u)+\epsilon f'(u)v)=0.$$

If f is Gateaux differentiable at all points $u\in\mathscr{W}$ we can consider the Gateaux derivative as a mapping f': $\mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W}_1)$. Suppose f' is again Gateaux differentiable in $u\in\mathscr{W}$. The second derivative of f in $u\in\mathscr{W}$ is then a mapping $f''(u)\in L(\mathscr{W},L(\mathscr{W},\mathscr{W}_1))$. This mapping can be considered as a bilinear mapping $f''(u)$: $\mathscr{W}\times\mathscr{W}\rightarrow\mathscr{W}_1$. Under certain conditions (see, for instance, Yamamuro[11]) this mapping is symmetric: $f''(u)(v,w)=f''(u)(w,v)$, for all $v,w\in\mathscr{W}$.

Suppose B: $\mathscr{W}\rightarrow\mathscr{W}$ is (represents) a vector field. The Gateaux derivative in $u\in\mathscr{W}$ is a linear mapping $B'(u)\in L(\mathscr{W},\mathscr{W})$. The dual of this mapping is denoted by $B'^*(u)\in L(\mathscr{W}^*,\mathscr{W}^*)$. The Lie derivatives in the direction of a vector field B of a function F: $\mathscr{W}\rightarrow\mathbb{R}$ and of the various tensor fields (vector fields, one-forms) considered in (A1) are defined by

$$\begin{aligned}
\mathscr{L}_B F(u) &= F'(u)B\equiv\langle F'(u),B\rangle,\\
\mathscr{L}_B A(u) &\equiv [B,A](u) = A'(u)B(u) - B'(u)A(u),\\
\mathscr{L}_B \alpha(u) &= \alpha'(u)B(u) + B'^*(u)\alpha(u),\\
\mathscr{L}_B \Phi(u) &= (\Phi'(u)B(u)) + \Phi(u)B'(u) + B'^*(u)\Phi(u),\\
\mathscr{L}_B \Lambda(u) &= (\Lambda'(u)B(u)) + \Lambda(u)B'(u) - B'(u)\Lambda(u),\\
\mathscr{L}_B \Psi(u) &= (\Psi'(u)B(u)) - \Psi(u)B'^*(u) - B'(u)\Psi(u).
\end{aligned}\tag{A2}$$

First some remarks on the notation in these expressions. Consider the formula for $\mathscr{L}_B\Phi$. Since Φ: $\mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W}^*)$ we have $\Phi'(u)\in L(\mathscr{W},L(\mathscr{W},\mathscr{W}^*))$. So $(\Phi'(u)B)\in L(\mathscr{W},\mathscr{W}^*)$ and $(\Phi'(u)B)C\in\mathscr{W}^*$ for $B,C\in\mathscr{W}$. By definition,

$$(\Phi'(u)B)C=\lim_{\epsilon\rightarrow 0}(1/\epsilon)(\Phi(u+\epsilon B)C-\Phi(u)C).$$

Of course, in general this expression is not symmetric in B and C. Therefore we shall always insert brackets in expressions of this type. It is easily seen that the Lie derivative of an object yields again an object of the same type. Note that the expressions given in (A2) strongly resemble the formulas for Lie derivatives in terms of local coordinates on a finite-dimensional manifold.

Now we turn to exterior derivatives of differential forms. Two-forms will be identified with skew-symmetric (0,2) tensor fields, i.e., Φ: $\mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W}^*)$ represents a two-form if $\Phi(u) = -\Phi^*(u)$ for all $u\in\mathscr{W}$. Let F: $\mathscr{W}\rightarrow\mathbb{R}$ be a function (= zero-form), α: $\mathscr{W}\rightarrow\mathscr{W}^*$ be a one-form, and Φ: $\mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W}^*)$ be a two-form. Then the exterior derivatives of F, α, and Φ are the one-, two-, and three-forms defined by

$$\begin{aligned}
&dF:\ \mathscr{W}\rightarrow\mathscr{W}^*,\ u\mapsto F'(u)\quad [\text{so } dF(u)=F'(u)],\\
&d\alpha:\ \mathscr{W}\rightarrow L(\mathscr{W},\mathscr{W}^*),\ u\mapsto\alpha'(u)-\alpha'^*(u),\\
&d\Phi:\ \mathscr{W}\rightarrow L(\mathscr{W},L(\mathscr{W},L(\mathscr{W},\mathbb{R}))),
\end{aligned}\tag{A3}$$

given by

$$d\Phi(u)(A,B,C) = \langle((\Phi'(u)A)B,C\rangle$$

$$+\langle (\Phi'(u)B)C,A\rangle + \langle (\Phi'(u)C)A,B\rangle.$$

Also these definitions strongly resemble the expressions in local coordinates of exterior derivatives of differential forms on a finite-dimensional manifold.

Definitions as above can, of course, always be given. The important observation, however, is that all formulas from classical differential geometry on a finite-dimensional manifold also hold in this case. The proofs of all used formulas are identical to the proofs in terms of local coordinates of the corresponding formulas on a finite-dimensional manifold. In particular we often used that for a closed two-form Φ and an arbitrary vector field A, the identity $\mathscr{L}_A\Phi = d(\Phi A)$ holds.

In the case of the massive Thirring model the duality map $\langle\cdot,\cdot\rangle$ between $\mathscr{W}$ and $\mathscr{W}^*$ is the L_2 inner product. In that case the derivative $F'(u)$ of a function (functional) F on $\mathscr{W}$ is usually denoted as $\delta F/\delta u$, the variational derivative of F. In terms of partial derivatives this means

$$dF(u) = \begin{pmatrix} \dfrac{\delta F}{\delta u_1} \\ \dfrac{\delta F}{\delta u_2} \\ \dfrac{\delta F}{\delta v_1} \\ \dfrac{\delta F}{\delta v_2} \end{pmatrix}.$$

APPENDIX B: SOME EXPLICIT FORMULAS

The symmetries X_1 and X_{-1} are given by

$$X_1 = \frac{1}{m}\begin{pmatrix} mv_2 - v_1R_2 \\ -2u_{2x} + mv_1 - v_2R_1 \\ -mu_2 + u_1R_2 \\ -2v_{2x} - mu_1 + u_2R_1 \end{pmatrix},$$

$$X_{-1} = \frac{1}{m}\begin{pmatrix} 2u_{1x} + mv_2 - v_1R_2 \\ mv_1 - v_2R_1 \\ 2v_{1x} - mu_2 + u_1R_2 \\ -mu_1 + u_2R_1 \end{pmatrix}.$$

The constants of the motion F_1, F_{-1}, F_2, and F_{-2} are given by

$$F_1 = \frac{1}{m}\int_{-\infty}^{\infty}\left(-\frac{1}{2}R_1R_2 + mR - 2u_{2x}v_2\right)dx,$$

$$F_{-1} = \frac{1}{m}\int_{-\infty}^{\infty}\left(-\frac{1}{2}R_1R_2 + mR + 2u_{1x}v_1\right)dx,$$

$$\begin{aligned} F_2 = \frac{1}{m^2}\int_{-\infty}^{\infty}\Big(&\frac{1}{2}R_1R_2(R_1+R_2) - mR(R_1+R_2) \\ &+ \frac{1}{2}m^2(R_1+R_2) - 2m(u_{2x}v_1 + u_{1x}v_2) \\ &+ 2u_{2x}^2 + 2v_{2x}^2 + 2R_1(u_{2x}v_2 - v_{2x}u_2) \\ &+ 4u_2v_2(u_{2x}u_2 - v_{2x}v_2)\Big)dx, \end{aligned}$$

$$\begin{aligned} F_{-2} = \frac{1}{m^2}\int_{-\infty}^{\infty}\Big(&\frac{1}{2}R_1R_2(R_1+R_2) - mR(R_1+R_2) \\ &+ \frac{1}{2}m^2(R_1+R_2) + 2m(u_{1x}v_2 + u_{2x}v_1) \\ &+ 2u_{1x}^2 + 2v_{1x}^2 + 2R_2(v_{1x}u_1 - u_{1x}v_1) \\ &+ 4u_1v_1(v_{1x}v_1 - u_{1x}u_1)\Big)dx. \end{aligned}$$

To reduce the expressions for the recursion operators we introduce the following abbreviations:

$$\begin{aligned} &(ij\,kl) = u_i\,\partial^{-1}u_jR_k + u_iR_l\,\partial^{-1}u_j, \\ &\quad i,j = 1,2,3,4,\ \ k,l = 1,2, \\ &\{ij\,kl\} = mu_i\,\partial^{-1}u_j + mu_k\,\partial^{-1}u_l, \\ &\quad i,j,k,l = 1,2,3,4, \\ &(ij\,) = 2u_{ix}\,\partial^{-1}u_j, \quad i,j = 1,2,3,4, \\ &\{ij\,\} = 2u_i\,\partial^{-1}u_{jx}, \quad i,j = 1,2,3,4, \end{aligned}$$

where $u_3 = v_1$ and $u_4 = v_2$. The recursion operators for symmetries of the massive Thirring model Λ_1 and Λ_{-1} are now given by

$$\Lambda_1 = \frac{1}{m}\begin{pmatrix} -(3122)+\{3241\} & -(3212)+\{3142\} + \{34\} + m - 2v_1v_2 & -(3322)+\{3443\} & -(3412)+\{3344\} - \{32\} + 2u_2v_1 \\ -(4121)+\{3142\} - (21) + m - 2u_1u_2 & -(4211)+\{3241\} + \{44\} - (22) - R_1 - R_2 & -(4321)+\{3344\} - \{23\} - 2u_2v_1 & -(4411)+\{3443\} - \{42\} - (24) + 2\partial \\ (1122)-\{1221\} & (1212)-\{1122\} - \{14\} + 2u_1v_2 & (1322)-\{1423\} & (1412)-\{1324\} + \{12\} + m - 2u_1u_2 \\ (2121)-\{1122\} - (41) - 2u_1v_2 & (2211)-\{1221\} - \{24\} - (42) - 2\partial & (2321)-\{1324\} - (43) + m - 2v_1v_2 & (2411)-\{1423\} + \{22\} - (44) - R_1 - R_2 \end{pmatrix},$$

$$\Lambda_{-1} = \frac{1}{m}\begin{pmatrix} (3122)-\{3241\} + \{33\} - (11) - R_1 - R_2 & (3212)-\{3142\} - (12) + m - 2u_1u_2 & (3322)-\{3443\} - \{31\} - (13) - 2\partial & (3412)-\{3344\} - (14) - 2u_1v_2 \\ (4121)-\{3142\} + \{43\} + m - 2v_1v_2 & (4211)-\{3241\} & (4321)-\{3344\} - \{41\} + 2u_1v_2 & (4411)-\{3443\} \\ -(1122)+\{1221\} - \{13\} - (31) + 2\partial & -(1212)+\{1122\} - (32) - 2u_2v_1 & -(1322)+\{1423\} + \{11\} - (33) - R_1 - R_2 & -(1412)+\{1324\} - (34) + m - 2v_1v_2 \\ -(2121)+\{1122\} - \{23\} + 2u_2v_1 & -(2211)+\{1221\} & -(2321)+\{1324\} + \{21\} + m - 2u_1u_2 & -(2411)+\{1423\} \end{pmatrix}.$$

[1]P. H. M. Kersten, Ph.D. thesis, Twente University of Technology, The Netherlands, 1985.
[2]P. H. M. Kersten and R. Martini, J. Math. Phys. **26**, 822 (1985).
[3]P. H. M. Kersten and R. Martini, J. Math. Phys. **26**, 1775 (1985).
[4]H. M. M. Ten Eikelder, *Symmetries for Dynamical and Hamiltonian Systems*, CWI tract 17 (Centre for Mathematics and Computer Science, Amsterdam, 1985).
[5]A. Nijenhuis, Ned. Akad. Wetensch. Proc. Ser. A **54**, 200 (1951) [Indag. Math. **13**, 200 (1951)].
[6]J. A. Schouten, *Ricci Calculus* (Springer, Berlin, 1954), 2nd ed.
[7]F. Magri, "A geometrical approach to the nonlinear solvable equations," in *Nonlinear Evolution Equations and Dynamical Systems, Lecture Notes in Physics*, Vol. 120 (Springer, Berlin, 1980).
[8]B. Fuchssteiner, Nonlinear Anal. Theor. Meth. Appl. **3**, 849 (1979).
[9]B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[10]I. M. Gel'fand and I. Ya. Dorfman, Funct. Anal. Appl. **13**, 248 (1979); **14**, 223 (1980).
[11]S. Yamamuro, *Differential Calculus in Topological Linear Spaces, Lecture Notes in Mathematics*, Vol. 374 (Springer, Berlin, 1974).

Dielectric description of hadrons with anti-de Sitter symmetry

E. van Beveren
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104
T. A. Rijken and C. Dullemond
Institute of Theoretical Physics, University of Nijmegen, The Netherlands

(Received 10 October 1985; accepted for publication 8 January 1986)

In the SO(3,2) symmetric strong curvature and related dielectric description of the quantum chromodynamic vacuum, the classical Maxwell equations are solved for charged particles moving along geodesics. An alternative five-dimensional description of the quark and gluon dynamics is studied in detail. Some interesting features of the anti-de Sitter bag are discussed.

I. INTRODUCTION

Although space curvature is intimately connected with a dielectric description for electromagnetism as shown in Ref. 1, its relevance for strong interactions has only been developed recently.[2] As recognized by Salam and Strathdee[3] several years ago, the best candidate for strong curvature is the closed SO(3,2) symmetric anti-de Sitter universe. Two superimposed geometries, one for electromagnetism and another for strong interactions, describe in this view the dynamics of quarks, gluons, electrons, and photons. However, the connection with the standard quantum chromodynamic (QCD) theory for strong interactions was hard to make. In that area of physics most theoreticians preferred the approach of Kogut and Susskind[4] and Lee,[5] which describes the QCD vacuum in terms of a chromodielectric field. The first sign of a possible "vierbein coupling" for quarks with the dielectric field appeared in the work of Nielsen and Patkós.[6] In an elaborated investigation,[7] we showed that the concept of symmetry breaking gives an interesting strategy for the derivation of the SO(3)×SO(2) symmetric background field from QCD.

A model for hadrons, in which quarks, antiquarks, and gluons move inside a finite spherical and strongly curved anti-de Sitter universe, is discussed in a series of papers.[7–10] The radius of this spherical baglike structure is related to the average level splittings of hadron spectra, which we know are nearly flavor mass independent. Consequently the radii are about the same for all hadrons (or at least for all mesons) and can be taken to be universal for a first inspection of the properties of the hadrons described by the anti-de Sitter bag. The associated spectrum is flavor mass independent and equidistant, which at first sight does not seem to be in agreement with the experimental observations. However, in the nonrelativistics limit of our model, where the valence partons are bound by harmonic oscillator forces with universal frequency, we have shown that an equidistant flavor mass independent bare spectrum can lead in a natural way to the physical spectra once the influence of hadronic decay has been taken into account. In several papers the properties of bare hadrons versus physical hadrons are discussed.[1,11,12] The main aim of these investigations was to show that the properties of bare hadrons cannot be easily understood from the spectra and the decay properties of the physical hadrons, but that one has to derive and agree on a method for the subtraction of strong decay effects.

The testing ground for this approach is the spectra and other properties of hadrons. A rigorous theoretical framework does not exist, and therefore many models have been suggested. In the bag model[13] quarks and gluons move freely inside a cavity with a sharp boundary. In QCD inspired nonrelativistic Schrödinger models,[14–18] interquark potentials rise to infinity. All these models consider hadrons as quark bound states, a concept that has been fruitful in other areas of physics. Moreover they all have quark confinement built in *a priori*. They all do not seem to support the harmonic oscillator description that follows from the anti-de Sitter dynamics, however, although it cannot be stated that no attention at all has been paid to strong decay (see, e.g., Refs. 18–20), we feel that in most investigations the effect of it on the theoretical predictions has been underestimated. In Ref. 21 our results in the nonrelativistic limit are compared with the experimental data of heavy quarkonia and in Refs. 12, 22, and 23 it is shown that even the light mesons with, however, not very light constituent valence quarks, fit well in our scheme. Our calculations show that the properties of the bare hadrons can be considerably different from those of the physical hadrons due to strong decay.

The connection between the anti-de Sitter bag and QCD is extensively dealt with in Ref. 7. There we assume that the Lagrangian of the scalar field, which is to be associated with the chromodielectric, has the naive conformal symmetry of the gluon sector. Fubini[24] showed that breaking to an SO(3,2) symmetric solution is possible if one starts from a conformal invariant Lagrangian for a scalar field σ. We assume that the solution for σ is relevant for the description of hadrons. In conformal coordinates the Fubini solution has singularities in space and time. For the description of confinement we use the central projection coordinates,[8] for which the singularities become spacelike and static, defining the closed and localized domain in space, which we called the anti-de Sitter bag before. Our study of σ leads to the strong interaction picture of quark confinement.

We assume that broken conformal symmetry is a valuable framework to understand the hadronic spectra. In our choice we are left with anti-de Sitter symmetry, which does not contain Poincaré symmetry. Note, however, that the broken O(4,2) group describes part of the internal structure of a hadron, rather than the whole world. Our σ solution, which represents the gluon sector of a localized system like a hadron, by itself does not need to preserve Poincaré symmetry. In Ref. 7 we describe in detail how the coupling of photons to the scalar field σ via a tensor $t_{\mu\nu}$ leads to the above mentioned central projection. The tensor $t_{\mu\nu}$ can be considered as the photon metric and becomes Minkowskian (i.e., diagonal $+,-,-,-$) under the central projection. This way we obtain two metric fields: one curved metric to which the strong interacting fields are coupled and one flat metric which couples to ordinary electromagnetism.

The precise connection between the color electromagnetic properties of the gluon background medium and our strong curvature description is discussed in Refs. 1 and 2. There we have shown that our dielectric and magnetic susceptibility functions differ from those suggested in the main literature (see, e.g., Ref. 5). The differences being that in our case these quantities are tensors rather than scalars and equal rather than inversely proportional. The consequences of these properties are, however, in agreement with what we expect to find as we will show briefly in this investigation.

We conclude that it is worthwhile to study the anti-de Sitter bag in more detail. The final goal might be to study perturbation theory in this strongly curved universe. For the moment we suffice with the study of ordinary electromagnetism in the five- and four-dimensional descriptions of anti-de Sitter space. (See also Ref. 25.)

In this paper we investigate the advantage of the five-dimensional approach as a basis for further calculations. For this we restrict ourselves to the Abelian case because the extension to the non-Abelian case is straightforward but not necessary for a first inspection.

II. ABELIAN CHROMODYNAMICS AND MAXWELL'S EQUATIONS IN ANTI-DE SITTER SPACE

The static version of the anti-de Sitter bag with finite spatial radius R embedded in Minkowskian coordinates x^μ (throughout this paper it is understood that Greek indices run from 0 to 3, that Latin indices $a,b,c,d,\ldots$ run from 1 to 5, and that Latin indices $i,j,k,l,\ldots$ run from 1 to 3) is given by the metric

$$g_{00}(x^0,\mathbf{x}) = (1-\alpha r^2)^{-1}, \quad g_{0i} = g_{i0} = 0,$$

and

$$g_{ij}(x^0,\mathbf{x}) = (1-\alpha r^2)^{-1}\{\eta_{ij} - (1-\alpha r^2)^{-1}\alpha x^i x^j\}, \tag{2.1}$$

where

$$\alpha = R^{-2},$$

$$r^2 = -\eta_{ij}x^i x^j,$$

and

$$\eta_{\mu\nu} = \text{diag}(+,-,-,-).$$

Metric (2.1) describes the O(3,2) invariant four-dimensional curved surface $\xi^a\xi_a = R^2$ embedded in flat five dimensions with coordinates ξ^a and metric $\eta_{ab} = \text{diag}(-,-,-,+,+)$, under the central projection[8] given by

$$(\xi^i,\xi^4,\xi^5) = (1-\alpha r^2)^{-1/2} \times (x^i, R\sin(x^0/R), R\cos(x^0/R)). \tag{2.2}$$

In the following, we study electromagnetism simultaneously in the x^μ and ξ^a coordinates.

We assume that the Lagrangian for a charged scalar field $\phi(\xi^a)$ in the four-dimensional hypersurface $\xi^a\xi_a = R^2$ can be written

$$\mathscr{L}(\phi) = \eta^{ab}(\mathring{\partial}_a\phi)(\mathring{\partial}_b\phi^*) - m(m+(3/R))\phi\phi^*, \tag{2.3}$$

where the tangentional derivative $\mathring{\partial}_a$ is defined by

$$\mathring{\partial}_a = \frac{\partial}{\partial\xi^a} - \alpha\xi_a\xi^b\frac{\partial}{\partial\xi^b}, \tag{2.4}$$

and which satisfies the condition

$$\xi^a\mathring{\partial}_a = 0. \tag{2.5}$$

The generalized Klein–Gordon equation for $\phi(\xi^a)$, which follows via the Euler–Lagrange equation

$$\frac{\partial\mathscr{L}}{\partial\phi^*} - \mathring{\partial}_a\frac{\partial\mathscr{L}}{\mathring{\partial}_a\phi^*} = 0, \tag{2.6}$$

from the Lagrangian (2.3) reads

$$M^2\phi = mR(mR+3)\phi. \tag{2.7}$$

In (2.7) the generalized angular momentum operator in five dimensions M^2 is defined by

$$\begin{aligned} M^2 &= \tfrac{1}{2}M^{ab}M_{ab} \\ &= -\frac{1}{2}\left(\xi^a\frac{\partial}{\partial\xi_b} - \xi^b\frac{\partial}{\partial\xi_a}\right)\left(\xi_a\frac{\partial}{\partial\xi^b} - \xi_b\frac{\partial}{\partial\xi^a}\right) \\ &= -R\eta^{ab}\mathring{\partial}_a\mathring{\partial}_b. \end{aligned} \tag{2.8}$$

The five-dimensional electromagnetic vector field $\widetilde{A}_a$ defined on the hypersurface $\xi^a\xi_a = R^2$ might be chosen such that it satisfies the analog of Eq. (2.5), i.e.,

$$\xi^a\widetilde{A}_a = 0, \tag{2.9}$$

and transforms under the gauge transformation $\phi(\xi) \to \exp(ie\theta(\xi))\phi(\xi)$ according to

$$\widetilde{A}_a \to \widetilde{A}_a + \mathring{\partial}_a\theta(\xi). \tag{2.10}$$

From (2.10) it follows that the Lagrangian (2.3) can be attributed with electromagnetism by means of the minimal substitution

$$\mathring{\partial}_a \to \mathring{\partial}_a - ie\widetilde{A}_a. \tag{2.11}$$

Equation (2.11) allows the comparison of the five components of the vector field $\widetilde{A}_a$ and the four components of the usual vector field A_μ, which can be introduced by means of the standard minimal substitution

$$\partial_\mu = \frac{\partial}{\partial x^\mu} \to \partial_\mu - ieA_\mu. \tag{2.12}$$

We might use (2.2) to derive the relation between $\mathring{\partial}_a$ from (2.4) and ∂_μ to obtain

$$(\mathring{\partial}_i,\mathring{\partial}_4,\mathring{\partial}_5) = (1-\alpha r^2)^{1/2} \times \left(\partial_i, \cos\left(\frac{x^0}{R}\right)\partial_0 - \frac{1}{R}\sin\left(\frac{x^0}{R}\right)x^k\partial_k,\right.$$

$$-\sin\left(\frac{x^0}{R}\right)\partial_0-\frac{1}{R}\cos\left(\frac{x^0}{R}\right)x^k\partial_k\right). \tag{2.13}$$

This equation together with (2.11) and (2.12) leads to the relation between A_μ and $\widetilde{A}_a$:

$$(\widetilde{A}_i,\widetilde{A}_4,\widetilde{A}_5)=(1-\alpha r^2)^{1/2}\times\left(A_i,\cos\left(\frac{x^0}{R}\right)A_0-\frac{1}{R}\sin\left(\frac{x^0}{R}\right)x^kA_k,\ -\sin\left(\frac{x^0}{R}\right)A_0-\frac{1}{R}\cos\left(\frac{x^0}{R}\right)x^kA_k\right), \tag{2.14a}$$

and the reverse relations

$$(A_0,A_i)=\left(\widetilde{A}_4\frac{\xi_5}{R}-\widetilde{A}_5\frac{\xi_4}{R},(1+\alpha\rho^2)^{1/2}\widetilde{A}_i\right), \tag{2.14b}$$

where ρ is defined by

$$\rho^2=\boldsymbol{\xi}^2=-\eta_{ij}\xi^i\xi^j. \tag{2.15}$$

From (2.14a) together with (2.2) it is easy to check that $\widetilde{A}_a(\xi)$ satisfies the condition (2.9). Under hyperspherical rotations, $\widetilde{A}^a=\eta^{ab}\widetilde{A}_b$ transforms as a vector field. Moreover hyperspherical rotations induce coordinate transformations in the four-dimensional formulation of the anti-de Sitter bag (2.1). Under such coordinate transformations A_μ transforms according to

$$A'_\mu=\frac{\partial x^\nu}{\partial x'^\mu}A_\nu. \tag{2.16}$$

Next we want to construct an antisymmetric 5×5 electromagnetic tensor field $\widetilde{F}_{ab}$, for the five-field $\widetilde{A}_a$, that satisfies a condition similar to (2.9), i.e.,

$$\xi^a\widetilde{F}_{ab}=0, \tag{2.17}$$

and for which the kinetic term equals the similar term for the four-potential A_μ, i.e.,

$$-\tfrac{1}{4}\widetilde{F}^{ab}\widetilde{F}_{ab}=-\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{2.18}$$

A choice that satisfies these conditions is[26]

$$\widetilde{F}_{ab}=(\mathring{\partial}_a-\alpha\xi_a)\widetilde{A}_b-(\mathring{\partial}_b-\alpha\xi_b)\widetilde{A}_a. \tag{2.19}$$

Let us check the invariance of the choice (2.19) under the gauge transformation (2.10):

$$\widetilde{F}'_{ab}=\widetilde{F}_{ab}+[\mathring{\partial}_a,\mathring{\partial}_b]\theta(\xi)-\alpha(\xi_a\mathring{\partial}_b-\xi_b\mathring{\partial}_a)\theta(\xi)=\widetilde{F}_{ab}. \tag{2.20}$$

The relations between the components of $\widetilde{F}_{ab}$ and $F_{\mu\nu}$ read

$$\widetilde{F}_{ij}=(1-\alpha r^2)F_{ij},$$

$$\widetilde{F}_{4i}=(1-\alpha r^2)\left\{\cos\left(\frac{x^0}{R}\right)F_{0i}-\frac{1}{R}\sin\left(\frac{x^0}{R}\right)x^kF_{ki}\right\},$$

$$\widetilde{F}_{5i}=(1-\alpha r^2)\left\{-\sin\left(\frac{x^0}{R}\right)F_{0i}-\frac{1}{R}\cos\left(\frac{x^0}{R}\right)x^kF_{ki}\right\}, \tag{2.21a}$$

and

$$\widetilde{F}_{45}=(1-\alpha r^2)\frac{1}{R}x^kF_{k0},$$

and the reverse relations

$$F_{ij}=(1+\alpha\rho^2)\widetilde{F}_{ij},$$

and

$$F_{0i}=(1+\alpha\rho^2)^{1/2}\left\{\frac{\xi_5}{R}\widetilde{F}_{4i}-\frac{\xi_4}{R}\widetilde{F}_{5i}\right\}. \tag{2.21b}$$

With the use of Eq. (2.21) it is easy to verify that for the kinetic terms the equality (2.18) holds.

Starting from (2.18) the Euler–Lagrange equations (2.6) for $\widetilde{A}_a$ lead to the free field equations

$$\eta^{ab}\mathring{\partial}_a\mathring{\partial}_b\widetilde{A}_c-\eta^{ab}\mathring{\partial}_a\mathring{\partial}_c\widetilde{A}_b-3\alpha\widetilde{A}_c=0. \tag{2.22}$$

In the generalized Lorentz condition

$$0=\eta^{ab}\mathring{\partial}_a\widetilde{A}_b=g^{\mu\nu}A_{\mu;\nu}, \tag{2.23}$$

Eq. (2.22) can be reduced to

$$(M^2+2)\widetilde{A}_a=0. \tag{2.24}$$

Equation (2.24) is the conformal invariant form of Eq. (2.7) for $mR=-1$ (or $mR=-2$), the form of which has to be associated with "massless" fields.

The operator M^2, which is defined in (2.8), can be expressed in the Minkowskian coordinates x^μ via the central projection (2.2). We obtain

$$M^2=R^2(1-\alpha r^2)\{-\partial_0^2+\partial_i^2-\alpha(x^i\partial_i)^2\}. \tag{2.25}$$

So if we combine expressions (2.25) and (2.24) the differential equation in flat coordinates x^μ for each component of the field $\widetilde{A}_a$ reads

$$R^2[(1-\alpha r^2)\{-\partial_0^2+\partial_i^2-\alpha(x^i\partial_i)^2\}+2\alpha]\widetilde{A}_a=0. \tag{2.26}$$

In the following we will study solutions of the inhomogeneous generalization of Eq. (2.26).

III. THE DIELECTRIC DESCRIPTION

Suppose we place a unit charge somewhere inside the anti-de Sitter bag (2.1). Then in terms of a static vector potential for which we choose only one nonvanishing component

$$A_\mu(\mathbf{x})=(A_0(\mathbf{x}),\mathbf{A}(\mathbf{x})=0), \tag{3.1}$$

we expect to obtain an equation of the form

$$\mathscr{O}A_0(\mathbf{x})=\delta(\mathbf{x}-\mathbf{a}), \tag{3.2}$$

where $\mathscr{O}$ is a differential operator and $\mathbf{a}$ is the position of the unit charge. In the Appendix we show that the anti-de Sitter bag (2.1) can be described in terms of a local dielectric function $\epsilon_{ij}(\mathbf{x})$ for which the operator $\mathscr{O}$ of (3.2) would satisfy (see also Ref. 1)

$$\mathscr{O}A_0(\mathbf{x})\sim\partial_iD_i(\mathbf{x})=\partial_i\epsilon_{ij}(\mathbf{x})E_j(\mathbf{x})=-\partial_i\epsilon_{ij}(\mathbf{x})\partial_jV(\mathbf{x}), \tag{3.3}$$

where $\mathbf{D}$, $\mathbf{E}$, and V are the displacement field, the electric field, and the electric potential of the charge distribution (3.2), respectively. First we will derive the connection between (3.3) and the more general equation (2.26). From (2.14a) we obtain for the static vector potential (3.1) the corresponding $\widetilde{A}$ field

$$(\widetilde{A}_i,\widetilde{A}_4,\widetilde{A}_5)=(1-\alpha r^2)^{1/2}$$

$$\times \left(0,0,0,\cos\left(\frac{x^0}{R}\right)A_0, -\sin\left(\frac{x^0}{R}\right)A_0\right). \quad (3.4)$$

If we next apply (2.26) to, e.g., $\widetilde{A}_4$ of (3.4) then we obtain for the free field equation

$$R^2(1-ar^2)^{3/2}\cos(x^0/R) \times\{\partial_i^2 - a(x^i\partial_i)^2 - 2ax^i\partial_i\}A_0(\mathbf{x}) = 0. \quad (3.5)$$

In (3.5) the time derivative disappeared because $A_0(\mathbf{x})$ is static. So finally we end up with a differential equation for the potential $A_0(\mathbf{x})$, which might be written in the form

$$(1-ar^2)^{-1/2}\{\partial_i^2 - a(x^i\partial_i)^2 - 2ax^i\partial_i\}A_0(\mathbf{x}) = 0. \quad (3.6)$$

Equation (3.6) can again be rewritten to read

$$\partial_i(1-ar^2)^{-1/2}(\delta_{ij} - ax^ix^j)\partial_j A_0(\mathbf{x}) = 0. \quad (3.7)$$

Comparison of (3.7) with (3.3) gives for the dielectric tensor the form

$$\epsilon_{ij}(\mathbf{x}) = (1-ar^2)^{-1/2}\{\delta_{ij} - ax^ix^j\}, \quad (3.8)$$

which is in agreement with the result published in Ref. 1. In the Appendix it is also shown that the magnetic susceptibility tensor following from the medium (2.1) is equal to the dielectric tensor (3.8).

In the more general case where we do not start out from the free Lagrangian (2.18) but also allow source terms to be present, the field equations for the $\widetilde{A}$ field might read

$$(1/R^2)(M^2+2)\widetilde{A}_a = -\widetilde{J}_a. \quad (3.9)$$

For the A_μ field in the curved space (2.1), the similar equation read

$$(\sqrt{-g}F^{\mu\nu})_{,\mu} = \sqrt{-g}J^\nu. \quad (3.10)$$

In the Appendix it is shown how (3.10) for the metric (2.1) can be interpreted in terms of the electromagnetic constants (i.e., dielectric constant and magnetic susceptibility, respectively). It is straightforward to obtain the relations between the components of the currents $\widetilde{J}_a$ of (3.9) and J_μ of (3.10). In the gauge (2.23) we obtain the relations

$$(\widetilde{J}_i,\widetilde{J}_4,\widetilde{J}_5) = (1-ar^2)^{1/2} \times\left(J_i, \cos\left(\frac{x^0}{R}\right)J_0 - \frac{1}{R}\sin\left(\frac{x^0}{R}\right)x^kJ_k, -\sin\left(\frac{x^0}{R}\right)J_0 - \frac{1}{R}\cos\left(\frac{x^0}{R}\right)x^kJ_k\right), \quad (3.11a)$$

and the reverse relations

$$(J_0,J_i) = \left(\widetilde{J}_4\frac{\xi_5}{R} - \widetilde{J}_5\frac{\xi_4}{R}, (1+a\rho^2)^{1/2}\widetilde{J}_i\right). \quad (3.11b)$$

The relations (3.11) between $\widetilde{J}$ and J are similar to those for $\widetilde{A}$ and A (2.14). This is what we had to expect, because the transformation properties for $\widetilde{J}$ and J under hyperspherical rotations and associated coordinate transformations, respectively, are the same as for $\widetilde{A}$ and A.

With the results of this section we are sufficiently prepared to solve the field equations for the electromagnetic field in the anti-de Sitter bag for a given charge configuration and to select the most convenient description in each specific case. For instance in the dielectric description we can formulate the problem in terms of static Maxwell equations and define **E**, **D**, **H**, and **B** fields, which allow a transparent interpretation. Formulated in terms of $\widetilde{A}$ fields the interpretation might be obscure but the transformation properties under SO(3,2) hyperspherical rotations are very convenient. The concept of strong curvature and the use of x^μ coordinates connects the problem with general relativity and the machinery developed in that area of physics.

IV. A AND $\widetilde{A}$ FIELDS FOR CHARGES MOVING ALONG GEODESICS

The practical use of the five-dimensional description is most conveniently studied from the solutions of (3.10) for electric charges moving along geodesics. Let us first consider the spherical symmetric case of an electric charge at the center of the space (2.1). In this case we assume for the current J_μ the form

$$J_\mu = (J_0 = q\delta(\mathbf{x}), \mathbf{J} = 0), \quad (4.1)$$

and for the vector potential A_μ the form

$$A_\mu = (A_0(r), \mathbf{A} = 0). \quad (4.2)$$

With (4.1), (4.2), and (2.1) Maxwell's equations (3.10) reduce to

$$\{\partial_i^2 - a(x^i\partial_i)^2 - 2ax^i\partial_i\}A_0 = -q\delta(\mathbf{x}), \quad (4.3a)$$

or in terms of the dielectric (3.11)

$$\partial_i\epsilon_{ij}(\mathbf{x})\partial_j A_0(r) = -q\delta(\mathbf{x}), \quad (4.3b)$$

the equations of which are solved by

$$A_0(r) = (q/4\pi)(1/r)(1-ar^2)^{1/2}. \quad (4.4)$$

In Sec. V we come back to an interesting feature of this solution.

From solution (4.4) one can calculate all electromagnetic fields for the case of a single charge in the center of the anti-de Sitter bag. For this the necessary equations are

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu,$$
$$E_i(\mathbf{x},t) = F_{0i}(x),$$
$$B_i(\mathbf{x},t) = -\tfrac{1}{2}\epsilon_{ijk}F_{jk}(x),$$
$$D_i(\mathbf{x},t) = -\sqrt{-g(x)}F^{0i}(x),$$

and

$$H_i(\mathbf{x},t) = -\tfrac{1}{2}\sqrt{-g(x)}\epsilon_{ijk}F^{jk}(x). \quad (4.5)$$

The **D** and **H** fields could equivalently be derived from the **E** and **B** fields, respectively, via the electromagnetic constants $\epsilon_{ij}(\mathbf{x})$ and $\mu_{ij}(\mathbf{x})$ with

$$D_i(\mathbf{x},t) = \epsilon_{ij}(\mathbf{x})E_j(\mathbf{x},t),$$

and

$$B_i(\mathbf{x},t) = \mu_{ij}(\mathbf{x})H_j(\mathbf{x},t). \tag{4.6}$$

If we insert the static solution (4.4) into Eqs. (4.5) and (4.6), then we obtain

$$F_{0i}(x) = (q/4\pi)(1-\alpha r^2)^{-1/2}(x^i/r^3),$$

$$F_{ij}(x) = 0,$$

$$\mathbf{D}(\mathbf{x},t) = (q/4\pi)(\mathbf{x}/r^3), \tag{4.7}$$

and

$$\mathbf{B}(\mathbf{x},t) = 0.$$

Let us next study the $\widetilde{J}$ current and the $\widetilde{A}$ and $\widetilde{F}$ fields in the five-dimensional description. From the Eqs. (3.11a) and (4.1), we find for the current

$$\widetilde{J}_a = q\delta(\mathbf{x})\cdot\left(0,0,0,\cos\left(\frac{x^0}{R}\right), -\sin\left(\frac{x^0}{R}\right)\right), \tag{4.8}$$

and from Eqs. (2.2), (2.14a) and (4.4), we find analogously for the vector field

$$\widetilde{A}_a = \frac{q}{4\pi}\frac{1-\alpha r^2}{r}\left(0,0,0,\cos\left(\frac{x^0}{R}\right), -\sin\left(\frac{x^0}{R}\right)\right),$$

$$= \frac{q}{4\pi}\frac{1}{\rho(1+\alpha\rho^2)}\left(0,0,0,\frac{\xi^5}{R}, -\frac{\xi^4}{R}\right), \tag{4.9}$$

where ρ is defined in (2.15). The tensor field can be found from (2.21b) and (4.7) and reads

$$\widetilde{F}_{ij} = 0,$$

$$\widetilde{F}_{4i} = (q/4\pi)(1/R\rho^3)\xi^5\xi^i,$$

$$\widetilde{F}_{5i} = -(q/4\pi)(1/R\rho^3)\xi^4\xi^i, \tag{4.10}$$

and

$$\widetilde{F}_{45} = -(q/4\pi)(1/R\rho).$$

The first observation from (4.9) is that $\widetilde{A}_a$ has two nonzero components in the four and five directions. This is quite unerstandable since a static charge in the anti-de Sitter bag is the projection of the line given by

$$(\xi^4)^2 + (\xi^5)^2 = R^2, \tag{4.11}$$

in the O(3,2) invariant hypersurface (2.2). The next observation is that the nonzero components of $\widetilde{A}$ approach zero faster than ρ^{-1} in the limit $\rho\to\infty$. We find

$$\widetilde{A}_4 \ (\text{and } \widetilde{A}_5) \underset{\rho\to\infty}{\longrightarrow} \sim (1/\rho^2). \tag{4.12}$$

We next come to the point of the electromagnetic fields for charged particles that move along geodesics in the anti-de Sitter bag. All geodesics in this space are connected via SO(3,2) transformations. So we can start out from the static charge in the center of the bag and boost it up via a SO(3,2) boost. Associated with such a boost is a coordinate transformation

$$x^\mu \to x'^\mu(x), \tag{4.13}$$

and since A_μ and $F_{\mu\nu}$ transform as a vector and a tensor, respectively, we can find the A'_μ and $F'_{\mu\nu}$ for the boosted particle via

$$A'_\mu = \frac{\partial x^\nu}{\partial x'^\mu}A_\nu \quad \text{and} \quad F'_{\mu\nu} = \frac{\partial x^\sigma}{\partial x'^\mu}\frac{\partial x^\rho}{\partial x'^\nu}F_{\sigma\rho}. \tag{4.14}$$

However, the complete calculation (4.14) even for the simple case (4.7) is in general quite tedious. For that reason, it is more convenient to consider the boost in five dimensions and its effects on the five-vector and five-tensor fields.

Let us consider the case of an electric charge oscillating along the x axis. For an SO(3,2) transformation in five space that boosts the electric charge (4.1) as to make geodesic oscillations along the x axis with frequency R^{-1} and amplitude a, we might choose

$$(\Lambda^a{}_b) = \left\|\begin{matrix} (1-\alpha a^2)^{-1/2} & 0 & 0 & 0 & (a/R)(1-\alpha a^2)^{-1/2} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ (a/R)(1-\alpha a^2)^{-1/2} & 0 & 0 & 0 & (1-\alpha a^2)^{-1/2} \end{matrix}\right\|. \tag{4.15}$$

The associated coordinate transformation in the anti-de Sitter bag reads

$$\left\|\begin{matrix} x' \\ y' \\ z' \\ \sin(t'/R) \\ \cos(t'/R) \end{matrix}\right\| = \left\{\sin^2\frac{t}{R} + \frac{(\cos(t/R)+\alpha a x)^2}{1-\alpha a^2}\right\}^{-1/2} \left\|\begin{matrix} \dfrac{x+a\cos(t/R)}{\sqrt{1-\alpha a^2}} \\ y \\ z \\ \sin(t/R) \\ \dfrac{\cos(t/R)+\alpha a x}{\sqrt{1-\alpha a^2}} \end{matrix}\right\|. \tag{4.16}$$

The tensor $\widetilde{F}_{ab}$ transforms according to

$$\widetilde{F}'_{ab} = \eta_{ac}\eta_{bd}\Lambda^c{}_e\Lambda^d{}_f\eta^{eg}\eta^{fh}\widetilde{F}_{gh}. \tag{4.17}$$

For the specific $\widetilde{F}$ of (4.10) under the transformation (4.15) we find

$$\widetilde{F}'_{12} = \frac{a}{R}\mathscr{A}\frac{\xi^4}{R}\xi^2,$$

$$\widetilde{F}'_{13} = \frac{a}{R}\mathscr{A}\frac{\xi^4}{R}\xi^3,$$

$$\tilde{F}'_{14} = -\mathscr{A}\left\{\frac{\xi^5}{R}\xi^1 + \frac{a\rho^2}{R^2}\right\},$$

$$\tilde{F}'_{15} = \mathscr{A}(1-\alpha a^2)^{1/2}\frac{\xi^4}{R}\xi^1,$$

$$\tilde{F}'_{23} = 0,$$

$$\tilde{F}'_{24} = -\mathscr{A}(1-\alpha a^2)^{1/2}\frac{\xi^5}{R}\xi^2, \qquad (4.18)$$

$$\tilde{F}'_{25} = \mathscr{A}\frac{\xi^4}{R}\xi^2,$$

$$\tilde{F}'_{34} = -\mathscr{A}(1-\alpha a^2)^{1/2}\frac{\xi^5}{R}\xi^3,$$

$$\tilde{F}'_{35} = \mathscr{A}\frac{\xi^4}{R}\xi^3,$$

and

$$\tilde{F}'_{45} = -\mathscr{A}\left\{\frac{a\xi^5}{R^2}\xi^1 + \frac{\rho^2}{R}\right\},$$

where $\mathscr{A}$ is defined by

$$\mathscr{A} = (q/4\pi)(1-\alpha a^2)^{-1/2}\rho^{-3}. \qquad (4.19)$$

Next we can express all this in the transformed coordinates (4.16). Finally we come to $F'_{\mu\nu}$ in the primed coordinates via (2.21b). If we then forget about all the primes, we have constructed the solution of (3.10) for an electric charge moving along a geodesic (in this case oscillating along the x axis). With (2.21b), (4.18), and (4.19) we obtain

$$F_{12} = \mathscr{B}(ay/R)\sin(t/R),$$

$$F_{13} = \mathscr{B}(az/R)\sin(t/R),$$

$$F_{23} = 0, \qquad (4.20)$$

$$F_{01} = \mathscr{B}\{x - a\cos(t/R) + \alpha a(y^2+z^2)\cos(t/R)\},$$

$$F_{02} = \mathscr{B}y\{1 - \alpha ax\cos(t/R)\},$$

and

$$F_{03} = \mathscr{B}z\{1 - \alpha ax\cos(t/R)\},$$

where $\mathscr{B}$ is defined by

$$\mathscr{B} = \mathscr{A}(1-\alpha r^2)^{-2}$$
$$= (q/4\pi)\{(1-\alpha r^2)(1-\alpha a^2)\}^{-1/2} \times \left\{\frac{(x-a\cos(t/R))^2}{1-\alpha a^2} + y^2 + z^2\right\}^{-3/2}. \qquad (4.21)$$

From (4.20) any physical quantity of interest can be derived. Let us, for example, consider the **D** and **B** fields of the oscillating charge:

$$\mathbf{D}(\mathbf{x},t) = \mathscr{B}(1-\alpha r^2)^{1/2}(x - a\cos(t/R),y,z), \qquad (4.22)$$

and

$$\mathbf{B}(\mathbf{x},t) = \mathscr{B}(a/R)\sin(t/R)(0,z,-y). \qquad (4.23)$$

We see that the **D** field at any instant and at all places inside the cavity points straight outward from the position of the charge at that instant and that the **B** field is circular around the x axis.

Another interesting quantity is the Poynting vector $\mathbf{s}(\mathbf{x},t)$, which describes the energy flow at each point,

$$\mathbf{s}(\mathbf{x},t) = \mathbf{E}(\mathbf{x},t)\times\mathbf{H}(\mathbf{x},t) = \mathscr{B}^2(1-\alpha r^2)^{1/2}(a/R)\sin(t/R) \times\{-(y^2+z^2)(1-\alpha ax\cos(t/R)), y\{x - a\cos(t/R) + \alpha(y^2+z^2)a\cos(t/R)\}, z\{x - a\cos(t/R) + \alpha(y^2+z^2)a\cos(t/R)\}\}, \qquad (4.24)$$

with

$$\mathbf{H} = (1-\alpha r^2)^{1/2}\mathscr{B}(a/R)\sin(t/r)(0,z,-y).$$

From (4.24) we see that at the surface of the bag the Poynting vector tends to infinity. However, its radial component tends to zero, so there is no net outward or inward energy flux.

V. SOME PROPERTIES OF THE DIELECTRIC MEDIUM

In order to demonstrate that the properties of the electromagnetic constants of the dielectric medium (3.8) do not lead to unexpected or even unphysical results, we might study, for instance, the velocity of gluons. First we recall that the dielectric tensor (3.8) for the medium described by the static metric (2.1) is equal to the magnetic susceptibility tensor for the same medium as shown in Ref. 1. The velocity of the massless gauge particles (in this case called gluons, although this is not totally correct since we study only the Abelian case without color degrees of freedom) is in classical electromagnetism given by the Fresnel equation.[27] This is in general a very complicated equation, but we might use a local approximation, which for the medium (3.8) reads

$$\{1 - (\delta_{ij} - \alpha x^i x^j)n_i n_j\}^2 = 0, \qquad (5.1)$$

where $\mathbf{n} = \mathbf{k}/\omega$, $\mathbf{k}$ being the wave vector and ω the frequency. For a gluon that moves on the z axis ($x^1 = x^2 = 0$) in the direction of the z axis ($n_1 = n_2 = 0$) we find from (5.1) the equation

$$\{1 - (1-\alpha z^2)n_z^2\}^2 = 0. \qquad (5.2)$$

Equation (5.2) is solved by

$$v_z = n_z^{-1} = (1-\alpha z^2)^{1/2}, \qquad (5.3)$$

which describes a harmonically oscillating gluon along the z axis, which just touches the surface of the sphere and has a frequency R^{-1}. For a gluon on the z axis moving perpendicular to it (e.g., $n_2 = n_3 = 0$) Eq. (5.1) reduces to

$$\{1 - n_x^2\}^2 = 0. \qquad (5.4)$$

In this case we deal with a gluon that moves with velocity 1 with respect to the inertial coordinates. In general the motion of gluons can be decomposed into three oscillations at least one of which just touches the surface of the sphere, which is precisely the motion found from the geometry (2.1).

We thus find that gluons can at most have velocity 1, which is perfectly acceptable. The prevention of gluon annihilation or creation at the surface requires special boundary conditions that guarantee total reflection.[28]

Another interesting feature of the medium can be studied from the solution for the vector field of a static charge in the center of the bag (4.4), the solution of which might be interpreted as the one "gluon" exchange potential. The behavior of this potential for small values of the distance r reads, for $q<0$,

$$V(r) \underset{r\to 0}{\rightarrow} -\frac{|q|}{4\pi r} + \frac{|q|\alpha}{8\pi} r. \tag{5.5}$$

We obtain a Coulombic part and a linear part near the origin. This is precisely the form of the potential that is used in many nonrelativistic quark models.[17,18] In our description of a bare hadron, this potential comes out as a first-order correction to the harmonic oscillator type of force, which still dominates at intermediate quark distances.

VI. SUMMARY

In this paper we have set up the machinery that allows further calculations in the anti-de Sitter bag. The mathematical elegance of an equivalent five-dimensional description has been shown in detail. We have derived and solved the classical electromagnetic field equations for moving charges inside a medium with the complicated electromagnetic properties of an SO(3,2) symmetric curved space in the central projection. The results are exciting and might stimulate the development of perturbation theory in this strongly curved background field.

ACKNOWLEDGMENTS

We wish to thank the members of the Department of Theoretical Physics in Nijmegen for many stimulating discussions. One of us (E.v.B.) wishes to thank the Department of Theoretical Physics of the University of Pennsylvania for the hospitality during his stay in Philadelphia.

APPENDIX: THE MAXWELL EQUATIONS IN CURVED SPACE

In a curved space the inhomogeneous Maxwell equations read

$$[\sqrt{-g(x)}F^{\mu\nu}(x)]_{,\mu} = \sqrt{-g(x)}J^{\nu}(x). \tag{A1}$$

In flat inertial coordinates the similar equations read

$$\nabla\cdot\mathbf{D}(\mathbf{x},t) = J^0(\mathbf{x},t)$$

and

$$\nabla\times\mathbf{H}(\mathbf{x},t) - \frac{\partial D(\mathbf{x},t)}{\partial t} = J(\mathbf{x},t). \tag{A2}$$

If we identify the conserved currents $J^{\mu}(\mathbf{x},t)$ from (A2) and $\sqrt{-g(x)}J^{\mu}(x)$ from (A1), we obtain the following identities

$$D_i(\mathbf{x},t) = \sqrt{-g(x)}F^{0i}(x)$$

and

$$H_i = -\tfrac{1}{2}\sqrt{-g(x)}\epsilon_{ijk}F^{jk}(x). \tag{A3}$$

The homogeneous Maxwell equations read (i) in curvilinear coordinates

$$F_{\mu\nu,\sigma}(x) + F_{\nu\sigma,\mu}(x) + F_{\sigma\mu,\nu}(x) = 0, \tag{A4}$$

and (ii) in flat coordinates

$$\nabla\cdot\mathbf{B}(\mathbf{x},t) = 0$$

and

$$\nabla\times\mathbf{E}(\mathbf{x},t) + \frac{\partial\mathbf{B}(\mathbf{x},t)}{\partial t} = 0. \tag{A5}$$

Comparison of (A4) with (A5) leads to the identifications

$$E_i(\mathbf{x},t) = F_{0i}(x)$$

and

$$B_i(\mathbf{x},t) = -\tfrac{1}{2}\epsilon_{ijk}F_{jk}(x). \tag{A6}$$

One might easily check that the identifications (A3) and (A6) lead to the identifications of the Lagrangians (i) in curvilinear coordinates

$$\mathscr{L}(x) = -\tfrac{1}{4}\sqrt{-g(x)}F^{\mu\nu}(x)F_{\mu\nu}(x), \tag{A7}$$

and (ii) in flat coordinates

$$\mathscr{L}(\mathbf{x},t) = \tfrac{1}{2}\{\mathbf{E}(\mathbf{x},t)\cdot\mathbf{D}(\mathbf{x},t) - \mathbf{H}(\mathbf{x},t)\cdot\mathbf{B}(\mathbf{x},t)\}. \tag{A8}$$

With (A6) and (A3) we obtain the relations between $\mathbf{E}(\mathbf{x},t)$, $\mathbf{B}(\mathbf{x},t)$, $\mathbf{D}(\mathbf{x},t)$, and $\mathbf{H}(\mathbf{x},t)$:

$$\begin{aligned} D_i(\mathbf{x},t) = -\sqrt{-g(x)}\{ & g^{00}(x)g^{ij}(x)E_j(\mathbf{x},t) \\ & - g^{0j}(x)g^{i0}(x)E_j(\mathbf{x},t) \\ & - \tfrac{1}{2}g^{0j}(x)g^{ik}(x)\epsilon_{jkl}B_l(\mathbf{x},t)\} \end{aligned} \tag{A9}$$

and similarly for $\mathbf{H}(\mathbf{x},t)$.

In the lowest-order approximation of the strong interaction Lagrangian, which is given in Ref. 7, we have chosen the stationary metric $g(\mathbf{x},t) = g(\mathbf{x})$, for which $g_{i0}(x) = g_{0i}(x) = 0$. In that case (A9) reduces to

$$D_i(\mathbf{x},t) = -\sqrt{-g(\mathbf{x})}g^{00}(\mathbf{x})g^{ij}(\mathbf{x})E_j(\mathbf{x},t), \tag{A10}$$

and for $\mathbf{H}(\mathbf{x},t)$ and $\mathbf{B}(\mathbf{x},t)$ we obtain, similarly,

$$B_i(\mathbf{x},t) = -\sqrt{-g(\mathbf{x})}g^{00}(\mathbf{x})g^{ij}(\mathbf{x})H_j(\mathbf{x},t). \tag{A11}$$

So if we interpret the metric description of Ref. 7 in terms of the color electromagnetic properties of the hadronic medium, then we obtain for the color dielectric constant $\epsilon_{ij}(\mathbf{x})$ and the magnetic susceptibility $\mu_{ij}(\mathbf{x})$

$$\epsilon_{ij}(x) = \mu_{ij}(x) = -\sqrt{-g(\mathbf{x})}g^{00}(\mathbf{x})g^{ij}(\mathbf{x}). \tag{A12}$$

[1]E. van Beveren, T. A. Rijken, C. Dullemond, and G. Rupp, in *Proceedings of the Conference on "Resonances—Models and Phenomena,"* held at the Zentrum für Interdisziplinäre Forschung, Bielefeld, Germany (Springer, Berlin, 1984).

[2]E. van Beveren, T. A. Rijken, and C. Dullemond, Nijmegen report No.

THEF-NYM-84.02, 1984; which appeared in *Proceedings of the XIX Rencontre de Moriond*, edited by J. Tran Thanh Van (Editions Frontières, Gif sur Yvette, 1984), p. 331.
[3] A. Salam and J. Strathdee, Phys. Lett. B **67**, 429 (1977).
[4] J. B. Kogut and L. Susskind, Phys. Rev. D **9**, 3501 (1974).
[5] T. D. Lee, Phys. Rev. D **19**, 1802 (1979).
[6] H. B. Nielsen and A. Patkós, Nucl. Phys. B **195**, 137 (1982).
[7] C. Dullemond, T. A. Rijken, and E. van Beveren, Nuovo Cimento A **80**, 401 (1984).
[8] E. van Beveren, C. Dullemond, and T. A. Rijken, Nijmegen report No. THEF-NYM-79.11, 1979; [revised version published in Phys. Rev. D **30**, 1103 (1984)].
[9] C. Dullemond and E. van Beveren, Phys. Rev. D **28**, 1028 (1983).
[10] C. Dullemond, J. Math. Phys. **25**, 2638 (1984).
[11] C. Dullemond, T. A. Rijken, E. van Beveren, and G. Rupp, Nijmegen report No. THEF-NYM-83.09, 1983; E. van Beveren, presented at the VI Warsaw Symposium on Elementary Particle Physics, Kazimierz, Poland, 28 May–3 June 1983 (unpublished).
[12] E. van Beveren, "The influence of strong decay on the spectra of hadrons," Ph.D. thesis, University of Nijmegen, The Netherlands, 1983.
[13] A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, Phys. Rev. D **9**, 3471 (1974).
[14] N. Isgur and G. Karl, Phys. Lett. B **72**, 109 (1977); Phys. Rev. D **18**, 4187 (1978); Phys. Lett. B **74**, 353 (1978); Phys. Rev. D **19**, 2653 (1979); **20**, 1191 (1979).
[15] W. A. Bardeen, M. S. Chanowitz, S. D. Drell, M. Weinstein, and T.-M. Yan, Phys. Rev. D **11**, 1094 (1975).
[16] For many references to the literature see, e.g., A. Martin, Phys. Lett. B **93**, 338 (1980); H. Grosse and A. Martin, Phys. Rep. **60**, 341 (1980).
[17] T. Appelqvist, A. De Rujúla, S. L. Glashow, and H. D. Politzer, Phys. Rev. Lett. **34**, 365 (1975); R. Barbieri, R. Gatto, R. Kögerler, and Z. Kunzst, Phys. Lett. B **57**, 445 (1975); J. S. Kang and H. J. Schnitzer, Phys. Rev. D **12**, 841 (1975); J. Pumplin, W. Repko, and A. Sato, Phys. Rev. Lett. **35**, 1538 (1975); J. L. Richardson, Phys. Lett. B **82**, 272 (1979).
[18] E. Eichten, K. Gottfried, T. Kinoshita, and T.-M. Yan, Phys. Rev. Lett. **36**, 500 (1976); K. D. Lane and E. Eichten, Phys. Rev. Lett **37**, 477 (1976); E. Eichten, K. Gottfried, K. D. Lane, T. Kinoshita, and T.-M. Yan, Phys. Rev. D **17**, 3090 (1978); **21**, 313 (E), 203 (1980).
[19] S. Jacobs, K. J. Miller, and M. G. Olsson, Phys. Rev. Lett. **50**, 1181 (1983).
[20] N. A. Törnqvist, Phys. Rev. Lett. **49**, 624 (1982); K. Heikkilä, N. A. Törnqvist, and S. Ono, Phys. Rev. D **29**, 110 (1984).
[21] E. van Beveren, C. Dullemond, and G. Rupp, Phys. Rev. D **21**, 772 (1980); **22**, 787 (E) (1980).
[22] E. van Beveren, G. Rupp, T. A. Rijken, and C. Dullemond, Phys. Rev. D **27**, 1527 (1983).
[23] G. Rupp, "Spectra and decay properties of pseudoscalar and vector mesons in a multichannel quark model," Ph.D. thesis, University of Nijmegen, The Netherlands, 1982.
[24] S. Fubini, Nuovo Cimento A **34**, 521 (1976); V. de Alfaro, S. Fubini, and S. Furlan, Nuovo Cimento A **34**, 569 (1976).
[25] B. Binegar, C. Fronsdal, and W. Heidenreich, Ann. Phys. (NY) **149**, 254 (1983); J. Math. Phys. **24**, 2828 (1983).
[26] J. P. Gazeau, J. Math. Phys. **26**, 1847 (1985).
[27] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960), Chap. XI.
[28] C. Dullemond and E. van Beveren, J. Math. Phys. **26**, 2050 (1985).

Baker–Campbell–Hausdorff relations for supergroups

V. Alan Kostelecký[a)] and Michael Martin Nieto[b)]
Theoretical Division, Los Alamos National Laboratory, University of California, Los Alamos, New Mexico 87545

Rodney Truax[c)]
Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, University of California, Los Alamos, New Mexico 87545

(Received 30 May 1985; accepted for publication 8 January 1986)

One-parameter subgroups, the exponential mapping, and canonical coordinates have previously been defined for connected supergroups. In this paper, a technique for obtaining analytical expressions linking different coordinate schemes for supergroups is presented. These are the Baker–Campbell–Hausdorff relations. The method is illustrated in detail with the examples of supergroups based on the supersymmetric quantum-mechanical superalgebra sqm (2) and on the Inönü–Wigner contraction isop(1/2) of the simple superalgebra osp(1/2).

I. INTRODUCTION

Frequently, in applications of Lie group theory, group operators are most conveniently defined in a particular coordinate scheme that may make practical calculations difficult. For example, in the theory of coherent states, the displacement operator is most easily defined as $D(\alpha) = \exp(\alpha a^\dagger - \alpha^* a)$. However, $D(\alpha) = \exp(-\frac{1}{2}|\alpha|^2) \times \exp(\alpha a^\dagger)\exp(-\alpha^* a)$ is a more useful expression. The analytical relationship between these two forms is an example of a Baker–Campbell–Hausdorff (BCH) relation.[1]

The theory of BCH relations for Lie groups has been extensively developed.[2,3] There now exist several different techniques for obtaining such formulas. Perhaps the most widely used technique[2] invokes relationships between different exponentiations of matrix representations of the algebra associated with the Lie group. Another technique,[3] which does not rely on the existence of finite-dimensional matrix representations of the group, involves obtaining appropriate differential equations whose solution yields the desired BCH relations directly.

In contrast, little is known about the theory of BCH relations for supergroups. This may be due partly to difficulties with the earliest attempts at defining supergroups,[4] which lacked the abstract notion of a group. Subsequent efforts[5,6] did not treat the manifold structure of supergroups, although it is interesting to note that the approach of Ref. 5 was based on BCH relations. A rigorous definition of a superanalytic supermanifold was introduced by Rogers[7] in 1980. This formed the basis for a description[8] of supergroups defined both as abstract groups and as superanalytic supermanifolds. Rogers's analysis is general enough to incorporate as special limiting cases other definitions[9] of supermanifolds and supergroups. Several authors[10,11] have further extended the ideas of Refs. 7 and 8, and the approach has found applications in physics.[12] In spite of the interest aroused by Rogers's work, BCH relations have not been extensively studied for supergroups.

Our plan is to extend further the range of practical calculations that may be done using supergroups by expounding a general technique for obtaining BCH relations, applicable to the case of supergroups. The method we shall present is an extension of the approach based on differential equations[3] mentioned above. We prefer this method for supergroups, since in general it may be inappropriate to use matrix methods. For supergroups, the recurring presence of noncompact Lie subgroups means that unitary matrix representations will be infinite-dimensional. In this paper, we shall build upon a previous work[11] in which properties of one-parameter subgroups of supergroups were established and three types of canonical coordinates were defined.

In Sec. II, the basic results of Refs. 7, 8, and 11 relevant to the construction of BCH relations for supergroups are summarized. In particular, we recall the three types of canonical coordinates for supergroups. In Sec. III, a theorem relating canonical coordinates of the second and third kinds is stated and proved.

Section IV contains a derivation of some results on differentiation of the exponential mapping. These results are used in a method for obtaining BCH relations for supergroups, outlined in Sec. V.

Then, we present two examples in detail. Section VI contains derivations of various BCH relations for a connected supergroup that we denote by CSQM(2). This supergroup is based on the supersymmetric quantum-mechanical superalgebra sqm(2). We consider both canonical and noncanonical parametrizations and use them to illustrate our technique. Section VI contains an analogous discussion for a connected supergroup based on the superalgebra iosp(1/2). This superalgebra is an Inönü–Wigner contraction[13] of the simple superalgebra osp(1/2). We denote the associated connected supergroup by CIOSP(1/2).

To aid the reader, we have included an Appendix, containing a glossary of frequency used symbols. Our conventions are those of Ref. 11 and are based on those of Rogers.[7,8] In particular, the reader should note that we do *not* use the summation convention, due to the different types of summation that occur. Also, observe that a number of our theorems

---

[a)] Address after 1 September 1985: Department of Physics, Indiana University, Bloomington, Indiana 47405.
[b)] Address during academic year 1985–1986: Niels Bohr Institutet, Blegdamsvej 17, 2100-København Ø, Denmark.
[c)] On leave of absence from the Department of Chemistry, University of Calgary, Calgary, Alberta, Canada T2N 1N4.

and lemmas are carried over directly from the theory of Lie groups, due to the Lie group structure of supergroups.

II. SUMMARY OF EARLIER WORKS

Let us first recall a few definitions from Refs. 7, 8, and 11 that are pertinent for our later discussion. A more detailed discussion of the following may be found in Ref. 11, which also contains a summary of the relevant results of Rogers.[7,8] For ease of reference, in the Appendix we give a glossary of frequently used symbols.

Denote the real Grassmann algebra over **R** by B_L. This algebra has 2^L generators and may be written as the direct sum of two parts, ${}^0B_L$ and ${}^1B_L$. Elements of ${}^0B_L$, called even, commute among themselves and with elements of ${}^1B_L$. Elements of ${}^1B_L$ anticommute among themselves and are called odd.

Flat superspace, $B_L^{m,n}$, is defined to be the Cartesian product of m copies of ${}^0B_L$ with n copies of ${}^1B_L$. It can be regarded as $d = 2^{L-1}(m+n)$-dimensional vector space over **R**. An (m,n)-dimensional superanalytic supermanifold $S_L^{m,n}$ is defined to be a Hausdorff space with an atlas such that $S_L^{m,n}$ is locally homeomorphic to $B_L^{m,n}$ and the transition functions are superanalytic.

An (m,n)-dimensional supergroup H may be defined as an abstract group that is also an (m,n)-dimensional superanalytic supermanifold with a superanalytic map $H \times H \to H$: $(h_1,h_2) \to h_1 h_2^{-1}$. In analogy with the theory of Lie groups and Lie algebras, the infinitesimal left translations on the supergroup form a left B_L supermodule, which we denote by W. The basis elements of W will be written $\{X_1,\ldots, X_m, X_{m+1},\ldots,X_n\}$.

There exists a homeomorphism $\iota: B_L^{m,n} \to \mathbf{R}^d$. The supergroup H can therefore be viewed as a d-dimensional Lie group with Lie algebra **h**. It may be shown that the even part W_0 of W, viewed as a d-dimensional Lie algebra, is isomorphic to **h**. If β_μ is a basis element of B_L (see the Appendix) a basis for W_0 is $\{X_{i\mu} \mid X_{i\mu} = \beta_\mu X_i, i = 1,\ldots,m+n, \mu \in M_{L,|i|}\}$.

In standard practice, two types of canonical coordinates are delineated for Lie groups. For supergroups, we have proposed three.[11] We define canonical coordinates of the first and second kinds for supergroups in analogy with the definitions for Lie groups:

$$g_{\rm I} = \exp\left(\sum_{j=1}^{m+n} a^j X_j\right), \tag{2.1}$$

and

$$g_{\rm II} = \prod_{j=1}^{m+n} \prod_{\mu \in M_{L,|j|}} \exp(a^{j\mu}\beta_\mu X_j), \tag{2.2}$$

where $a^j \in {}^0B_L$ if $1 \leqslant j \leqslant m$ and $a^j \in {}^1B_L$ if $m+1 \leqslant j \leqslant m+n$. In these equations,

$$a^j = \sum_{\mu \in M_{L,|j|}} a^{j\mu}\beta_\mu, \quad 1 \leqslant j \leqslant m+n, \quad a^{j\mu} \in \mathbf{R}. \tag{2.3}$$

There exists a third natural parametrization for supergroups. In terms of canonical coordinates of the third kind, a supergroup element is given by

$$g_{\rm III} = \sum_{j=1}^{m+n} \exp(a^j X_j). \tag{2.4}$$

There is no canonical analog of Eq. (2.4) in standard Lie group theory, except under special conditions to be discussed later.

We shall call "normal" the sequence in which the generators X_j appear in Eqs. (2.1), (2.2), and (2.4). Other sequences are called "non-normal."

III. RELATIONSHIP BETWEEN CANONICAL COORDINATES OF SECOND AND THIRD KINDS

Having established the background for the discussions to follow, we shall now present a theorem enabling the partial reduction of canonical coordinates of the second kind of those of the third kind.

Theorem 1: Let H be an (m,n)-dimensional supergroup with left B_L supermodule W. Denote the even part of W by W_0 and the Lie algebra isomorphic to W_0 by **h**. Let

$$g_{\rm II} = \prod_{j=1}^{m+n} \prod_{\mu \in M_{L,|j|}} \exp(a^{j\mu}\beta_\mu X_j), \tag{3.1}$$

be a representation of H in terms of canonical coordinates of the second kind. Then,

$$g_{\rm II} = \prod_{j=1}^{m} \exp(a^j X_j) \times \prod_{j=m+1}^{m+n} \left\{ \exp\left[- \sum_{\mu,\nu \in M_{L,|j|}}{}'' a^{j\mu} a^{j\nu} \beta_\mu \beta_\nu \{X_j, X_j\} \right] \times \exp(a^j x_j) \right\}. \tag{3.2}$$

The restricted summation Σ'' will be defined in the proof.

Corollary: If $\{X_j, X_j\} = 0$, $m+1 \leqslant j \leqslant m+n$, then

$$g_{\rm II} = \prod_{j=1}^{m+n} \exp(a^j X_j) = g_{\rm III}, \tag{3.3}$$

i.e., canonical coordinates of the second and third kinds are identical.

To prove the theorem, it is convenient to establish the following four lemmas.

Lemma 1: For $X,Y,Z, \in \mathbf{h} \cong W_0$,

$$\exp X \exp Y = \exp Z, \tag{3.4}$$

where

$$Z = X + Y + \tfrac{1}{2}[X,Y] + \tfrac{1}{12}[X,[X,Y]] - \tfrac{1}{12}[Y,[Y,X]] + \cdots. \tag{3.5}$$

Proof: Lemma 1 holds[14] for a Lie algebra **h** with Lie group H. Since the supergroup H associated with W is also a Lie group with Lie algebra **h**, the lemma follows.

Lemma 2: Let $X,Y \in \mathbf{h} \cong W_0$. If $[X,Y] = 0$, then

$$\exp X \exp Y = \exp(X+Y) = \exp Y \exp X. \tag{3.6}$$

Aso, if

$$[X,[X,Y]] = [Y,[X,Y]] = 0, \tag{3.7}$$

then

$$\exp X \exp Y = \exp(X+Y)\exp(\tfrac{1}{2}[X,Y]) = \exp(\tfrac{1}{2}[X,Y])\exp(X+Y) \tag{3.8}$$

and

$$\exp X \exp Y = \exp Y \exp X \exp([X,Y]) = \exp([X,Y])\exp Y \exp X. \tag{3.9}$$

Proof: These relations follow from the straightforward application of Lemma 1.

Lemma 3: For $|j| = 0$, we have the result

$$[a^{j\mu}\beta_\mu X_j, a^{j\nu}\beta_\nu X_j] = 0,$$
$$1 \leqslant j \leqslant m, \quad \mu,\nu \in M_{L,0}. \tag{3.10}$$

Proof: Since $a^{j\mu}, a^{j\nu} \in \mathbf{R}$ and since $\beta_\mu, \beta_\nu \in {}^0B_L$, we can write

$$[a^{j\mu}\beta_\mu X_j, a^{j\nu}\beta_\nu X_j] = a^{j\mu}a^{j\nu}\beta_\mu \beta_\nu [X_j, X_j] = 0. \tag{3.11}$$

Lemma 4: For $|j| = 1$, we have the expressions

$$[a^{j\mu}\beta_\mu X_j, a^{j\nu}\beta_\nu X_j] = -a^{j\mu}a^{j\nu}\beta_\mu \beta_\nu \{X_j, X_j\} \tag{3.12}$$

and

$$[a^{j\mu}\beta_\mu X_j, a^{j\nu}a^{j\sigma}\beta_\nu \beta_\sigma \{X_j, X_j\}] = 0, \tag{3.13}$$

for $m+1 \leqslant j \leqslant m+n$, $\mu,\nu \in M_{L,1}$.

Proof: Since $a^{j\mu}, a^{j\nu} \in \mathbf{R}$ and since $\beta_\mu, \beta_\nu \in {}^1B_L$, we can write

$$[a^{j\mu}\beta_\mu X_j, a^{j\nu}\beta_\nu X_j] = a^{j\mu}a^{j\nu}(\beta_\mu X_j \beta_\nu X_j - \beta_\nu X_j \beta_\mu X_j)$$
$$= -a^{j\mu}a^{j\nu}\beta_\mu \beta_\nu (X_j X_j + X_j X_j) = -a^{j\mu}a^{j\nu}\beta_\mu \beta_\nu \{X_j, X_j\}. \tag{3.14}$$

For Eq. (3.13), since $a^{j\sigma} \in \mathbf{R}$ and $\sigma \in M_{L,1}$, we find that

$$[a^{j\mu}\beta_\mu X_j, a^{j\nu}a^{j\sigma}\beta_\nu \beta_\sigma \{X_j, X_j\}] = a^{j\mu}a^{j\nu}a^{j\sigma}(\beta_\mu X_j \beta_\nu \beta_\sigma \{X_j, X_j\} - \beta_\nu \beta_\sigma \{X_j, X_j\}\beta_\mu X_j)$$
$$= a^{j\mu}a^{j\nu}a^{j\sigma}\beta_\mu \beta_\nu \beta_\sigma [X_j, \{X_j, X_j\}] = 0. \tag{3.15}$$

Proof of Theorem 1: Now that Lemmas 1–4 have been proved, we can establish Theorem 1. First, we consider the even part of the product (3.1), i.e., $|j| = 0$. By Lemmas 2 and 3, we have for each $j = 1,\ldots,m$ the result

$$\prod_{\mu \in M_{L,0}} \exp(a^{j\mu}\beta_\mu X_j) = \exp\left(\sum_{\mu \in M_{L,0}} a^{j\mu}\beta_\mu X_j\right) = \exp(a^j X_j). \tag{3.16}$$

Thus,

$$\prod_{j=1}^{m} \prod_{\mu \in M_{L,0}} \exp(a^{j\mu}\beta_\mu X_j) = \prod_{j=1}^{m} \exp(a^j X_j). \tag{3.17}$$

Next, consider the case $|j| = 1$. From Eq. (3.1), the product of odd elements can be expressed as

$$\prod_{j=m+1}^{m+n} \prod_{\mu \in M_{L,|j|}} \exp(a^{j\mu}\beta_\mu X_j) = \prod_{j=m+1}^{m+n} \prod_{r=1}^{2^{L-1}} \exp(a^{j\mu^r}\beta_{\mu^r} X_j). \tag{3.18}$$

In this equation, $\mu^r \in M_{L,1}$, $1 \leqslant r \leqslant 2^{L-1}$, i.e., we allow r to count all the odd sequences in M_L, of which there are 2^{L-1}. Consider the product over r in Eq. (3.18). We must provide a method for ordering the sequences in $M_{L,1}$. We adopt the following conventions. Recall from Ref. 7 that associated with any nonempty sequence $\mu = (\mu_1,\ldots,\mu_p) \in M_L$ is a unique integer $z(\mu)$ given by

$$Z(\mu) = \tfrac{1}{2}(2^{\mu_1} + 2^{\mu_2} + \cdots + 2^{\mu_p}). \tag{3.19}$$

Then, we write $\mu^r < \mu^s$ or $r < s$ if and only if $z(\mu^r) < z(\mu^s)$, where μ^r and μ^s are any nonequivalent sequences in M_L. Since $M_{L,1}$ is a subset of M_L, we can order the odd sequences using the same device.

Expanding the product over r and using Lemmas 2–4, we obtain

$$\prod_{r=1}^{2^{L-1}} \exp(a^{j\mu^r}\beta_{\mu^r} X_j) = \exp(a^{j\mu^1}\beta_{\mu^1} X_j)\exp(a^{j\mu^2}\beta_{\mu^2} X_j) \prod_{r=3}^{2^{L-1}} \exp(a^{j\mu^r}\beta_{\mu^r} X_j)$$
$$= \exp(-a^{j\mu^1}a^{j\mu^2}\beta_{\mu^1}\beta_{\mu^2}\{X_j, X_j\})\exp\left(\sum_{r=1}^{2} a^{j\mu^r}\beta_{\mu^r} X_j\right) \prod_{r=3}^{2^{L-1}} \exp(a^{j\mu^r}\beta_{\mu^r} X_j). \tag{3.20}$$

Proceeding iteratively, and by repeated applications of Lemmas 2–4, we find

$$\prod_{r=1}^{2^{L-1}} \exp(a^{j\mu^r}\beta_{\mu^r} X_j) = \exp\left(-\sum_{r<s}^{2^{L-1}} a^{j\mu^r}a^{j\mu^s}\beta_{\mu^r}\beta_{\mu^s}\{X_j, X_j\}\right)\exp(a^j X_j). \tag{3.21}$$

Combining Eqs. (3.17) and (3.21), we get the desired result

$$\prod_{j=1}^{m+n} \prod_{\mu \in M_{L,|j|}} \exp(a^{j\mu}\beta_\mu X_j) = \prod_{j=1}^{m} \exp(a^j X_j) \prod_{j=m+1}^{m+n} \left\{\exp\left(-\sum_{\mu,\nu \in M_{L,|j|}}'' a^{j\mu}a^{j\nu}\beta_\mu\beta_\nu\{X_j, X_j\}\right)\exp(a^j X_j)\right\},$$

where the restricted summation means that $\mu < \nu$ whenever $z(\mu) < z(\nu)$.

This completes the proof of Theorem 1.

The proof of the corollary is immediate upon substitution of $\{X_j, X_j\} = 0$ into Eq. (3.2).

At this stage, it is appropriate to remark on the order of the exponentials in the product over μ in expression (3.1). Now, when $|j| = 0$ the order is immaterial, as may be seen from Eq. (3.16). However, if $|j| = 1$, then the order of the exponentials in the product over μ is of some consequence. The ordering in Eq. (3.22) is based on Eq. (3.19). Other methods for sequencing the product are possible, but they affect at most the order of the product $\beta_{\mu'}\beta_{\mu^s}$ in the first exponential of Eq. (3.21). This produces at most a sign change for each such term. Therefore, any other choice for sequencing an odd product produces an equation of the same form as Eq. (3.2), differing only by signs.

Finally, observe that according to Eq. (3.17) and Lemma 3, canonical coordinates of the second and third kinds are identical for ordinary Lie groups. Indeed, from the corollary to Theorem 1, these coordinate schemes are distinguishable only for supergroups with $\{X_j, X_j\} \neq 0$ for some $j = m+1,...,m+n$.

IV. DIFFERENTIATION OF THE EXPONENTIAL MAPPING

In this section, we derive tools essential for our calculations. We begin with three lemmas concerning differentiation of the exponential mapping.

Lemma 5: For $X \in W_0$,

$$\frac{d}{dt}\exp(tX) = x\exp(tX) = \exp(tX)X. \tag{4.1}$$

Proof: We have, by definition,

$$\begin{aligned}\frac{d}{dt}\exp(tX) &= \lim_{s\to 0}\left\{\frac{\exp[(t+s)X] - \exp(tX)}{s}\right\}\\ &= \left\{\lim_{s\to 0}\left[\frac{\exp(sX)-1}{s}\right]\right\}\exp(tX) = X\exp(tX)\\ &= \exp(tX)\left\{\lim_{s\to 0}\left[\frac{\exp(sX)-1}{s}\right]\right\} = \exp(tX)X.\end{aligned} \tag{4.2}$$

The second and fourth equalities follow because $\exp(tX)$ is a one-parameter subgroup of the subgroup H and from the expansion[11] of $\exp(sX)$.

Lemma 6: For $x(t) \in {}^0B_L$, $X \in W_0$, we have

$$\frac{d}{dt}\exp(x(t)X) = \dot{x}(t)X\exp(x(t)X). \tag{4.3}$$

In contrast, for $x(t) \in {}^1B_L$, $X \in W_1 \equiv W - W_0$, we have

$$\frac{d}{dt}\exp(x(t)X) = \dot{x}(t)X. \tag{4.4}$$

Proof: From the expansion of $\exp x(t)X$, $x(t) \in {}^0B_L$, and $X \in W_0$, we have

$$\begin{aligned}\frac{d}{dt}\exp(x(t)X) &= \frac{d}{dt}\left(1 + x(t)X + \frac{1}{2!}x^2(t)X^2 + \cdots\right)\\ &= \dot{x}(t)X\exp(x(t)X).\end{aligned} \tag{4.5}$$

Similarly, if $x(t) \in {}^1B_L$ and $X \in W_1$, then by the expansion of $\exp x(t)X$, we obtain

$$\begin{aligned}\frac{d}{dt}\exp(x(t)X) &= \frac{d}{dt}(1 + x(t)X)\\ &= \dot{x}(t)X.\end{aligned} \tag{4.6}$$

Lemma 7: Let $x(t) \in {}^1B_L$ and $X \in W_1$. Then

$$\dot{x}(t)X\exp(-x(t)X) = \dot{x}(t)(X + \tfrac{1}{2}x(t)\{X,X\}). \tag{4.7}$$

Proof: Expanding the exponential, recalling that $xx = 0$, and realizing that $XX = \frac{1}{2}\{X,X\}$ yields immediately Eq. (4.7).

The remaining tools we shall need are the extension to Grassmann variables of two theorems from Lie group theory.

Theorem 2 (Campbell–Hausdorff): Let $X, Y \in W_0$. Then

$$\exp(X)Y\exp(-X) = Y + [X,Y] + (1/2!)[X,[X,Y]] + \cdots. \tag{4.8}$$

Proof: Define $Y(t) = \exp(tX)Y\exp(-tX)$. Then $Y(0) = Y$. Expanding $Y(t)$ in a Taylor series about $t = 0$ yields

$$\begin{aligned}Y(t) = Y(0) &+ \left.\frac{dY}{dt}\right|_{t=0} t + \frac{1}{2!}\left.\frac{d^2Y}{dt^2}\right|_{t=0} t^2 + \cdots\\ &+ \frac{1}{k!}\left.\frac{d^kY}{dt^k}\right|_{t=0} t^k + \cdots.\end{aligned} \tag{4.9}$$

However, by repeated application of Lemma 5, we find

$$\left.\frac{d^kY}{dt^k}\right|_{t=0} = [X,[X,[\cdots[X,Y]]\cdots], \tag{4.10}$$

where there are a total of k commutators on the right-hand side of Eq. (4.10). For $t = 1$, we have Eq. (4.8).

Theorem 3: Let $X, Y \in W_0 \cong \mathbf{h}$. Then,

$$\exp X\exp Y\exp(-X) = \exp[(\exp X)Y(\exp(-X))]. \tag{4.11}$$

Proof: This follows[14] from the fact that W_0 is a Lie algebra and from Theorem 2.

V. ALGORITHM FOR GENERAL BCH RELATIONS

Now, we turn to the presentation of our algorithm for obtaining formulas that express superanalytic relationships between different canonical or noncanonical coordinates. These formulas are BCH relations for supergroups.

As stated in the Introduction, although matrix techniques such as those available for Lie groups[2] are applicable to supergroups, we shall adopt an alternative approach that uses a system of first-order differential equations. This method[3] avoids potential pitfalls of the matrix representation technique, particularly those associated with infinite-dimensional unitary representations.

With Theorem 2 and Lemmas 5–7, we can outline our algorithm for obtaining BCH relations for supergroups. For simplicity, let us first assume that we are interested in the BCH relation between canonical coordinates of the first and third kinds, Eqs. (2.1) and (2.4), respectively. Thus, we seek an equation of the form

$$\exp\left(\sum_{j=1}^{m+n} a^j X_j\right) = \prod_{k=1}^{m} \exp(p^k X_k) \prod_{k=m+1}^{m+n} \exp(q^k X_k), \tag{5.1}$$

where the p^k and q^k are even and odd Grassmann variables, respectively, that are to be determined in terms of superfunctions of the a^j.

We proceed by introducing an additional parameter t so that

$$g_{\rm I}(t) = \exp\left(t \sum_{j=1}^{m+n} a^j X_j\right) = \prod_{k=1}^{m} \exp(p^k(t) X_k) \times \prod_{k=m+1}^{m+n} \exp(q^k(t) X_k) = g_{\rm III}(t), \tag{5.2}$$

where $p^k(t)$ and $q^k(t)$ are functions of t. Next, differentiate both sides with respect to t. Using Lemmas 5 and 6, we obtain

$$\begin{aligned}\left(\sum_{j=1}^{m+n} a^j X_j\right) g_{\rm I}(t) &= \dot p^1 X_1 g_{\rm III}(t) + \exp(p^1X_1)\dot p^2 X_2 \exp(-p^1X_1) g_{\rm III}(t) + \cdots + \left(\prod_{k=1}^{m}\exp(p^kX_k)\right)\dot q^1 X_{m+1} \\ &\times \exp(-q^1X_{m+1})\left(\prod_{k=m}^{1}\exp(-p^kX_k)\right) g_{\rm III}(t) + \cdots + \left(\prod_{k=1}^{m}\exp(p^kX_k)\right) \\ &\times\left(\prod_{k=1}^{n-1}\exp(q^kX_{m+k})\right)\dot q^n X_{m+n}\exp(-q^nX_{m+n})\left(\prod_{k=n-1}^{1}\exp(-q^kX_{m+k})\right)\left(\prod_{k=m}^{1}\exp(-p^kX_k)\right).\end{aligned} \tag{5.3}$$

We denote differentiation with respect to t by a dot over the quantity being differentiated. Equation (5.3) has $m+n$ terms. Note that the results of differentiating even and odd variables are different. This is a direct consequence of Eqs. (4.3) and (4.4) of Lemma 6.

We have expressed Eq. (5.3) so that both sides have as right multipliers one-parameter subgroup elements $g_{\rm I}$ or $g_{\rm III}$. However, by Eq. (5.2), $g_{\rm I} = g_{\rm III}$. Therefore, these multipliers may be eliminated from Eq. (5.3) by right multiplication with $g_{\rm III}^{-1}$. Equation (5.3) may then be further simplified by means of Theorem 2 and Lemma 7, until the superalgebra basis elements X_j appear only linearly rather than in exponentials. We shall illustrate this method in detail for specific cases in the next two sections.

Since the X_j are independent by assumption, we can proceed to collect their coefficients. In general, this results in a system of $m+n$ coupled nonlinear first-order differential equations with the following form:

$$\begin{aligned}&\dot p^1 + \sum_{l=2}^{m} F^1_l(p^1,\dots,q^n)\dot p^l + \sum_{r=1}^{n} D^1_r(p^1,\dots,q^n)\dot q^r = a^1,\\ &\sum_{l=2}^{m} F^k_l(p^1,\dots,q^n)\dot p^l + \sum_{r=1}^{n} D^k_r(p^1,\dots,q^n)\dot q^r = a^k,\\ &\quad 2\leqslant k\leqslant m,\\ &\sum_{l=2}^{m} F^s_l(p^1,\dots,q^n)\dot p^l + \sum_{r=1}^{n} D^s_r(p^1,\dots,q^n)\dot q^r = a^s,\\ &\quad 1\leqslant s\leqslant n,\end{aligned} \tag{5.4}$$

where the $m+n$ dependent variables are Grassmann valued. The initial conditions are $p^k=0$, $1\leqslant k\leqslant m$, and $q^r(0)=0$, $1\leqslant r\leqslant n$. Integration of these equations will yield expressions for $p^k(t)$ and $q^r(t)$ as superfunctions of the a^k as well as the parameter t. The desired BCH relation is then obtained by setting $t=1$ in Eq. (5.2).

It would be useful to have a general technique for solving Eqs. (5.4). Unfortunately, as is the case for real variables, such a method for constructing solutions is unavailable. Nonetheless, under certain conditions we can make statements concerning the existence and uniqueness of the solutions to Eqs. (5.4). We must assume that the matrix of the coefficients of the derivatives in (5.4) is invertible. Then, the system (5.4) may be written as

$$\begin{aligned}\dot p^k &= \sum_{l=1}^{m} \mathscr{F}^k_l(p^1,\dots,q^n)a^l + \sum_{s=1}^{n} \mathscr{F}^k_s(p^1,\dots,q^n)a^s, \quad 1\leqslant k\leqslant m,\\ \dot q^r &= \sum \mathscr{D}^r_l(p^1,\dots,q^n)a^l + \sum_{s=1}^{n} \mathscr{D}^r_s(p^1,\dots,q^n)a^s, \quad 1\leqslant r\leqslant n.\end{aligned} \tag{5.5}$$

Assuming this is possible, we can then proceed by decomposing the Grassmann variables appearing in the differential equations in terms of the basis β_μ. Thus, we write

$$\begin{aligned}a^k &= \sum_{\mu\in M_{L,|k|}} a^{k\mu}\beta_\mu, \quad 1\leqslant k\leqslant m+n,\\ p^k(t) &= \sum_{\mu\in M_{L,|k|}} p^{k\mu}(t)\beta_\mu, \quad 1\leqslant k\leqslant m,\\ q^k(t) &= \sum_{\mu\in M_{L,|k|}} q^{k\mu}(t)\beta_\mu, \quad m+1\leqslant k\leqslant m+n.\end{aligned} \tag{5.6}$$

For each p^k or q^k, there are 2^{L-1} coefficients $p^{k\mu}$ or $q^{k\mu}$. The $m+n$ differential equations in Grassmann-valued dependent variables that are obtained from Eq. (5.5) may therefore be separated into $d = 2^{L-1}(m+n)$ coupled nonlinear *ordinary* differential equations, by collecting coefficients of the β_μ.

Once this has been done, however, we can rely on a theorem of ordinary differential equation theory that states[15] that it is always possible in principle to integrate a set of such equations, provided the analyticity conditions are satisfied and provided suitable initial conditions are imposed. Thus, we see that a BCH relation of the type in Eq. (5.2) may be found.

Note that, in the examples we shall consider, it is possible both to invert the matrix of coefficients of derivatives in Eq. (5.4) and to solve directly the resulting Grassmann differential equations.

It is possible to obtain BCH relations other than those in Eq. (5.2) by an extension of the general method. For instance, a noncanonical parametrization

$$g'_{\rm I} = \exp\left(\sum_{j=1}^{m} a^j X_j\right) \exp\left(\sum_{j=m+1}^{m+n} a^j X_j\right)$$

may be related to $g_{\rm III}$ as given in Eq. (5.2) by inserting a parameter t in front of *each* summation sign. Here, differentiation of $g'_{\rm I}$ with respect to t yields two terms. These may be written with $g'_{\rm I}$ as right multipliers, as in Eq. (5.3), and the resulting expression may be simplified by right multiplication with $g'^{-1}_{\rm III}$ and subsequent applications of Lemma 7 and Theorem 2. Collecting coefficients of the superalgebra basis elements again leads to $m+n$ coupled nonlinear first-order differential equations in Grassmann-valued dependent variables.

As we have outlined above, in principle we can relate any canonical or noncanonical form to canonical coordinates of the third kind. It is therefore possible to obtain a BCH relation between any two parametrizations of a one-parameter subgroup by passing through canonical coordinates of the third kind.

We remark, however, that in certain instances it may be simpler to pass directly from one parametrization to another. This will be the case when there is some simplification in the particular problem being considered, such as the vanishing of certain structure constants of the superalgebra. We shall see examples of this in the following sections.

VI. BCH RELATIONS FOR CSQM(2)

For our first example, we shall consider a connected supergroup based on the supersymmetric quantum-mechanical superalgebra[16] sqm(2). This superalgebra is spanned by the even generator X_1 and the two odd generators X_2 and X_3, which satisfy the graded commutation relations

$$[X_1, X_2] = [X_1, X_3] = 0, \quad \{X_2, X_3\} = X_1, \quad \{X_2, X_2\} = \{X_3, X_3\} = 0. \tag{6.1}$$

A connected supergroup, which we shall denote by CSQM(2), may be obtained by exponentiation of Eq. (6.1). As described in Ref. 11, a representation of a one-parameter subgroup of this supergroup in terms of canonical coordinates of the first kind is given by

$$T_1 = \exp(a^1X_1 + a^2X_2 + a^3X_3), \quad a^1 \in {}^0B_L, \quad a^2, a^3 \in {}^1B_L, \tag{6.2}$$

while a representation in terms of canonical coordinates of the third kind in normal sequence can be expressed by

$$T_3 = \exp(p^1X_1)\exp(p^2X_2)\exp(p^3X_3), \quad p^1 \in {}^0B_L, \quad p^2, p^3 \in {}^1B_L. \tag{6.3}$$

We shall begin by finding those p^j such that $T_1 = T_3$.

First, consider the one-parameter subgroup

$$T_1(t) = \exp t(a^1X_1 + a^2X_2 + a^3X_3), \tag{6.4}$$

and let the coordinates p^j depend on the parameter t. Thus,

$$T_3(t) = \exp(p^1(t)X_1)\exp(p^2(t)X_2)\exp(p^3(t)X_3). \tag{6.5}$$

We proceed as outlined in Sec. V, by differentiating $T_1(t) = T_3(t)$ with respect to t. By Lemmas 5 and 6, we have

$$(a^1X_1 + a^2X_2 + a^3X_3)T_1(t) = \dot p^1 X_1 T_3(t) + \exp(p^1X_1)\dot p^2 X_2 \exp(-p^2X_2)\exp(-p^1X_1)T_3(t) + \exp(p^1X_1)\exp(p^2X_2)\dot p^3 X_3 \exp(-p^3X_3)\exp(-p^2X_2)\exp(-p^1X_1)T_3(t). \tag{6.6}$$

Now, multiplying from the right by T_3^{-1} and applying Lemma 7 and Eq. (6.1), we obtain

$$a^1X_1 + a^2X_2 + a^3X_3 = \dot p^1X_1 + \exp(p^1X_1)\dot p^2X_2\exp(-p^1X_1) + \exp(p^1X_1)\exp(p^2X_2)\dot p^3X_3 \times\exp(-p^2X_2)\exp(-p^1X_1), \tag{6.7}$$

due to Eq. (6.1).

By Theorem 2, we find

$$\exp(p^1X_1)\dot p^2X_2\exp(-p^1X_1) = \dot p^2X_2, \tag{6.8}$$

using Eq. (6.1). Similarly,

$$\exp(p^1X_1)\exp(p^2X_2)\dot p^3X_3\exp(-p^2X_2)\exp(-p^1X_1) = \dot p^3X_3 - p^2\dot p^3X_1. \tag{6.9}$$

Substituting Eqs. (6.8) and (6.9) into Eq. (6.7) yields

$$a^1X_1 + a^2X_2 + a^3X_3 = (\dot p^1 - p^2\dot p^3)X_1 + \dot p^2X_2 + \dot p^3X_3. \tag{6.10}$$

Since X_1, X_2, and X_3 form an independent basis spanning the superalgebra sqm(2), we can identify coefficients of corresponding elements of the superalgebra. Therefore, we obtain

$$\dot p^1 - p^2\dot p^3 = a^1, \tag{6.11}$$

$$\dot p^2 = a^2, \tag{6.12}$$

and

$$\dot p^3 = a^3. \tag{6.13}$$

Equations (6.11)–(6.13) form an system of first-order coupled differential equations for the Grassmann variables $p^j(t)$, $j = 1,2,3$, subject to the initial conditions $p^j(0) = 0$.

Integrating Eq. (6.12) and using the initial condition $p^2(0) = 0$, we find

$$p^2(t) = a^2 t. \tag{6.14}$$

Similarly, we have

$$p^3(t) = a^3(t). \tag{6.15}$$

Finally, substituting for $p^2(t)$ and $\dot{p}^3(t)$ in Eq. (6.11) and integrating yields

$$p^1(t) = a^1 t + \tfrac{1}{2} a^2 a^3 t^2, \tag{6.16}$$

where the constant of integration vanishes due to the initial conditions.

The BCH rleation between canonical coordinates of the first and third kinds for CSQM(2) can be expressed by setting $t = 1$ in Eqs. (6.14)–(6.16). Thus, we obtain

$$\exp(a^1 X_1 + a^2 X_2 + a^3 X_3) = \exp[(a^1 + \tfrac{1}{2} a^2 a^3) X_1] \exp(a^2 X_2) \exp(a^3 X_3). \tag{6.17}$$

This is the desired result.

A canonical BCH relation may be obtained immediately by noting that Eq. (6.1) implies $[a^1 X_1, a^2 X_2 + a^3 X_3] = 0$, so that by Lemma 2

$$\exp(a^1 X_1 + a^2 X_2 + a^3 X_3) = \exp(a^1 X_1) \exp(a^2 X_2 + a^3 X_3). \tag{6.18}$$

Therefore, we have a second BCH relation:

$$\exp(a^2 X_2 + a^3 X_3) = \exp(\tfrac{1}{2} a^2 a^3 X_1) \exp(a^2 X_2) \exp(a^3 X_3). \tag{6.19}$$

Other BCH formulas.

$$\exp(a^1 X_1 + a^2 X_2 + a^3 X_3) = \exp[(a^1 - \tfrac{1}{2} a^2 a^3) X_1] \exp(a^3 X_3) \exp(a^2 X_2) \tag{6.20}$$

and

$$\exp(a^2 X_2 + a^3 X_3) = \exp(-\tfrac{1}{2} a^2 a^3 X_1) \exp(a^3 X_3) \exp(a^2 X_2), \tag{6.21}$$

follow from a similar analysis. Note that Eqs. (6.20) and (6.21) are *not* in normal sequence.

Next, we relate canonical coordinates of the second and third kinds for CSQM(2). We have immediately from Eq. (6.1), Theorem 1, and its corollary, the result

$$T_2 = T_3. \tag{6.22}$$

Therefore, for CSQM(2), representations in terms of canonical coordinates of the second and third kinds are identical. Furthermore, canonical coordinates of the first and second kinds are also related by Eq. (6.14)–(6.16) when t is set to unity.

Choosing a non-normal sequence for the product in canonical coordinates of the second kind may cause the two coordinate schemes to differ. As a simple example of this, let us restrict ourselves momentarily to B_2. Then, defining T'_2 by

$$T'_2 = \exp(a^{1\Omega} \beta_\Omega X_1) \exp(a^{2(1)} \beta_{(1)} X_2) \times \exp(a^{3(1)} \beta_{(1)} X_3) \exp(a^{2(2)} \beta_{(2)} X_2) \times \exp(a^{3(2)} \beta_{(2)} X_3) \exp(a^{1(12)} \beta_{(12)} X_1), \tag{6.23}$$

we find

$$T'_2 = \exp(a^1 X_1) \exp(a^{2(1)} \beta_{(1)} X_2) \times \exp(a^{3(1)} \beta_{(1)} X_3 \exp(a^{2(2)} \beta_{(2)} X_2) \times \exp(a^{3(2)} \beta_{(2)} X_3). \tag{6.24}$$

Since $\beta_{(1)} X_3$ and $\beta_{(2)} X_2$ are elements of a Lie algebra, we can employ Lemmas 2 and 4 to obtain the identity

$$\exp(a^{3(1)} \beta_{(1)} X_3) \exp(a^{2(2)} \beta_{(2)} X_2) = \exp(-a^{3(1)} a^{2(2)} \beta_{(12)} X_1) \exp(a^{2(2)} \beta_{(2)} X_2) \times \exp(a^{3(1)} \beta_{(1)} X_3). \tag{6.25}$$

When substituted in Eq. (6.24), this gives the relation

$$T'_2 = \exp[(a^1 - a^{3(1)} a^{2(2)} \beta_{(12)}) X_1] \exp(a^2 X_2) \exp(a^3 X_3). \tag{6.26}$$

This result can be extended to any L. Clearly, different orders in the product of T_2 will yield different relationships between canonical coordinates of the second and third kinds.

To find a BCH relation between canonical coordinates of the first and second kinds, we can exploit the relationship between canonical coordinates of the first and third kinds. For example, in B_2 we find

$$\exp(a^1 X_1 + a^2 X_2 + a^3 X_3) = \exp\{(a^\Omega \beta_\Omega + [a^{(12)} + \tfrac{1}{2}(a^{2(1)} a^{3(2)} + a^{2(2)} a^{3(1)})] \beta_{(12)}) X_1\} \times \exp(a^{2(1)} \beta_{(1)} X_2) \exp(a^{3(1)} \beta_{(1)} X_3) \exp(a^{2(2)} \beta_{(2)} X_2) \exp(a^{3(2)} \beta_{(3)} X_3), \tag{6.27}$$

from Eqs. (6.17) and (6.26).

This completes the analysis for our first example. Note that all our results are valid even in the limit where Grassmann parameters are nilpotent or are taken to zero.

VII. BCH RELATIONS FOR CIOSP(1/2)

For our second example we treat a connected supergroup, which we denote by CIOSP(1/2), that is based on the inhomogeneous superalgebra iosp(1/2). This superalgebra arises from a contraction of the simple superalgebra osp(1/2). Its increased complexity over sqm(2) presents some of the novel features associated with differential equations in Grassmann variables without overwhelming the reader with the many technical details that arise in the consideration of a more complicated superalgebra, such as osp(1/2).

First, we derive the iosp(1/2) commutation relations by an Inönü–Wigner contraction[13] of osp(1/2). Many such contractions are possible. We have selected the one that contains the Poincaré algebra in two space-time dimensions as a subalgebra, both because its commutation relations make discussions of BCH relations especially worthwhile, and because its status as "two-dimensional supersymmetry" makes it physically interesting.

Subsequently, we develop the BCH formulas relating the different parametrization schemes. We remark here that all our BCH relations will be valid even in the limits where the Grassmann parameters become zero or nilpotent.

The superalgebra osp(1/2) has graded commutation relations

$$\begin{aligned}&[X_1,X_2]=-2X_3,\quad [X_3,X_1]=X_1,\quad [X_3,X_2]=-X_2,\\&\{X_4,X_4\}=X_1,\quad \{X_5,X_5\}=X_2,\quad \{X_4,X_5\}=X_3,\\&[X_1,X_4]=0,\quad [X_2,X_4]=X_5,\quad [X_3,X_4]=\tfrac{1}{2}X_4,\\&[X_1,X_5]=-X_4,\quad [X_2,X_5]=0,\quad [X_3,X_5]=-\tfrac{1}{2}X_5,\end{aligned}\tag{7.1}$$

where X_1, X_2, and X_3 are even and X_4 and X_5 are odd. We redefine the generators as

$$\begin{aligned}&X_4=\lambda^a X_4',\quad X_5=\lambda^a X_5',\\&X_1=\lambda^b X_1',\quad X_2=\lambda^b X_2',\quad X_3=\lambda^c X_3',\end{aligned}\tag{7.2}$$

where $\lambda,a,b,c\in\mathbb{R}$. For $b=2a<0$, $c=0$, and taking the limit $\lambda\to 0$, we obtain the graded commutation relations

$$\begin{aligned}&[X_1,X_2]=0,\quad [X_3,X_1]=X_1,\quad [X_3,X_2]=-X_2,\\&\{X_4,X_4\}=X_1,\quad \{X_5,X_5\}=X_2,\quad \{X_4,X_5\}=0,\\&[X_1,X_4]=0,\quad [X_2,X_4]=0,\quad [X_3,X_4]=\tfrac{1}{2}X_4,\\&[X_1,X_5]=0,\quad [X_2,X_5]=0,\quad [X_3,X_5]=-\tfrac{1}{2}X_5.\end{aligned}\tag{7.3}$$

These are associated with the algebra that we call iosp(1/2).

We begin our discussion of the supergroup CIOSP(1/2) by deriving the BCH relations between canonical coordinates of the first and third kinds. The parametrization of T_1 is given by

$$T_1=\exp\Big(\sum_{j=1}^{5}a^jX_j\Big),\qquad a^1,a^2,a^3\in{}^0B_L,\quad a^4,a^5\in{}^1B_L.\tag{7.4}$$

For computational convenience, we consider first the relationship between T_1 and T_3', given by

$$\begin{aligned}T_3'&=\exp(p^3X_3)\exp(p^1X_1)\exp(p^2X_2)\\&\quad\times\exp(p^4X_4)\exp(p^5X_5),\\&\quad p^1,p^2,p^3\in{}^0B_L,\quad p^4,p^5\in{}^1B_L,\end{aligned}\tag{7.5}$$

where the prime indicates a non-normal sequence for the product in Eq. (7.5).

To calculate the BCH relation, consider the one-parameter subgroup of CIOSP (1/2) written as

$$T_1(t)=\exp\Big(t\sum_{j=1}^{5}a^jX_j\Big).\tag{7.6}$$

Let the parameters p^j in Eq. (7.5) depend upon t, also. Following the technique outlined in Sec. V, we set $T_1(t)=T_3'(t)$ and differentiate both sides of this equation with respect to t. By Lemmas 5 and 6, we have

$$\begin{aligned}\sum_{j=1}^{5}a^jX_j&=\dot p^3X_3+e^{p^3X_3}\dot p^1X_1e^{-p^2X_3}+e^{p^3X_3}e^{p^1X_1}\dot p^2X_2e^{-p^1X_1}e^{-p^3X_3}+e^{p^3X_3}e^{p^1X_1}e^{p^2X_2}\dot p^4X_4\\&\quad\times e^{-p^4X_4}e^{-p^2X_2}e^{-p^1X_1}e^{-p_3X_3}+e^{p^3X_3}e^{p^1X_1}e^{p^2X_2}e^{p^4X_4}\dot p^5X_5*e^{-p^5X_5}e^{-p^4X_4}e^{-p^2X_2}e^{-p^1X_1}e^{-p^3X_3},\end{aligned}\tag{7.7}$$

where we have multiplied from the right by $T_3'^{-1}(t)$. Using Lemma 7, Theorem 2, and Eq. (7.3), we find

$$\sum_{j=1}^{5}a^jX_j=\dot p^3X_3+(\dot p^1e^{p^3}+\tfrac{1}{2}\dot p^4p^4e^{p^3})X_1+(\dot p^2e^{-p^3}+\tfrac{1}{2}\dot p^5p^5e^{-p^3})X_2+\dot p^4e^{p^3/2}X_4+\dot p^5e^{-p^3/2}X_5.\tag{7.8}$$

Identifying corresponding elements of the basis of iosp(1/2) yields

$$\dot p^3=a^3,\tag{7.9}$$

$$\dot p^1e^{p^3}+\tfrac{1}{2}\dot p^4p^4e^{p^3}=a^1,\tag{7.10}$$

$$\dot p^2e^{-p^3}+\tfrac{1}{2}\dot p^5p^5e^{-p^3}=a^2,\tag{7.11}$$

$$\dot p^4e^{p^3/2}=a^4,\tag{7.12}$$

$$\dot p^5e^{-p^3/2}=a^5,\tag{7.13}$$

subject to the initial condition $p^j(0)=0, j=1,...,5$.

Equations (7.9)–(7.13) form a system of coupled first-order differential equations that can be integrated in a straightforward way. Solving (7.9), we have

$$p^3(t)=a^3t.\tag{7.14}$$

Substituting for $p^3(t)$ into Eqs. (7.12) and integrating gives the solution for p^4 as

$$p^4(t)=(2a^4/a^3)(1-e^{-a^3t/2}),\tag{7.15}$$

where the initial conditions have been applied. Equation (7.15) is also valid in the limit $a^3\to 0$. Furthermore, it is valid even in the limit where a^3 is nilpotent, as may be seen by expanding the exponential. This appears to be an example of L'Hôpital's rule as extended to Grassmann variables. Similarly, we find

$$p^5(t) = (2a^5/a^3)(e^{a^3t/2} - 1). \tag{7.16}$$

To integrate Eqs. (7.10) and (7.11), we note that[17] $p^4\dot{p}^4 = p^5\dot{p}^5 = 0$ because $a^4, a^5 \in {}^1B_L$. Therefore, Eqs. (7.10) and (7.11) become

$$\dot{p}^1 = a^1 e^{-a^3t}, \quad \dot{p}^2 = a^2 e^{a^3t}. \tag{7.17}$$

Integrating these equations and applying the initial conditions, we obtain

$$p^1(t) = (a^1/a^3)(1 - e^{-a^3t}), \tag{7.18}$$

and

$$p^2(t) = (a^2/a^3)(e^{a^3t} - 1). \tag{7.19}$$

Setting $t = 1$ in the equation $T_1 = T'_3$, we find that Eqs. (7.14)–(7.16), (7.18), and (7.19) imply the BCH relation

$$\exp\left(\sum_{j=1}^{5} a^j X_j\right) = \exp(a^3 X_3)\exp\left[\frac{a^1}{a^3}(1 - e^{-a^3})X_1\right] \times \exp\left[\frac{a^2}{a^3}(e^{a^3} - 1)X_2\right]\exp\left[\frac{2a^4}{a^3}(1 - e^{-a^3/2})X_4\right]\exp\left[\frac{2a^5}{a^3}(e^{a^3/2} - 1)X_5\right]. \tag{7.20}$$

This is the desired result.

Now, we demonstrate a means of obtaining other BCH relations from Eq. (7.20) and by judicious application of the results we have derived in Secs. III and IV.

We begin by showing that the sequence of products in Eq. (7.20) can be rearranged to the normal sequence. Consider the formulas

$$\exp(a^3X_3)\exp(b^1X_1)\exp(-a^3X_3) = \exp(b^1 e^{a^3}X_1), \tag{7.21}$$

and

$$\exp(a^3X_3)\exp(b^2X_2)\exp(-a^3X_3) = \exp(b^2 e^{-a^3}X_2), \tag{7.22}$$

both of which follow from Theorem 3. With these, and by inserting the identity element

$$\exp(-a^3X_3)\exp(a^3X), \tag{7.23}$$

in the appropriate positions in Eq. (7.20), we arrive at the expression

$$\exp\left(\sum_{j=1}^{5} a^j X_j\right) = \exp\left[\frac{a^1}{a^3}(e^{a^3} - 1)X_1\right]\exp\left[\frac{a^2}{a^3}(1 - e^{-a^3})X_2\right] \times \exp(a^3X_3)\exp\left[\frac{2a^4}{a^3}(1 - e^{-a^3/2})X_4\right]\exp\left[\frac{2a^5}{a^3}(e^{a^3/2} - 1)X_5\right]. \tag{7.24}$$

This agrees with the BCH formula computed by repeating the algorithm with a sequential product replacing the product in Eq. (7.5).

To relate canonical coordinates of the second and third kinds, we apply Theorem 1. For the non-normal sequences, this gives

$$\prod_{\alpha\in M_{L,0}} \exp(a^{3\alpha}\beta_\alpha X_3) \prod_{\gamma\in M_{L,0}} \exp(a^{1\gamma}\beta_\gamma X_1) \prod_{\epsilon\in M_{L,0}} \exp(a^{2\epsilon}\beta_\epsilon X_2) \prod_{\kappa\in M_{L,0}} \exp(a^{4\kappa}\beta_\kappa X_4) \prod_{\lambda\in M_{L,0}} \exp(a^{5\lambda}\beta_\lambda X_5)$$
$$= \exp(a^3X_3)\exp(a^1X_1)\exp(a^2X_2)\exp(x^4X_1)\exp(a^4X_4)\exp(x^5X_5)\exp(a^5X_5), \tag{7.25}$$

where

$$x^k = -\sum_{\mu,\nu\in M_{L,1}}{}'' a^{j\mu}a^{j\nu}\beta_\mu\beta_\nu, \tag{7.26}$$

for $k = 4,5$. Since $[X_1, X_2] = [X_2, X_4] = 0$, we can rearrange Eq. (7.25) using Lemma 2 to yield

$$T'_2 = \exp(a^3X_3)\exp[(a^1 + x^4)X_1] \times \exp[(a^2 + x^5)X_2]\exp(a^4X_4)\exp(a^5X_5). \tag{7.27}$$

This is now in the form of canonical coordinates of the third kind, in non-normal sequence.

However, if we choose to adopt, instead, the normal sequence

$$T_2 = \prod_{j=1}^{5}\prod_{\mu\in M_{L,|j|}} \exp(a^{j\mu}\beta_\mu X_j), \tag{7.28}$$

then Theorem 1 gives

$$T_2 = \prod_{j=1}^{3} \exp(a^jX_j)\exp(x^4X_1)\exp(a^4X_4) \times \exp(x^5X_2)\exp(a^5X_5), \tag{7.29}$$

where x^4 and x^5 are given by Eq. (7.26). Since $[X_2, X_4] = 0$, we can transform Eq. (7.29) to the form

$$T_2 = \exp(a^1X_1)\exp(a^2X_2)\exp(a^3X_3)$$
$$\times\exp(x^4X_1)\exp(x^5X_2)\exp(a^4X_4)\exp(a^5X_5). \quad (7.30)$$

By inserting the identity element (7.23) in the appropriate places in Eq. (7.30) and by making use of Eqs. (7.21) and (7.22), we have

$$T_2 = (a^1X_1)\exp(a^2X_2)\exp(x^4e^{a^3}X_1)$$
$$\times\exp(x^5e^{-a^3}X_2)\exp(a^3X_3)$$
$$\times\exp(a^4X_4)\exp(a^5X_5). \quad (7.31)$$

Since $[X_{12}, X_2] = 0$, by Lemma 2 we find

$$T_2 = \exp[(a^1 + x^4e^{a^3})X_1]\exp[(a^2 + x^5e^{-a^3})X_2]$$
$$\times\exp(a^3X_3)\exp(a^4X_4)\exp(a^5X_5). \quad (7.32)$$

This is an expression in terms of canonical coordinates of the third kind in normal sequence.

Using expressions (7.27) and (7.32), it is possible to obtain a BCH relation between T_2' and T_2, i.e., we can calculate the conditions that permit the equality between the two parametrizations. Transforming T_2' in Eq. (7.27) into normal sequence, we obtain

$$T_2' = \exp[(a^1 + x^4)e^{a^3}X_1]\exp[(a^2 + x^5)e^{-a^3}X_2]$$
$$\times\exp(a^3X_3)\exp(a^4X_4)\exp(a^5X_5) = T_2$$
$$= \exp[(b^1 + y^4e^{b^3})X_1]\exp[(b^2 + y^5e^{-b^3})X_2]$$
$$\times\exp(b^3X_3)\exp(b^4X_4)\exp(b^5X_5). \quad (7.33)$$

Here, y^4 and y^5 are given by Eq. (7.26), except that the $a^{j\mu}$ are to be replaced by $b^{j\mu}$. Equation (7.33) implies

$$(a^1 + x^4)e^{a^3} = b^1 + y^4e^{b^3},$$
$$(a^2 + x^5)e^{-a^3} = b^2 + y^5e^{-b^3}, \quad (7.34)$$
$$a^3 = b^3, \quad a^4 = b^4, \quad a^5 = b^5.$$

It follows from these equations and Eq. (7.26) that $y^4 = x^4$, $y^5 = x^5$, and hence

$$b^1 = a^1e^{a^3}, \quad b^2 = a^2e^{-a^3}. \quad (7.35)$$

Thus, Eq. (7.35) and the last three identities in (7.34) provide conditions for which $T_2' = T_2$. Note that, for arbitrary L, converting T_2' directly into the normal sequence T_2 given by Eq. (7.28) would be tedious. It can be accomplished most economically via conversion to canonical coordinates of the third kind, Eq. (7.32), as we have shown.

Finally, it is expedient to use coordinates of the third kind to give expressions for BCH relations between canonical coordinates of the first and second kinds.

We begin by finding the conditions on b such that $T_1 = T_2'$, where the parametrizations are given by Eqs. (7.4) and (7.25). We do this through the BCH relations (7.20) and (7.27). This results in the identifications

$$(a^1/a^3)(1 - e^{-a^3}) = b^1 + y^4,$$
$$(a^2/a^3)(e^{a^3} - 1) = b^2 + y^5,$$
$$a^3 = b^3, \quad (7.36)$$
$$(2a^4/a^3)(1 - e^{-a^3/2}) = b^4,$$
$$(2a^5/a^3)(e^{a^3/2} - 1) = b^5.$$

Next, we repeat the procedure to find conditions such that $T_1 = T_2$, where the parametrizations are given by Eqs. (7.4) and (7.28). Using the BCH relations (7.24) and (7.32) yields the criteria for equality as

$$(a^1/a^3)(e^{a^3} - 1) = b^1 + y^4e^{b^3},$$
$$(a^2/a^3)(1 - e^{-a^3}) = b^2 + y^5e^{-b^3},$$
$$a^3 = b^3, \quad (7.37)$$
$$(2a^4/a^3)(1 - e^{-a^3/2}) = b^4,$$
$$(2a^5/a^3)(e^{a^3/2} - 1) = b^5.$$

In Eqs. (7.36) and (7.37), we solve for the components of b^4 and b^5 and then compute y^4 and y^5. With the resulting expressions, we can find the components of b^1 and b^2.

ACKNOWLEDGMENTS

The authors acknowledge the support of the United States Department of Energy. D.R.T. is grateful for the financial support of the Natural Sciences and Engineering Reseach Council of Canada.

APPENDIX: GLOSSARY OF FREQUENTLY USED SYMBOLS

This glossary serves to define frequently used symbols. We adhere as closely as possible to the conventions of Rogers.[7,8] More detailed discussions of quantities listed may be found in Refs. 7, 8, and 11.

B_L	real Grassmann algebra with L generators
M_L	set of integer sequences, with $\mu = (\mu_1,\ldots,\mu_k)$, $1 \leqslant \mu_1 < \mu_2 < \cdots < \mu_k \leqslant L$, $1 \leqslant k \leqslant L$; Ω is the null sequence
β_μ	basis element in B_L; $\beta_\Omega = 1$
$\lvert i \rvert$	$\lvert i \rvert = 0$ for even elements, $\lvert i \rvert = 1$ for odd elements
$M_{L,0}$ ($M_{L,1}$)	the subset of M_L with all elements even (odd)
$B_L^{m,n}$	"flat superspace" the Cartesian product of m copies of the even part, ${}^0B_L$, of B_L with n copies of the odd part, ${}^1B_L$
$d = 2^{L-1}(m+n)$	dimension of $B_L^{m,n}$ viewed as a vector space over $\mathbb{R}$
H	supergroup; also, supergroup viewed as a Lie group
W	left B_L supermodule
X_j	generators of W; in general, $1 \leqslant j \leqslant m+n$
$\mathbf{h}$	Lie algebra with Lie group H
W_0	even part of W, isomorphic to $\mathbf{h}$
$X_{j\mu} = \beta_\mu X_j$	generators of W_0
x	element in $B_L^{m,n}$; note that for convenience we have dropped the tilde under x used in Ref. 11

[1]J. E. Campbell, Proc. London Math. Soc. (1) **28**, 381 (1897); H. F. Baker, Proc. London Math. Soc. (1) **34**, 347 (1902); **35**, 333 (1903); (2) **2**, 293 (1905); F. Hausdorff, Ber. Verh. Saechs. Akad. Wiss. Leipzig Math. Phys. Kl. **58**, 19 (1906).
[2]See, for example, R. M. Wilcox, J. Math. Phys. **8**, 962 (1967); B. Mielnik, Ann. Inst. H. Poincaré A **12**, 215 (1970); R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).
[3]D. R. Truax, Phys. Rev. D **31**, 1988 (1985).
[4]F. A. Berezin and G. I. Kac. Math. USSR-Sb. **11**, 311 (1970); B. Kostant, in *Lecture Notes in Mathematics*, Vol. 570 (Springer, New York, 1977), p. 617.
[5]A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).
[6]V. Rittenberg, in *Lecture Notes in Physics*, Vol. 79 (Springer, Berlin, 1978), p. 3; F. A. Berezin and V. N. Tolstoy, Commun. Math. Phys. **78**, 409 (1981); see also I. Bars, in *Supersymmetry in Physics*, edited by V. A. Kostelecký and D. K. Campbell (North-Holland, Amsterdam, 1985), p. 42 [also published as Physica D **15**, 42 (1985)].
[7]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[8]A. Rogers, J. Math. Phys. **22**, 939 (1981).
[9]F. A. Berezin and D. A. Leites, Sov. Math. Dokl. **16**, 1218 (1975); M. Batchelor, Trans. Am. Math. Soc. **253**, 329 (1979); **258**, 257 (1980).
[10]A. Rogers, J. Math. Phys. **22**, 443 (1981); **26**, 385 (1985); C. P. Boyer and S. Gitler, Trans. Am. Math. Soc. **258**, 241 (1984); J. Hoyos, M. Quiros, J. Ramirez Mittelbrunn, and F. J. de Urries, J. Math. Phys. **25**, 833, 841, 847 (1984); J. M. Rabin and L. Crane, Commun. Math. Phys. **102**, 123 (1985).
[11]D. R. Truax, V. A. Kostelecký, and M. M. Nieto, J. Math. Phys. **27**, 354 (1986).
[12]V. A. Kostelecký, Nucl. Phys. B **219**, 167 (1983); V. A. Kostelecký and J. M. Rabin, J. Math. Phys. **25**, 2744 (1984).
[13]E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[14]See, for example, W. Miller, Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972).
[15]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[16]This superalgebra has recently enjoyed great popularity. See, for example, E. Witten, Nucl. Phys. B **185**, 513 (1981); M. de Crombrugghe and V. Rittenberg, Ann. Phys. (NY) **151**, 99 (1983); P. Salomonson and J. W. van Holten, Nucl. Phys. B **196**, 509 (1982); F. Cooper and B. Freedman, Ann. Phys. (NY) **146**, 262 (1983); F. Cooper, B. Freedman, and C. M. Bender, Nucl. Phys. B **219**, 61 (1983); D. Lancaster, Nuovo Cimento A **79**, 28 (1984); M. M. Nieto, Phys. Lett. B **145**, 208 (1984). Proposed physical applications are discussed in L. F. Urrutia and E. Hernandez, Phys. Rev. Lett. **51**, 755 (1983); V. A. Kostelecký and M. M. Nieto, Phys. Rev. Lett. **53**, 2285 (1984).
[17]We remark that this is not true in general. Although $pp = 0$ for $p \in {}^1B_L$, $p\dot{p}$ may be nonvanishing.

Symmetry breaking of $u(6/2j+1)$ supersymmetric models

M. Baake and P. Reinicke
Universität Bonn, Physikalisches Institut, Nussallee 12, D-5300 Bonn 1, West Germany

(Received 1 October 1985; accepted for publication 11 December 1985)

In this paper, the group theory of models with broken $u(6/2j+1)$ supersymmetry described by the chain $u(6/2j+1) \supset u_B(6) \times u_F(2j+1) \supset u_B(6) \times sp_F(2j+1) \supset \cdots \supset so_B(3) \times su_F(2) \supset su_{B+F}(2)$, which has recently been suggested for application to nuclear physics, is presented. All invariants that are needed for the construction of the general Hamiltonian for this model are presented.

I. INTRODUCTION

Recently, a simple supersymmetric Hamiltonian was proposed that describes the ground states and collective excitations in a region of nuclei with high-spin, single-j orbitals.[1] Its physical relevance was shown by an explicit application to the isotopes of xenon.[1] The corresponding symmetry breaking scheme is given by the following chain of algebras:

$$u(6/2j+1) \supset u_B(6) \times u_F(2j+1) \supset u_B(6) \times sp_F(2j+1) \supset u_B(6) \times su_F(2)\,. \tag{1.1}$$

For the bosonic algebra $u_B(6)$ one can now insert any of the three possible limiting cases of the interacting boson approximation (IBA[2,3]),

$$\begin{aligned} &(\mathrm{I})\ \ u_B(6) \supset u_B(5) \supset so_B(5) \supset so_B(3)\,,\\ &(\mathrm{II})\ \ u_B(6) \supset so_B(6) \supset so_B(5) \supset so_B(3)\,,\\ &(\mathrm{III})\ \ u_B(6) \supset su_B(3) \supset so_B(3)\,. \end{aligned} \tag{1.2}$$

In a last step, the angular momenta of bosons and fermions are combined to total angular momentum,

$$so_B(3) \times su_F(2) \supset su_{B+F}(2)\,, \tag{1.3}$$

since this gives the best quantum number of the chain.

In the approach of Ref. 1, the Hamiltonian has been restricted to a linear combination of the linear and quadratic Casimir operators of the symmetry chain. Although this has the advantage that the Hamiltonian can be diagonalized analytically, there are, of course, further scalars that may in general play an important role. Most of these additional invariants clearly do not commute with all the Casimir operators of the chain but there is *a priori* no reason to omit them.

In what follows, we present the full set of Hermitian scalars that are at most quartic in the creation and annihilation operators and conserve the total particle number. This automatically leads to the most general Hamiltonian including two (quasi-) particle effective interactions. The invariants presented below are arranged so that it is impossible to obtain—for any given algebra of the chain—an additional, linearly independent scalar by taking a certain combination of the scalars of the subalgebras.

The paper is organized as follows. After some preliminaries in Sec. II, where the generators and the Casimir operators of the considered algebras are given explicitly, we give the invariants for each algebra in Sec. III. The result can be seen in the Tables III and IV together with Eqs. (3.3)–(3.5). Then, in Sec. IV, we give the matrix elements of all $so_B(5) \times sp_F(2j+1)$ scalars.

II. NOTATION AND PRELIMINARIES

The fermionic degrees of freedom of the chain (1.1) are denoted by a_μ, $\mu = -j,\ldots,j$, where $\{a_\mu, a_\nu^+\} = \delta_{\mu\nu}$ and the corresponding su(2) tensor operators are given by a_μ^+ and $\tilde{a}_\mu := (-1)^{j-\mu} a_{-\mu}$, $\mu = -j,\ldots,j$.

As usual, the s- and d-bosons are denoted by s and d_μ, $\mu = -2,\ldots,2$, respectively, with $[s,s^+] = 1$, $[d_\mu, d_\nu^+] = \delta_{\mu\nu}$, and $\tilde{d}_\mu := (-1)^\mu d_{-\mu}$. Bosonic and fermionic operators commute.

The coupling of su(2) tensor operators reads

$$(T^{(k)} \times T^{(l)})_\mu^{(m)} = \sum_\nu C_{\nu\,\mu-\nu\,\mu}^{\,k\,l\,m}\, T_\nu^{(k)}\, T_{\mu-\nu}^{(l)}\,, \tag{2.1}$$

where $C_{m_1 m_2 m}^{\,j_1 j_2 j}$ are the usual su(2) Clebsch–Gordan coefficients.

The generators and the Casimir operators of all algebras appearing in Eqs. (1.1)–(1.3) in terms of the creation and annihilation operators are given in Tables I and II. We can now turn to the construction of the invariants.

III. CONSTRUCTION OF THE GENERAL HAMILTONIAN

If all Hermitian $su_{B+F}(2)$ scalars H_i are known that are at most quartic expressions in the creation and annihilation operators and conserve the total particle number, one directly obtains the corresponding, most general Hamiltonian just as the general linear combination of these terms,

$$H = \sum_i a_i H_i\,, \tag{3.1}$$

where the coefficients a_i play the role of coupling constants.

For a given symmetry breaking chain one expects a certain hierarchy, i.e., the energy splitting due to a step down the chain must be small in comparison to the splitting of earlier steps. This behavior has of course to be tested experimentally. To this end, the scalars H_i have to be rearranged for maximal symmetry, i.e., one has to construct a full set of scalars for every member of the chain from those of $su_{B+F}(2)$. This leads to an equivalent but slightly modified Hamiltonian,

$$H = \sum_{\substack{\text{chain}\\ \text{numbers}\\ \mathscr{A}_i}} \sum_j a_j^{(i)} \cdot H_j^{(i)}\,, \tag{3.2}$$

 0022-2488/86/051430-05$02.50

TABLE I. Generators in terms of creation and annihilation operators for the algebras of Eqs. (1.1)–(1.3). The coupling of tensor operators is defined in Eq. (2.1).

Algebra	Generators
$u(6/2j+1)$	$(\alpha^+ \times \tilde{\beta})^{(k)}_\mu;\ \alpha, \beta = s,d,a$
$u_F(2j+1)$	$(a^+ \times \tilde{a})^{(k)}_\mu$
$sp_F(2j+1)$	$(a^+ \times \tilde{a})^{(k)}_\mu$, k odd
$su_F(2)$	$(a^+ \times \tilde{a})^{(1)}_\mu$
$u_B(6)$	$(\alpha^+ \times \tilde{\beta})^{(k)}_\mu;\ \alpha, \beta = s,d$
$u_B(5)$	$(d^+ \times \tilde{d})^{(k)}_\mu$
$so_B(6)$	$(d^+ \times \tilde{d})^{(k)}_\mu$, k odd, $s^+\tilde{d}_\mu + d^+_\mu s$
$so_B(5)$	$(d^+ \times \tilde{d})^{(k)}_\mu$, k odd
$su_B(3)$	$Q_\mu := s^+\tilde{d}_\mu + d^+_\mu s - (\sqrt{7}/2)(d^+ \times \tilde{d})^{(2)}_\mu;\ (d^+ \times \tilde{d})^{(1)}_\mu$
$so_B(3)$	$(d^+ \times \tilde{d})^{(1)}_\mu$
$su_{B+F}(2)$	$I_\mu := \sqrt{10}(d^+ \times \tilde{d})^{(1)}_\mu - \alpha_j (a^+ \times \tilde{a})^{(1)}_\mu$, where $\alpha_j = \{\frac{1}{3} j(j+1)(2j+1)\}^{1/2}$

where the $H^{(i)}_j$ are scalars with respect to $\mathscr{A}_i$. Now, the hierarchy can be tested by comparison of the coupling constants with different upper index i.

(a) For the presentation of the invariants, let us start with the purely *bosonic part*. There are—except for a constant—nine $so_B(3)$ scalars satisfying the conditions imposed above (a detailed analysis of these terms has been given in Ref. 3). Table III shows the invariance properties of these scalars with respect to $u_B(6)$ and its subalgebras that occur in the three chains of Eq. (1.2). As mentioned above, there are no combinations that give new invariants on a higher stage. Furthermore, one should notice that—in contrast to what is found in some discussions of the IBA—all terms can certainly be expressed by Casimir operators but that one term is not a Casimir operator itself but a product of two linear ones.

(b) For the purely *fermionic part*, one finds $j + \frac{1}{2}$ linearly independent $su_F(2)$ scalars, namely

$$\hat{N}_F \text{ and } A_k := ((a^+ \times a^+)^{(2k)} \times (\tilde{a} \times \tilde{a})^{(2k)})^{(0)}, \quad k = 0, 1, ..., j - \tfrac{1}{2}. \tag{3.3}$$

Among them there are two $u_F(2j+1)$ scalars, $\hat{N}_F$ and $C_2^{u_F(2j+1)}$ (cf. Table II), and together with $C_2^{sp_F(2j+1)}$ one already has all purely fermionic $sp_F(2j+1)$ invariants. A proof of this statement is given in Appendix A. For the remaining $su_F(2)$ scalars one can, without loss of generality, choose A_k with $k = 1, 2, ..., j - \frac{3}{2}$.

(c) In the *Bose–Fermi part* of the chain one finds the following 12 $su_{B+F}(2)$ scalars:

$$\begin{aligned} &s^+(\tilde{d} \times (a^+ \times \tilde{a})^{(2)})^{(0)} + \text{h.c.}, \\ &((d^+ \times \tilde{d})^{(k)} \times (a^+ \times \tilde{a})^{(k)})^{(0)}, \quad k = 1, 2, 3, 4, \\ &s^+(d^+ \times (\tilde{a} \times \tilde{a})^{(2)})^{(0)} + \text{h.c.}, \\ &((d^+ \times d^+)^{(k)} \times (\tilde{a} \times \tilde{a})^{(k)})^{(0)} + \text{h.c.}, \quad k = 2,4, \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} &\hat{A} := s^2 \cdot (a^+ \times a^+)^{(0)} + \text{h.c.}, \\ &\hat{B} := (\sqrt{5}\,(\tilde{d} \times \tilde{d})^{(0)} - s^2) \cdot (a^+ \times a^+)^{(0)} + \text{h.c.}, \\ &\hat{N}_B \cdot \hat{N}_F, \\ &\hat{n}_d \cdot \hat{N}_F. \end{aligned} \tag{3.5}$$

Notice that five of them do not conserve the nucleon number, which is a direct consequence of the introduction of quasiparticles. Only the four scalars of Eq. (3.5) are also invariants with respect to $so_B(3) \times su_F(2)$, and no linear combination of the operators of (3.4) would result in an additional $so_B(3) \times su_F(2)$ scalar. Table IV shows the invariance properties of the operators (3.5) at the different stages. The only $u(6/2j+1)$ scalars are indeed the Casimir

TABLE II. Linear (C_1) and quadratic (C_2) Casimir operators with their eigenvalues for the oscillator representations defined by Table I. For the notation, see Eq. (2.1); Q, I, and α_j are given in Table I.

Algebra	Linear Casimir operator	Eigenvalue
$u(6/2j+1)$	$\hat{\mathscr{N}} = \hat{N}_B + \hat{N}_F$	$\mathscr{N}$
$u_F(2j+1)$	$\hat{N}_F = -\sqrt{2j+1}\,(a^+ \times \tilde{a})^{(0)}$	N_F
$u_B(6)$	$\hat{N}_B = \hat{n}_d + s^+ s$	N_B
$u_B(5)$	$\hat{n}_d = \sqrt{5}(d^+ \times \tilde{d})^{(0)}$	n_d
Algebra	**Quadratic Casimir operator**	**Eigenvalue**
$u(6/2j+1)$	$\hat{\mathscr{N}}(\hat{\mathscr{N}} + 5 - (2j+1))$	$\mathscr{N}(\mathscr{N} + 5 - (2j+1))$
$u_F(2j+1)$	$\hat{N}_F(2j+2-\hat{N}_F)$	$N_F(2j+2-N_F)$
$sp_F(2j+1)$	$\hat{N}_F(2j+3-\hat{N}_F) + (2j+1)(a^+ \times a^+)^{(0)}(\tilde{a} \times \tilde{a})^{(0)}$	$\nu(2j+3-\nu)$
$su_F(2)$	$-\sqrt{3}\,\alpha_j^2((a^+ \times \tilde{a})^{(1)} \times (a^+ \times \tilde{a})^{(1)})^{(0)}$	$L_F(L_F+1)$
$u_B(6)$	$\hat{N}_B(\hat{N}_B + 5)$	$N_B(N_B+5)$
$u_B(5)$	$\hat{n}_d(\hat{n}_d + 4)$	$n_d(n_d+4)$
$so_B(6)$	$\hat{N}_B(\hat{N}_B + 4) - [\sqrt{5}\,(d^+ \times d^+)^{(0)} - s^+ s^+][\sqrt{5}\,(\tilde{d} \times \tilde{d})^{(0)} - ss]$	$\sigma(\sigma+4)$
$so_B(5)$	$\hat{n}_d(\hat{n}_d + 3) - 5(d^+ \times d^+)^{(0)}(\tilde{d} \times \tilde{d})^{(0)}$	$\tau(\tau+3)$
$su_B(3)$	$\frac{4}{3}\sqrt{5}\,(Q \times Q)^{(0)} - 5\sqrt{3}\,(d^+ \times \tilde{d})^{(1)} \times (d^+ \times \tilde{d})^{(1)})^{(0)}$	$\frac{2}{3}[(\lambda+\mu)(\lambda+\mu+3) - \lambda\mu]$
$so_B(3)$	$-10\sqrt{3}((d^+ \times \tilde{d})^{(1)} \times (d^+ \times \tilde{d})^{(1)})^{(0)}$	$L_B(L_B+1)$
$su_{B+F}(2)$	$-\sqrt{3}\,(I \times I)^{(0)}$	$I(I+1)$

TABLE III. Purely bosonic invariants with respect to the algebras of Eq. (1.2). The definition of the operators in terms of creation and annihilation operators can be taken from Table II. A plus (minus) sign means that the corresponding operator on the left is (is not) an invariant with respect to the corresponding algebra on top.

Operator \ Algebra	$u_B(6)$	$u_B(5)$	$so_B(6)$	$so_B(5)$	$su_B(3)$	$so_B(3)$
$C_1^{u_B(6)}$	+	+	+	+	+	+
$C_2^{u_B(6)}$	+	+	+	+	+	+
$C_1^{u_B(5)}$	−	+	−	+	−	+
$C_2^{u_B(5)}$	−	+	−	+	−	+
$C_2^{so_B(6)}$	−	−	+	+	−	+
$C_2^{so_B(5)}$	−	−	−	+	−	+
$C_2^{su_B(3)}$	−	−	−	−	+	+
$C_2^{so_B(3)}$	−	−	−	−	−	+
$C_1^{u_B(6)} \cdot C_1^{u_B(5)}$	−	+	−	+	−	+

operators $C_1^{u(6/2j+1)}$ and $C_2^{u(6/2j+1)}$.

IV. SOME REMARKS ON THE MATRIX ELEMENTS

For the diagonalization of the general Hamiltonian one has to know the matrix elements of all invariants in a certain basis. To this end, it is advantageous to start with a basis where as many relevant operators as possible are diagonal, i.e., to take the basis of the limiting case one is dealing with. For the sake of simplicity, we will restrict ourselves to the $u_B(5)$ and $so_B(6)$ limits [cases (I) and (II) of Eq. (1.2)]. In principle, the bosonic parts of these limits could be taken from Ref. 3, but one should take care of the fact that there the states of the $u_B(5)$ limit are not orthonomalized. Therefore we give the relevant formulas again in Appendix B.

In what follows, we only consider $so_B(5) \times sp_F(2j+1)$ scalars since—roughly speaking—the effect of invariants on lower stages are nearly negligible in the framework of a successive, dynamic symmetry breaking according to case (I) or (II) of Eq. (1.2). As a consequence, the matrix elements of an $so_B(5)\times sp_F(2j+1)$ scalar H_i have the general form

$$\langle \mathscr{N}, N_B, \nu, \lambda, \tau, (\alpha)|H_i|\mathscr{N}', N_B', \nu', \lambda', \tau', (\alpha')\rangle = \delta_{\mathscr{N}\mathscr{N}'}\delta_{\nu\nu'}\delta_{\tau\tau'}\delta_{\alpha\alpha'} \times \langle \mathscr{N}, N_B, \nu, \lambda, \tau, (\alpha)|H_i|\mathscr{N}, N_B', \nu, \lambda', \tau, (\alpha)\rangle, \tag{4.1}$$

where (α) denotes the additional set of quantum numbers necessary for a unique state labeling and $\lambda = n_d$ [for case (I)] or $\lambda = \sigma$ [for case (II)]. Note that H_i is diagonal in $\mathscr{N}$ due to the imposed conservation of the total particle number and that the matrix elements of an $so_B(5)\times sp_F(2j+1)$ scalar do not depend on the quantum numbers (α) in Eq. (4.1).

In Appendix B, we give the transformation between the bases—which are orthonormalized in contrast to those of Ref. 3—of the two limiting cases considered. Consequently, if H_i is diagonal in one basis, its matrix elements in the other one can be calculated straightforwardly. Since $\hat{A}$ and $\hat{B}$ are the only $so_B(5)\times sp_F(2j+1)$ scalar that are not diagonal in both bases, we give their matrix elements explicitly:

$$\langle \mathscr{N}, N_B', \nu, n_d, \tau, (\alpha)|\hat{A}|\mathscr{N}, N_B, \nu, n_d, \tau, (\alpha)\rangle = \delta_{2,|N_B-N_B'|}[(x-n_d+1)(x-n_d)(\mathscr{N}-x-\nu+1)(2j+2+x-\mathscr{N}-\nu)/(2j+1)]^{1/2}, \tag{4.2}$$

and

$$\langle \mathscr{N}, N_B', \nu, \sigma, \tau, (\alpha)|\hat{B}|\mathscr{N}, N_B, \nu, \sigma, \tau, (\alpha)\rangle = \delta_{2,|N_B-N_B'|}[(x-\sigma+1)(x+\sigma+5)(\mathscr{N}-x-\nu+1)(2j+2+x-\mathscr{N}-\nu)/(2j+1)]^{1/2}, \tag{4.3}$$

where $x = \frac{1}{2}(N_B + N_B')$. These formulas can be come to by means of the states given in Appendix B.

V. DISCUSSION

Let us first sum up our results. For the symmetry breaking scheme (1.1), which has been proposed recently,[1] we have constructed all Hermitian scalars that conserve the total particle number and consist of expressions at most quartic in the bosonic and fermionic creation and annihilation operators. This automatically leads to the most general Hamiltonian for this supersymmetry scenario.

For the $so_B(5)\times sp_F(2j+1)$ scalars all formulas are given that are needed for the calculation of the matrix ele-

TABLE IV. Invariance property of the Bose–Fermi operators of Eq. (3.5). A plus (minus) sign means that the corresponding operator on the left is (is not) an invariant with respect to the corresponding algebra on top.

Operator \ Algebra	$u_F(2j+1)$	$sp_F(2j+1)$	$u_B(6)$	$so_B(6)$	$su_B(3)$	$u_B(5)$	$so_B(5)$
$\hat{N}_B \cdot \hat{N}_F$	+	+	+	+	+	+	+
$\hat{n}_d \cdot \hat{N}_F$	+	+	−	−	−	+	+
$\hat{A}$	−	+	−	−	−	+	+
$\hat{B}$	−	+	−	+	−	−	+

ments in the $u_B(5)$ and $so_B(6)$ limits [cases (I) and (II) of Eq. (1.2)]. We did not carry out the diagonalization of the general Hamiltonian, which is a numerical problem. Nevertheless, this should be done in the future if more data are available to test the relevance of the different terms. *A priori*, there is no argument why nondiagonal invariants should be less important than diagonal ones, but it is astonishing that in several cases[1,4] the diagonal terms alone produce a remarkably good description of the experimental data. A physical understanding of these phenomena remains an open question.

ACKNOWLEDGMENTS

The authors would like to thank V. Rittenberg and A. Gelberg for valuable discussions and for critically reading the manuscript.

APPENDIX A: THE $sp_F(2j+1)$ SCALARS

In what follows, we shall prove that $\hat{N}_F$, $C_2^{u_F(2j+1)}$, and $C_2^{sp_F(2j+1)}$ are the only fermionic $sp_F(2j+1)$ invariants that satisfy our general conditions.

If we suppose the existence of a further linearly independent invariant, I, we can, without loss of generality, choose

$$I=\sum_{l=1}^{j-3/2}\beta_l A_l\,,\quad A_l=((a^+\times a^+)^{(2l)}\times(\tilde a\times\tilde a)^{(2l)})^{(0)}\,. \tag{A1}$$

With the generators of $sp_F(2j+1)$ (see Table I), we must have

$$[I,(a^+\times\tilde a)^{(k)}_\mu]=0\,,\quad k\ \text{odd}\,, \tag{A2}$$

and hence

$$\begin{aligned}0&=\prod_{\nu=-j}^{j-2}a^+_\nu\ [I,(a^+\times\tilde a)^{(2j)}_{2j}]\\&=\prod_{\nu=-j}^{j}a^+_\nu\sum_{s=3/2}^{j}\tilde a_s\tilde a_{1-s}\sum_{l=1}^{j-3/2}\alpha_l C^{jj\ \ 2l}_{s1-s1}\,,\end{aligned} \tag{A3}$$

where

$$\alpha_l=(4\beta_l/\sqrt{4l+1})\,C^{jj\ \ 2l}_{j1-j1}\,,\quad l=1,2,\ldots,j-\tfrac{3}{2}\,. \tag{A4}$$

Consequently,

$$\sum_{l=1}^{j-3/2}\alpha_l\,C^{jj\ \ 2l}_{s1-s1}=0\,,\quad s=\tfrac{3}{2},\tfrac{5}{2},\ldots,j\,. \tag{A5}$$

Now, using the orthogonality relations of the Clebsch–Gordan coefficients, we obtain

$$\begin{aligned}0&=\sum_{s=3/2}^{j}\sum_{l=1}^{j-3/2}\alpha_l C^{jj\ \ 2k}_{s1-s1}C^{jj\ \ 2l}_{s1-s1}\\&=\frac{1}{2}\sum_{l=1}^{j-3/2}\alpha_l\sum_{s=-j}^{j}C^{jj\ \ 2k}_{s1-s1}C^{jj\ \ 2l}_{s1-s1}\\&=\tfrac{1}{2}\alpha_k\,,\quad k=1,2,\ldots,j-\tfrac{3}{2}\,.\end{aligned} \tag{A6}$$

From (A4) one can now conclude that $\beta_l=0$, for $l=1,2,\ldots,j-\frac{3}{2}$, i.e., $I\equiv 0$, which completes the proof.

APPENDIX B: SOME REMARKS ON THE FOCK BASIS

In this appendix, we give the relevant formulas that are needed for the calculation of the matrix elements of all $so_B(5)\times sp_F(2j+1)$ scalars in the $u_B(5)$ and $so_B(6)$ limits of Eq. (1.2). Our aim is first the construction of the purely fermionic states, $|N_F,\nu,(\beta)\rangle$, and the purely bosonic ones, $|N_B,\lambda,\tau,(\gamma)\rangle$, $\lambda=n_d,\sigma$, in the Fock space. To this end, we assume the orthonormal states $|\nu,\nu,(\beta)\rangle$ and $|\tau,\tau,\tau,(\gamma)\rangle$ to be given.

Acting on these states with the Casimir operators of $sp_F(2j+1)$ and $so_B(5)$ (see Table II), respectively, one finds

$$(\tilde a\times\tilde a)^{(0)}\,|\nu,\nu,(\beta)\rangle=0\,, \tag{B1}$$

and, with $B:=\sqrt{5}(\tilde d\times\tilde d)^{(0)}$,

$$B\,|\tau,\tau,\tau,(\gamma)\rangle=0\,. \tag{B2}$$

Using now the commutation relations

$$\begin{aligned}&[(\tilde a\times\tilde a)^{(0)},(a^+\times a^+)^{(0)}]\\&\quad=[2/(2j+1)](2\hat N_F-2j-1)\,,\\&[B,B^+]=10+4\hat n_d\,,\end{aligned} \tag{B3}$$

one obtains after normalization

$$\begin{aligned}&|N_F,\nu,(\beta)\rangle\\&\quad=\frac{1}{2^k}\left[\frac{(2j+1)^k(j+\frac{1}{2}-\nu-k)!}{k!(j+\frac{1}{2}-\nu)!}\right]^{1/2}\\&\qquad\times[(a^+\times a^+)^{(0)}]^k\,|\nu,\nu,(\beta)\rangle\,,\end{aligned} \tag{B4}$$

where $k=\frac{1}{2}(N_F-\nu)$, and

$$\begin{aligned}&|n_d,n_d,\tau,(\gamma)\rangle\\&\quad=\left[\frac{(2\tau+3)!!}{2^k k!(n_d+\tau+3)!!}\right]^{1/2}(B^+)^k\,|\tau,\tau,\tau,(\gamma)\rangle\,,\end{aligned} \tag{B5}$$

where $k=\frac{1}{2}(n_d-\tau)$. Together with

$$\begin{aligned}&|N_B,n_d,\tau,(\gamma)\rangle\\&\quad=\left[\frac{1}{(N_B-n_d)!}\right]^{1/2}(s^+)^{N_B-n_d}\,|n_d,n_d,\tau,(\gamma)\rangle\,,\end{aligned} \tag{B6}$$

we have the states in the $u_B(5)$ basis.

Now let us construct the bosonic states in the $so_B(6)$ basis. Starting again from the states $|\tau,\tau,\tau,(\gamma)\rangle$ we first give the states $|\sigma,\sigma,\tau,(\gamma)\rangle$. Acting on such a state with the Casimir operators of $so_B(5)$ and $so_B(6)$, one obtains the conditions

$$\begin{aligned}&|\sigma,\sigma,\tau,(\gamma)\rangle\\&\quad=\sum_{i=0}^{[(\sigma-\tau)/2]}\alpha_i(\sigma,\tau)(B^+)^i(s^+)^{\sigma-\tau-2i}\,|\tau,\tau,\tau,(\gamma)\rangle\end{aligned} \tag{B7}$$

and

$$(B-s^2)\,|\sigma,\sigma,\tau,(\gamma)\rangle=0\,, \tag{B8}$$

respectively. The normalization then yields

$$\begin{aligned}&\alpha_i(\sigma,\tau)\\&\quad=\left[\frac{(\sigma+\tau+3)!(2\tau+3)!!}{2^{\sigma+1}(\sigma+1)!(\sigma-\tau)!}\right]^{1/2}\\&\qquad\times\frac{(2i+1)!!}{(2i+1)(2i+2\tau+3)!!}\binom{\sigma-\tau}{2i}\,.\end{aligned} \tag{B9}$$

Finally, one gets

$$|N_B, \sigma, \tau, (\gamma)\rangle = \left[\frac{(\sigma+2)!2^{\sigma-N_B}}{((N_B-\sigma)/2)!((N_B+\sigma)/2+2)!}\right]^{1/2} \times [B^+ - (s^+)^2]^{(N_B-\sigma)/2}|\sigma, \sigma, \tau, (\gamma)\rangle . \quad \text{(B10)}$$

The Bose–Fermi states can now easily be obtained by the well-known angular momentum coupling.

By some algebraic manipulations one gets the following formula for the basis transformation:

$$\langle N'_B, n_d, \tau', (\alpha') \, |N_B, \sigma, \tau, (\alpha)\rangle = \delta_{N_B,N'_B}\delta_{\tau,\tau'}\delta_{\alpha,\alpha'}\xi^{N_B,\sigma}_{n_d,\tau} , \quad \text{(B11)}$$

where

$$\xi^{N_B,\sigma}_{n_d,\tau} = (-1)^{(N_B-\sigma)/2}\left[\frac{(\sigma+2)(N_B-n_d)!(\sigma+\tau+3)!((n_d-\tau)/2)!(n_d+\tau+3)!!}{2^{N_B+1-(1/2)(n_d-\tau)}(\sigma-\tau)!((N_B-\sigma)/2)!((N_B+\sigma)/2+2)!}\right]^{1/2} \times \sum_{j=0}^{(N_B-\sigma)/2}(-1)^j\binom{(N_B-\sigma)/2}{j}\binom{\sigma-\tau}{n_d-\tau-2j}\prod_{k=0}^{\tau+1}\frac{1}{n_d-\tau-2j+1+2k} . \quad \text{(B12)}$$

Notice that this formula does not coincide with that of Ref. 3 because there the authors use a nonorthonormal basis for the states of the $u_B(5)$ limit.

[1]M. Baake, P. Reinicke, and A. Gelberg, Phys. Lett. B **166**, 10 (1986).

[2]A. Arima and F. Iachello, Adv. Nucl. Phys. **13**, 139 (1984), and references therein.

[3]A. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).

[4]R. F. Casten and P. von Brentano, Phys. Lett. **152**, 22 (1985).

Scattering of electromagnetic waves from a perfectly conductive slightly random surface: Depolarization in backscatter

J. Nakayama, K. Mitzutani, and M. Tsuneoka
Department of Electronics, Kyoto Institute of Technology, Matsugasaki, Kyoto 606, Japan

(Received 10 September 1985; accepted for publication 11 December 1985)

The scattering of electromagnetic waves from a perfectly conductive slightly random surface is studied by a probabilistic method developed recently. For a plane wave incident on a homogeneous, isotropic, Gaussian random surface, a stochastic wave solution involving multiple scattering effects is approximately obtained by use of the Wiener–Hermite expansion technique in the probability theory. Then the backscattering cross section per unit surface is derived for both like and cross polarization in a closed form and is shown in the figures. The surface waves propagating along the random surface are discussed.

I. INTRODUCTION

The scattering of waves from randomly rough surfaces has been investigated by many authors, because it is closely related to practical problems such as propagation over rough sea and land, backscatter from a rough terrain, optical properties of a random metal surface, etc. For references, see the literature.[1–4] Classical approaches of analysis are the small perturbation method[5,6] for a slightly rough surface and the Kirchhoff approximation[1–3] for a large-scale rough surface. They give a closed form solution for the scattering cross section. As is well known, however, the small perturbation method yields a solution that unphysically diverges for a perfectly conductive random surface and the Kirchhoff approximation gives no polarization effect. To overcome these drawbacks and to clarify the physical mechanism of the scattering precisely, several scattering theories have been developed recently. They are, for example, the diagram technique,[7–9] use of the extinction theorem,[10,11] the full wave analysis,[12,13] a probabilistic method,[14–20] and others[21–24] by which significant progress has been made. However, little effort seems to have been made concerning the depolarization in backscatter, which is often observed in experiments.

Valenzuela[6] derived a formula for the depolarized backscattering cross section by Rice's small perturbation approach.[5] His solution works well for a dielectric-air interface but gives a diverging scattering cross section for a perfectly conductive surface. Bahar[12,13] and Nieto-Vesperinas[11] investigate the depolarization effect by the full wave analysis and by use of the extinction theorem, respectively. But they were unsuccessful in obtaining depolarization in backscatter. Several multiple scattering theories introduced recently have not been successfully developed for depolarization in backscatter yet.[21–24]

The purpose of the present paper is to derive a mathematical formula representing the depolarized backscattering cross section for a perfectly conductive Gaussian random surface. For analysis we employ a probabilistic method[14–20] developed recently, which was originally introduced in the theory of waves in random media.[25–27] By use of the translation invariance property of a homogeneous random surface, we look for a possible form of the scattered wave, which proves to be a homogeneous random function multiplied by an exponential phase factor. Using the Wiener–Hermite expansion technique[28–30] in the probability theory, which is summarized in the Appendix, we calculate the scattered wave field as a stochastic functional of the Gaussian random surface. Then we definitely obtain the backscattering cross section with depolarization.

Only the monochromatic wave is considered but the time dependence $\exp(-i2\pi f_0 t)$ assumed is suppressed throughout the paper.

II. FORMULATION OF THE PROBLEM

Let us consider the scattering of an electromagnetic wave from a slightly random surface as is shown in Fig. 1. We denote a three-dimensional vector by $z\mathbf{e}_z + \mathbf{r}$, $\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y$ being a two-dimensional vector in the x-y plane $R^2 = (-\infty < x, y < \infty)$, and $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ being unit vectors along the Cartesian coordinates. We assume that the perfectly conductive surface is mathematically given by a homogeneous random function as

$$z = f(\mathbf{r}, \omega), \quad \mathbf{r} \in R^2, \quad \omega \in \Omega, \tag{1}$$

where ω is a probability parameter denoting a sample point in the sample space Ω. We assume a strong condition such

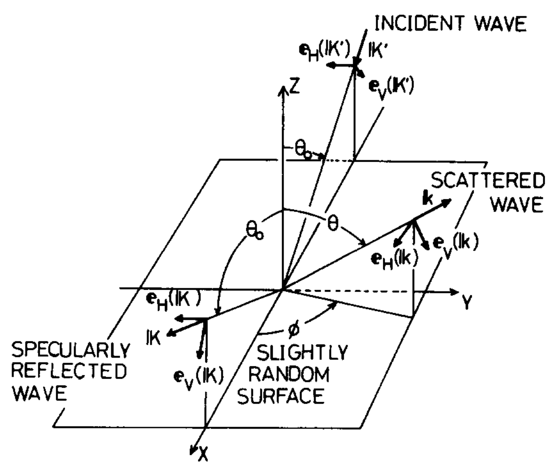

FIG. 1. Scattering of electromagnetic plane waves from a slightly random perfectly conductive surface. The wave vectors of an incident wave, of a specularly reflected wave, and of a scattered wave are denoted by $\mathbf{K}'$, $\mathbf{K}$, and $\mathbf{k}$, respectively. The vertical and horizontal polarization vectors are represented by $\mathbf{e}_H$ and $\mathbf{e}_V$, respectively.

 0022-2488/86/051435-14$02.50

that the sample space Ω is of the functional space type[31] derived from $f(\mathbf{r},\omega)$. Under this condition, a shift of the sample function by $\mathbf{a}$ generates a measure-preserving transformation $T^{\mathbf{a}}$ in the sample space such that[25,28,31]

$$f(\mathbf{r}+\mathbf{a},\omega)=f(\mathbf{r},T^{\mathbf{a}}\omega)\ ,\quad \mathbf{a}\in R^2\ . \tag{2}$$

Here, the measure-preserving transformation $T^{\mathbf{a}}$, which is often called the shift, becomes a group: $T^0=1$ (identity); $T^{\mathbf{a}}T^{\mathbf{b}}=T^{\mathbf{a}+\mathbf{b}}$. Once the shift $T^{\mathbf{a}}$ is so defined, $g(T^{\mathbf{r}}\omega)$ becomes a strictly homogeneous random function of $\mathbf{r}$ for any random variable $g(\omega)$. We call such a random function $g(T^{\mathbf{r}}\omega)$ a homogeneous random function generated by the shift. For example, if we put $f(\omega)\equiv f(0,\omega)$, we can write

$$z=f(\mathbf{r},\omega)=f(T^{\mathbf{r}}\omega)\ , \tag{3}$$

where $f(T^{\mathbf{r}}\omega)$ is a homogeneous random function generated by the shift. Note that $f(\mathbf{r},\omega)=f(T^{\mathbf{r}}\omega)$ does not have any translation invariance as a function of $\mathbf{r}$, i.e., as a sample function. From (2) or (3), however, the homogeneous random function $f(\mathbf{r},\omega)$ is invariable under a translation in the product space $R^2\times\Omega$ that takes $(\mathbf{r},\omega)$ into $(\mathbf{r}+\mathbf{a},\ T^{-\mathbf{a}}\omega)$. Such a translation invariance will be used below.

In what follows, however, we assume for concrete analysis that $f(T^{\mathbf{r}}\omega)$ is a homogeneous and isotropic Gaussian random function given by a Wiener integral[28,30]:

$$z=f(T^{\mathbf{r}}\omega)=\int_{R^2}F(\boldsymbol{\lambda})e^{i\boldsymbol{\lambda}\mathbf{r}}\,dB(\boldsymbol{\lambda},\omega)\ , \tag{4}$$

where $dB(\boldsymbol{\lambda},\omega)$ is the complex Gaussian random measure defined in (A4)–(A6). By (A5), Eq. (4) has zero average and the variance equal to

$$\sigma^2=\langle f^2(T^{\mathbf{r}}\omega)\rangle=\int_{R^2}|F(\boldsymbol{\lambda})|^2\,d\boldsymbol{\lambda}\ . \tag{5}$$

Here, the angle brackets denote the average over the sample space, σ is the rms surface height, and $|F(\boldsymbol{\lambda})|^2$ is the roughness spectrum satisfying

$$|F(\boldsymbol{\lambda})|^2=|F(\Lambda)|^2\ ,\quad |\boldsymbol{\lambda}|=\Lambda\ , \tag{6}$$

which makes $f(T^{\mathbf{r}}\omega)$ isotropic. For an arbitrary roughness spectrum we can calculate many statistical properties of the scattering. For numerical calculation, however, we will use the Gaussian roughness spectrum:

$$|F(\boldsymbol{\lambda})|^2=(\sigma^2\kappa^2/\pi)\exp(-\kappa^2\lambda^2)\ , \tag{7}$$

where κ is the correlation radius of the random surface.

We denote the electric field in free space above the surface by $\mathbf{E}(z,\mathbf{r},\omega)$, which satisfies Maxwell's equations, the divergence-free condition ($\operatorname{div}\mathbf{E}=0$), and the boundary condition $\mathbf{n}\times\mathbf{E}(z,\mathbf{r},\omega)=0$ on the random surface. For a slightly rough surface with gentle sloping, where $k\sigma\ll1$ and $k\kappa\gg1$, however, we employ the effective boundary condition[3,16,17] on the x-y plane at $z=0$:

$$\mathbf{E}_0(0,\mathbf{r},\omega)+f(T^{\mathbf{r}}\omega)\frac{\partial}{\partial z}\mathbf{E}_0(0,\mathbf{r},\omega)+\operatorname{grad}f(T^{\mathbf{r}}\omega)\cdot E_z(0,\mathbf{r},\omega)=0\ , \tag{8}$$

where $\mathbf{E}_0$ is the projection of the electric field $\mathbf{E}(z,\mathbf{r},\omega)$ on the x-y plane and E_z is its z component. We define the wave vector $\mathbf{k}$ and the polarization vectors associated with $\mathbf{k}$ as

$$\mathbf{k}=\mathbf{k}_0+k_z\mathbf{e}_z\ ,\quad k_0^2=\mathbf{k}_0^2\ ,\quad k_z=(k^2-\mathbf{k}_0^2)^{1/2}\ ,$$
$$\mathbf{e}_H(\mathbf{k})=\mathbf{k}_0\times\mathbf{e}_z/k_0\ ,\quad \mathbf{e}_V(\mathbf{k})=\mathbf{k}\times\mathbf{e}_H(\mathbf{k})/k\ , \tag{9}$$

where k is the wave number in free space, and a subscript H or V indicates the horizontal or vertical polarization, respectively. We write the electric field as

$$\mathbf{E}(z,\mathbf{r},\omega)=\mathbf{E}^0(z,\mathbf{r})+\mathbf{E}^s(z,\mathbf{r},\omega)\ , \tag{10}$$

where $\mathbf{E}^0$ is the unperturbed electric field over the nonfluctuating surface $z=0$ and $\mathbf{E}^s$ is the perturbed field due to the surface roughness. For a horizontally polarized plane wave incident we write

$$\mathbf{E}^0(z,\mathbf{r})=e^{iK_0x}[-\mathbf{e}_H(\mathbf{K}')e^{-iK_zz}+\mathbf{e}_H(\mathbf{K})e^{iK_zz}]\ , \tag{11}$$

$$\mathbf{K}'=\mathbf{K}_0-K_z\mathbf{e}_z\ ,\quad \mathbf{K}=\mathbf{K}_0+K_z\mathbf{e}_z\ ,$$
$$\mathbf{K}_0=K_0\mathbf{e}_x=k\sin(\theta_0)\mathbf{e}_x\ ,$$
$$K_z=(k^2-K_0^2)^{1/2}=k\cos(\theta_0)>0\ , \tag{12}$$

where the first term and the second one in the braces in (11) correspond to the incident plane wave and the specularly reflected wave respectively, the plane of incidence has been taken in the x-z plane, $\mathbf{K}'$ or $\mathbf{K}$ is the wave vector of incident wave or that of the specularly reflected wave, and θ_0 is the angle of incidence.

We have seen that the random surface $f(\mathbf{r},\omega)=f(T^{\mathbf{r}}\omega)$ is invariable under a translation in the product space $R^2\times\Omega$ that takes $(\mathbf{r},\omega)$ into $(\mathbf{r}+\mathbf{a},\ T^{-\mathbf{a}}\omega)$. Such a translation invariance is now employed in order to determine a possible form of the scattered wave. Following Ogura and coworkers,[25–27] we introduce the translation operator $D^{\mathbf{a}}$ by the relation

$$D^{\mathbf{a}}\mathbf{E}(z,\mathbf{r},\omega)=\mathbf{E}(z,\mathbf{r}+\mathbf{a},T^{-\mathbf{a}}\omega)\ ,\quad \mathbf{a}\in R^2\ . \tag{13}$$

By (10)–(13), the random surface, the incident plane wave, and the scattered wave are translated under $D^{\mathbf{a}}$ as follows:

$$z=f(T^{\mathbf{r}}\omega)\to D^{\mathbf{a}}f(T^{\mathbf{r}}\omega)=f(T^{\mathbf{r}}\omega)\ ,$$
$$\mathbf{E}^0(z,\mathbf{r})\to D^{\mathbf{a}}\mathbf{E}^0(z,\mathbf{r})=e^{i\mathbf{K}_0\mathbf{a}}\mathbf{E}^0(z,\mathbf{r})\ , \tag{14}$$
$$\mathbf{E}^s(z,\mathbf{r},\omega)\to D^{\mathbf{a}}\mathbf{E}^s(z,\mathbf{r},\omega)=\mathbf{E}^s(z,\mathbf{r}+\mathbf{a},T^{-\mathbf{a}}\omega)\ .$$

Because operating $D^{\mathbf{a}}$ gives only an additional phase factor to the incident plane wave for any value of $\mathbf{a}$ and the random surface remains invariable, $D^{\mathbf{a}}\mathbf{E}^s(z,\mathbf{r},\omega)$ should equal $\mathbf{E}^s(z,\mathbf{r},\omega)$ multiplied by the additional phase factor,[32] namely,

$$D^{\mathbf{a}}\mathbf{E}^s(z,\mathbf{r},\omega)=e^{i\mathbf{K}_0\mathbf{a}}\mathbf{E}^s(z,\mathbf{r},\omega)\ , \tag{15}$$

from which one easily finds a possible form of the electric field as

$$\mathbf{E}(z,\mathbf{r},\omega)=e^{i\mathbf{K}_0\mathbf{r}}[\mathbf{U}^0(z)+\mathbf{U}^s(z,T^{\mathbf{r}}\omega)]\ , \tag{16}$$

where we have put

$$\mathbf{U}^0(z)=\mathbf{E}^0(z,0)\ ,\quad \mathbf{U}^s(z,\omega)=\mathbf{E}^s(z,0,\omega)\ . \tag{17}$$

Note that $\mathbf{U}^s(z,T^{\mathbf{r}}\omega)$ in the right-hand side of (16) is a strictly homogeneous random function of $\mathbf{r}$, because it is described by the shift $T^{\mathbf{r}}$ in the sample space[31]: but $\mathbf{U}^s(z,T^{\mathbf{r}}\omega)$ is inhomogeneous in the z direction. Thus the scattered wave is written by a homogeneous random function multiplied by a phase factor; the phase factor is uniquely determined by the incident plane wave. This is analogous to the Floquet theorem in the theory of diffraction by a periodic surface, where the form of solution becomes a periodic function, with the same period as the surface, multiplied by a phase factor.

Because of (16), the problem is reduced to finding a homogeneous vector random function $\mathbf{U}^s(z,T^{\mathbf{r}}\omega)$. Generally speaking, there are two different ways to calculate such an unknown random function. In many scattering theories,[7–9,21] such an unknown random function is usually considered as a function of z and $\mathbf{r}$ for a fixed ω to derive an integral equation for the wave field. Then the integral equation is averaged over the sample space in order to get a moment equation such as the Bethe–Salpeter (BS) equation. However, this paper employs another idea, commonly used in the theory of stochastic processes, that considers $\mathbf{U}^s(z,T^{\mathbf{r}}\omega)$ as a random variable (a function of ω) for fixed z and $\mathbf{r}$. Because the sample space Ω is assumed to be a functional space type derived by $f(\mathbf{r},\omega)$, a function of ω means a functional of $f(\mathbf{r},\omega)$ in general, so that $\mathbf{U}(z,T^{\mathbf{r}}\omega)$ is a functional of the random surface $f(T^{\mathbf{r}}\omega)$. However, we will regard $\mathbf{U}^s(z,T^{\mathbf{r}}\omega)$ as a stochastic functional of $dB(\boldsymbol{\lambda},\omega)$, because $f(T^{\mathbf{r}}\omega)$ is a functional of $dB(\boldsymbol{\lambda},\omega)$ by (4). Then each component of $\mathbf{U}^s(z,T^{\mathbf{r}}\omega)$ can be represented in terms of the Wiener–Hermite expansion given by (A14), which holds as a random variable in the mean square sense [cf. (A15)]. However, we modify (A14) to make the scattered wave $\mathbf{E}^s(z,\mathbf{r},\omega) = e^{iK_0x}\mathbf{U}^s(z,T^{\mathbf{r}}\omega)$ satisfy Maxwell's equations and the divergence-free condition ($\operatorname{div}\mathbf{E}^s = 0$) term by term, and we put

$$\begin{aligned}\mathbf{E}^s(z,\mathbf{r},\omega) &= e^{iK_0x}\{\mathbf{e}_H(\mathbf{K})\,A_0e^{iK_zz} + \mathbf{e}_V(\mathbf{K})B_0e^{iK_zz}\}\\ &\quad + e^{iK_0x}\int_{R^2}\{A_1(\boldsymbol{\lambda})\mathbf{e}_H[\boldsymbol{\lambda}] + B_1(\boldsymbol{\lambda})\mathbf{e}_V[\boldsymbol{\lambda}]\}\\ &\quad \times e^{i\boldsymbol{\lambda}\mathbf{r} + ik_z(\boldsymbol{\lambda})z}h^{(1)}[dB(\boldsymbol{\lambda})]\\ &\quad + e^{iK_0x}\int_{R^2}\int_{R^2}\{A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\\ &\quad + B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\mathbf{e}_V[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\}\\ &\quad \times e^{i(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\mathbf{r} + ik_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')z}h^{(2)}[dB(\boldsymbol{\lambda}),dB(\boldsymbol{\lambda}')]\\ &\quad + \cdots,\end{aligned} \tag{18}$$

where the $h^{(n)}$'s are Wiener–Hermite differentials defined by (A7)–(A11), and $k_z(\boldsymbol{\lambda})$ is the z component of the wave vector $\mathbf{k}(\boldsymbol{\lambda})$ associated with a Bragg vector $\boldsymbol{\lambda}$ relating to a scattering direction,

$$\begin{aligned}&\mathbf{k}(\boldsymbol{\lambda}) = \mathbf{k}_0(\boldsymbol{\lambda}) + k_z(\boldsymbol{\lambda})\mathbf{e}_z\,,\quad \mathbf{k}^2(\boldsymbol{\lambda}) = k^2\,,\\ &\mathbf{k}_0(\boldsymbol{\lambda}) = \mathbf{K}_0 + \boldsymbol{\lambda} = (K_0+\lambda_x,\lambda_y)\,,\quad |\mathbf{k}_0(\boldsymbol{\lambda})| = k_0(\boldsymbol{\lambda})\,,\\ &k_z(\boldsymbol{\lambda}) = [k^2 - (\mathbf{K}_0+\boldsymbol{\lambda})^2]^{1/2},\quad \operatorname{Im}[k_z(\boldsymbol{\lambda})] \geqslant 0\,,\\ &K_0 = k_0(0)\,,\quad K_z = k_z(0)\,.\end{aligned} \tag{19}$$

A positive real and positive imaginary $k_z(\boldsymbol{\lambda})$ indicate an outgoing wave and an evanescent wave, respectively. The polarization vectors $\mathbf{e}_H[\boldsymbol{\lambda}]$ and $\mathbf{e}_V[\boldsymbol{\lambda}]$ associated with a Bragg vector $\boldsymbol{\lambda}$ are given by

$$\begin{aligned}\mathbf{e}_H[\boldsymbol{\lambda}] &= \mathbf{e}_H(\mathbf{k}(\boldsymbol{\lambda})) = \frac{\lambda_y}{k_0(\boldsymbol{\lambda})}\mathbf{e}_x - \frac{(K_0+\lambda_x)}{k_0(\boldsymbol{\lambda})}\mathbf{e}_y\,,\\ \mathbf{e}_V[\boldsymbol{\lambda}] &= \mathbf{e}_V(\mathbf{k}(\boldsymbol{\lambda}))\\ &= \frac{k_z(\boldsymbol{\lambda})[(K_0+\lambda_x)\mathbf{e}_x + \lambda_y\mathbf{e}_y]}{kk_0(\boldsymbol{\lambda})} - \frac{k_0(\boldsymbol{\lambda})}{k}\mathbf{e}_z\,.\end{aligned} \tag{20}$$

The unknown coefficients A_n and B_n, which are deterministic functions symmetrical with respect to their vector arguments, denote the amplitude of horizontally and vertically polarized partial waves. The first term in (18) is the coherent part of $\mathbf{E}^s$ and each integral represents an incoherent wave, which is written by a sum of outgoing plane waves and evanescent waves.

Once A_n and B_n are solved, we find the stochastic wave field from (18), in terms of which any statistical quantities of the scattering may be calculated. However, we write formulas only for the coherent scattering and the backscattering cross section per unit surface here. Since the $h^{(n)}$'s have zero average when $n\neq 0$ by (A9), we have from (10), (11), and (18) the coherent wave field

$$\begin{aligned}\langle\mathbf{E}(z,\mathbf{r},\omega)\rangle &= e^{iK_0x}\{-\mathbf{e}_H(\mathbf{K}')e^{-iK_zz}\\ &\quad + [(1+A_0)\mathbf{e}_H(\mathbf{K}) + B_0\mathbf{e}_V(\mathbf{K})]e^{iK_zz}\}\,.\end{aligned} \tag{21}$$

The backscattering cross section $\sigma^B(\theta)$ can be written in terms of the coefficients[16] as

$$\sigma^B(\theta) = \sigma^B_{hh}(\theta) + \sigma^B_{hv}(\theta)\,, \tag{22}$$

$$\begin{aligned}\sigma^B_{hh}(\theta) &= 4\pi k^2\cos^2(\theta)\Big\{|A_1(-2k\sin\theta\mathbf{e}_x)|^2\\ &\quad + 2!\int_{R^2}|A_2(-2k\sin\theta\mathbf{e}_x - \boldsymbol{\lambda}',\boldsymbol{\lambda}')|^2d\boldsymbol{\lambda}' + \cdots\Big\},\end{aligned} \tag{23}$$

$$\begin{aligned}\sigma^B_{hv}(\theta) &= 4\pi k^2\cos^2(\theta)\Big\{|B_1(-2k\sin\theta\mathbf{e}_x)|^2\\ &\quad + 2!\int_{R^2}|B_2(-2k\sin\theta\mathbf{e}_x - \boldsymbol{\lambda}',\boldsymbol{\lambda}')|^2d\boldsymbol{\lambda}' + \cdots\Big\},\end{aligned} \tag{24}$$

where θ is a scattering angle equal to the angle of incidence θ_0 (see Fig. 1), and subscripts hh and hv denote horizontal transmission–horizontal reception and horizontal transmission–vertical reception, respectively. We will calculate $\sigma^B_{hh}(\theta)$ and $\sigma^B_{hv}(\theta)$ below.

III. AN APPROXIMATE SOLUTION

In order to solve the coefficients A_n and B_n, we first derive equations for A_n and B_n from the boundary condition (8). Since the Wiener–Hermite expansion (18) holds in the mean square sense, the equality in (8) also should hold in the mean square sense, namely,

$$\Big\langle\Big|\mathbf{E}_0(0,\mathbf{r},\omega) + f(T^{\mathbf{r}}\omega)\frac{\partial}{\partial z}\mathbf{E}_0(0,\mathbf{r},\omega) + \operatorname{grad}f(T^{\mathbf{r}}\omega)\cdot E_z(0,\mathbf{r},\omega)\Big|^2\Big\rangle = 0\,. \tag{25}$$

Assuming that term by term derivative of the Wiener–Hermite expansion (18) converges to corresponding derivative of $\mathbf{E}^s$, we substitute (4), (10), (11), and (18) into (25) and calculate the left-hand side of (25) by use of the orthogonality relation (A9) and the recurrence formula (A8) of $h^{(n)}$. Then we find a set of equations for A_n and B_n, of which the lowest four equations are

$$\mathbf{e}_H(\mathbf{K})A_0+\mathbf{e}_{v0}(\mathbf{K})B_0+i\int_{R^2}k_z(\boldsymbol{\lambda})\{\mathbf{e}_H[\boldsymbol{\lambda}]A_1(\boldsymbol{\lambda})+\mathbf{e}_{v0}[\boldsymbol{\lambda}]B_1(\boldsymbol{\lambda})\}F^*(\boldsymbol{\lambda})\,d\boldsymbol{\lambda}-i\int_{R^2}\boldsymbol{\lambda}F^*(\boldsymbol{\lambda})e_{vz}[\boldsymbol{\lambda}]B_1(\boldsymbol{\lambda})\,d\boldsymbol{\lambda}=0, \tag{26}$$

$$\begin{aligned}&\mathbf{e}_H[\boldsymbol{\lambda}]A_1(\boldsymbol{\lambda})+\mathbf{e}_{v0}[\boldsymbol{\lambda}]B_1(\boldsymbol{\lambda})+iK_z[\mathbf{e}_H(\mathbf{K})(2+A_0)+\mathbf{e}_{v0}(\mathbf{K})B_0]F(\boldsymbol{\lambda})+i\boldsymbol{\lambda}F(\boldsymbol{\lambda})e_{vz}(\mathbf{K})B_0\\&\quad+2i\int_{R^2}F^*(\boldsymbol{\lambda}')k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\}\,d\boldsymbol{\lambda}'\\&\quad-2i\int_{R^2}\boldsymbol{\lambda}'F^*(\boldsymbol{\lambda}')e_{vz}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\,d\boldsymbol{\lambda}'=0,\end{aligned} \tag{27}$$

$$\begin{aligned}&\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\\&\quad+(i/2)F(\boldsymbol{\lambda})[k_z(\boldsymbol{\lambda}')\{\mathbf{e}_H[\boldsymbol{\lambda}']A_1(\boldsymbol{\lambda}')+\mathbf{e}_{v0}[\boldsymbol{\lambda}']B_1(\boldsymbol{\lambda}')\}+\boldsymbol{\lambda}e_{vz}[\boldsymbol{\lambda}']B_1(\boldsymbol{\lambda}')]\\&\quad+(i/2)F(\boldsymbol{\lambda}')[k_z(\boldsymbol{\lambda})\{\mathbf{e}_H[\boldsymbol{\lambda}]A_1(\boldsymbol{\lambda})+\mathbf{e}_{v0}[\boldsymbol{\lambda}]B_1(\boldsymbol{\lambda})\}+\boldsymbol{\lambda}'e_{vz}[\boldsymbol{\lambda}]B_1(\boldsymbol{\lambda})]\\&\quad+3i\int_{R^2}F^*(\boldsymbol{\lambda}'')k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']A_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']B_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')\}\,d\boldsymbol{\lambda}''\\&\quad-3i\int_{R^2}\boldsymbol{\lambda}''F^*(\boldsymbol{\lambda}'')e_{vz}[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']B_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')\,d\boldsymbol{\lambda}''=0,\end{aligned} \tag{28}$$

$$\begin{aligned}&\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']A_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']B_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')\\&\quad+(i/3)F(\boldsymbol{\lambda}'')k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\}\\&\quad+(i/3)F(\boldsymbol{\lambda}')k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'')\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}'']A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}'')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}'']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}'')\}\\&\quad+(i/3)F(\boldsymbol{\lambda})k_z(\boldsymbol{\lambda}''+\boldsymbol{\lambda}')\{\mathbf{e}_H[\boldsymbol{\lambda}''+\boldsymbol{\lambda}']A_2(\boldsymbol{\lambda}'',\boldsymbol{\lambda}')+\mathbf{e}_{v0}[\boldsymbol{\lambda}''+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda}'',\boldsymbol{\lambda}')\}+(i/3)\boldsymbol{\lambda}''F(\boldsymbol{\lambda}'')e_{vz}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\\&\quad+(i/3)\boldsymbol{\lambda}'F(\boldsymbol{\lambda}')e_{vz}[\boldsymbol{\lambda}+\boldsymbol{\lambda}'']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}'')+(i/3)\boldsymbol{\lambda}F(\boldsymbol{\lambda})e_{vz}[\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']B_2(\boldsymbol{\lambda}'',\boldsymbol{\lambda}')+\cdots=0,\end{aligned} \tag{29}$$

where $\mathbf{e}_{v0}[\boldsymbol{\lambda}]$ and $e_{Vz}[\boldsymbol{\lambda}]$ are the projection of $\mathbf{e}_V[\boldsymbol{\lambda}]$ on the x-y plane and its z component. Equations (26)–(28) are the same as (65)–(67) in a previous paper,[16] where the first-order solutions, including A_0, $A_1(\boldsymbol{\lambda})$, B_0, and $B_1(\boldsymbol{\lambda})$, were obtained with a rough approximation. However, we will obtain here the second-order solutions, involving A_0, B_0, $A_1(\boldsymbol{\lambda})$, $B_1(\boldsymbol{\lambda})$, $A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$, and $B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$, in a more precise manner.

Let us obtain A_2 and B_2 from (28). To obtain a solution involving multiple scattering, however, we approximately put

$$\begin{aligned}&\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']A_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']B_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')\\&\quad\simeq-(i/3)F(\boldsymbol{\lambda}'')k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')+\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\}-(i/3)\boldsymbol{\lambda}''F(\boldsymbol{\lambda}'')e_{vz}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}'),\end{aligned} \tag{30}$$

which has been obtained from (29) by neglecting the fourth, fifth, seventh, and eighth terms. Taking the $\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'']$ component of this gives

$$\begin{aligned}B_3(\boldsymbol{\lambda},\boldsymbol{\lambda}',\boldsymbol{\lambda}'')\simeq&\frac{-iF(\boldsymbol{\lambda}'')}{3k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')}[kk_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\boldsymbol{\lambda}''\}A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\\&+\{k_z^2(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')-k_0^2(\boldsymbol{\lambda}+\boldsymbol{\lambda}')[\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')\boldsymbol{\lambda}'']\}B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')].\end{aligned} \tag{31}$$

Substituting (30) and (31) into the fifth and sixth terms in (28), we obtain a vector equation for A_2 and B_2:

$$\begin{aligned}&A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\Big\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\Big[1+k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\int k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')|F(\boldsymbol{\lambda}'')|^2\,d\boldsymbol{\lambda}''\Big]\\&\quad+k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\int\frac{\boldsymbol{\lambda}''|F(\boldsymbol{\lambda}'')|^2(\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\boldsymbol{\lambda}'')}{k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')}\,d\boldsymbol{\lambda}''\Big\}\\&\quad+B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')\Big\{\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\Big[1+k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\int k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')|F(\boldsymbol{\lambda}'')|^2\,d\boldsymbol{\lambda}''\Big]\\&\quad+\int\frac{\boldsymbol{\lambda}''k[\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\boldsymbol{\lambda}'']}{k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}|F(\boldsymbol{\lambda}'')|^2\,d\boldsymbol{\lambda}''\Big\}\\&\quad=-(i/2)F(\boldsymbol{\lambda})[k_z(\boldsymbol{\lambda}')\{\mathbf{e}_H[\boldsymbol{\lambda}']A_1(\boldsymbol{\lambda}')+\mathbf{e}_{v0}[\boldsymbol{\lambda}']B_1(\boldsymbol{\lambda}')\}+\boldsymbol{\lambda}e_{vz}[\boldsymbol{\lambda}']B_1(\boldsymbol{\lambda}')]\\&\qquad-(i/2)F(\boldsymbol{\lambda}')[k_z(\boldsymbol{\lambda})\{\mathbf{e}_H[\boldsymbol{\lambda}]A_1(\boldsymbol{\lambda})+\mathbf{e}_{v0}[\boldsymbol{\lambda}]B_1(\boldsymbol{\lambda})\}+\boldsymbol{\lambda}'e_{vz}[\boldsymbol{\lambda}]B_1(\boldsymbol{\lambda})],\end{aligned} \tag{32}$$

where the limit of each integral is equal to R^2. Taking the $\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']$ and $\mathbf{e}_{v0}[\boldsymbol{\lambda}+\boldsymbol{\lambda}']$ components of this yields linear equations for A_2 and B_2,

$$[k+k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')]A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')+kQ(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$$
$$=-\frac{ikF(\boldsymbol{\lambda})[k_z(\boldsymbol{\lambda}')\mathbf{k}_0(\boldsymbol{\lambda}')\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')A_1(\boldsymbol{\lambda}')-kk_0(\boldsymbol{\lambda}')\{\boldsymbol{\lambda}\mathbf{e}_H[\boldsymbol{\lambda}']\}B_1(\boldsymbol{\lambda}')]}{2k_0(\boldsymbol{\lambda}')k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}$$
$$-\frac{ikF(\boldsymbol{\lambda}')[k_z(\boldsymbol{\lambda})\mathbf{k}_0(\boldsymbol{\lambda})\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')A_1(\boldsymbol{\lambda})-kk_0(\boldsymbol{\lambda})\{\boldsymbol{\lambda}'\mathbf{e}_H[\boldsymbol{\lambda}]\}B_1(\boldsymbol{\lambda})]}{2k_0(\boldsymbol{\lambda})k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}, \tag{33}$$

$$k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')Q(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')+[k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')+kZ_{sv}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')]B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$$
$$=-\frac{ikF(\boldsymbol{\lambda})}{2k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}\left[k_z(\boldsymbol{\lambda}')\boldsymbol{\lambda}\mathbf{e}_H[\boldsymbol{\lambda}']A_1(\boldsymbol{\lambda}')+\frac{[k^2\mathbf{k}_0(\boldsymbol{\lambda}')\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')-k_0^2(\boldsymbol{\lambda}')k_0^2(\boldsymbol{\lambda}+\boldsymbol{\lambda}')]B_1(\boldsymbol{\lambda}')}{kk_0(\boldsymbol{\lambda}')}\right]$$
$$-\frac{ikF(\boldsymbol{\lambda}')}{2k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}\left[k_z(\boldsymbol{\lambda})\boldsymbol{\lambda}'\mathbf{e}_H[\boldsymbol{\lambda}]A_1(\boldsymbol{\lambda})+\frac{[k^2\mathbf{k}_0(\boldsymbol{\lambda})\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')-k_0^2(\boldsymbol{\lambda})k_0^2(\boldsymbol{\lambda}+\boldsymbol{\lambda}')]B_1(\boldsymbol{\lambda})}{kk_0(\boldsymbol{\lambda})}\right], \tag{34}$$

where $Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$ and $Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$ are equivalent surface impedances for horizontal and vertical polarized partial waves, respectively, which are given by

$$Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')=k\int_{R^2}\left[k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')+\frac{\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\boldsymbol{\lambda}''\}^2}{k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')}\right]|F(\boldsymbol{\lambda}'')|^2\,d\boldsymbol{\lambda}'', \tag{35a}$$

$$=k\int_{R^2}\left[\mu(\boldsymbol{\lambda}+\boldsymbol{\lambda}',\boldsymbol{\Lambda})+\frac{\Lambda_H{}^2}{\mu(\boldsymbol{\lambda}+\boldsymbol{\lambda}',\boldsymbol{\Lambda})}\right]|F(\boldsymbol{\Lambda})|^2\,d\boldsymbol{\Lambda}, \tag{35b}$$

$$Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')=\frac{1}{k}\int_{R^2}\left\{k_z^2(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\cdot k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')+\frac{k^2[\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\boldsymbol{\lambda}'']^2}{k_0^2(\boldsymbol{\lambda}+\boldsymbol{\lambda}')k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')}\right\}|F(\boldsymbol{\lambda}'')|^2\,d\boldsymbol{\lambda}'', \tag{36a}$$

$$=\frac{1}{k}\int_{R^2}\left\{k_z^2(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\cdot\mu(\boldsymbol{\lambda}+\boldsymbol{\lambda}',\boldsymbol{\Lambda})+\frac{k^2\Lambda_V{}^2}{\mu(\boldsymbol{\lambda}+\boldsymbol{\lambda}',\boldsymbol{\Lambda})}\right\}|F(\boldsymbol{\Lambda})|^2\,d\boldsymbol{\Lambda}, \tag{36b}$$

and $Q(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$ is a coupling coefficient given by

$$Q(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')=k\int\frac{[\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\boldsymbol{\lambda}'']\{\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']\boldsymbol{\lambda}''\}}{k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\boldsymbol{\lambda}'')}|F(\boldsymbol{\lambda}'')|^2\,d\boldsymbol{\lambda}'', \tag{37a}$$

$$=k\int_{R^2}\frac{\Lambda_H\,\Lambda_V}{\mu(\boldsymbol{\lambda}+\boldsymbol{\lambda}',\boldsymbol{\Lambda})}|F(\boldsymbol{\Lambda})|^2\,d\boldsymbol{\Lambda}. \tag{37b}$$

Here, because $\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']$ and $\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')/k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')$ are unit vectors perpendicular to each other, we have

$$\boldsymbol{\lambda}''=\boldsymbol{\Lambda}=\Lambda_H\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']+\Lambda_V\frac{\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}{k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')},\quad d\boldsymbol{\lambda}''=d\boldsymbol{\Lambda}=d\Lambda_H\,d\Lambda_V, \tag{38}$$

$$\mu(\boldsymbol{\lambda}+\boldsymbol{\lambda}',\boldsymbol{\Lambda})=k_z\left(\boldsymbol{\lambda}+\boldsymbol{\lambda}'+\Lambda_H\mathbf{e}_H[\boldsymbol{\lambda}+\boldsymbol{\lambda}']+\Lambda_V\frac{\mathbf{k}_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}{k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}\right)=\{k^2-[k_0(\boldsymbol{\lambda}+\boldsymbol{\lambda}')+\Lambda_V]^2-\Lambda_H^2\}^{1/2}, \tag{39}$$

It should be noted, however, that (35b), (36b), and (37b) are valid only for an isotropic random surface with (6). For such an isotropic surface $Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$, $Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$, and $Q(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$ become isotropic functions with

$$Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')=Z_{sh}(|\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}'|),\quad Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')=Z_{sv}(|\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}'|), \tag{40}$$

$$Q(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')=Q(|\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}'|)=0, \tag{41}$$

where (41) holds because the integrand in (37b) is an odd function with respect to Λ_H. If the Gaussian roughness spectrum (7) is assumed, two-dimensional integrals in (35b) and (36b) are easily reduced to one-dimensional ones that are evaluated numerically. Numerical examples of $Z_{sh}(kp)$ and $Z_{sv}(kp)$ are illustrated in Fig. 2 as functions of kp for the Gaussian roughness spectrum (7). The real part (resistance) of $Z_{sh}(kp)$ or $Z_{sv}(kp)$ represents energy dissipation due to scattering, whereas the imaginary part (reactance) implies stored energy near the surface due to evanescent waves generated by the surface roughness. By (35) and (36) these surface impedances are proportional to σ^2, and hence they are considered as a double scattering effect. When the surface is slightly rough, they are small quantities of the order σ^2 for $0<kp<k$, but are not small for $kp\gg k$. In fact, Fig. 2 shows that, for $p\gg 1$, $\mathrm{Im}[Z_{sh}(kp)]$ becomes a large positive value proportional to p^1, and $\mathrm{Im}[Z_{sv}(kp)]$ also becomes a large negative value proportional to p^2. We will use these asymptotic property of the equivalent surface impedance below. From (33), (34), and (41) we find solutions for $A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$ and $B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$, which are symmetrical with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$:

$$A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')=\left[\frac{k}{k+k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')}\right][V_{2h}(\boldsymbol{\lambda},\boldsymbol{\lambda}')+V_{2h}(\boldsymbol{\lambda}',\boldsymbol{\lambda})], \tag{42a}$$

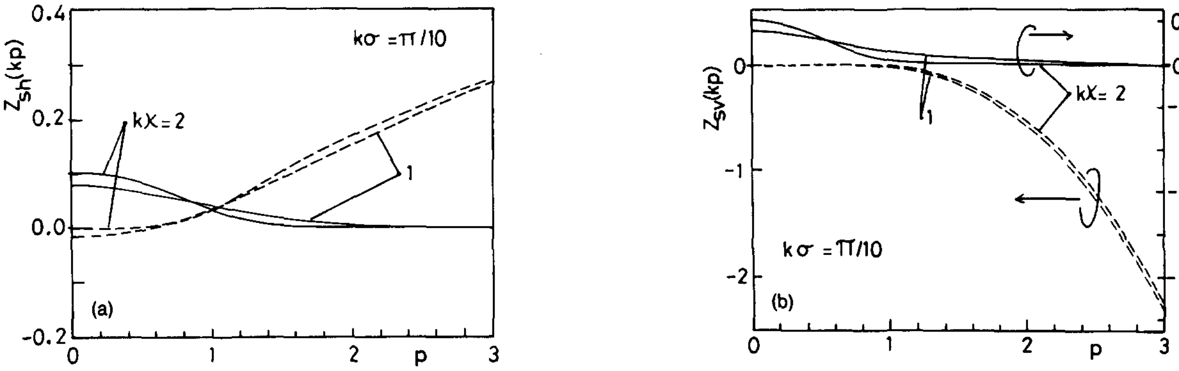

FIG. 2. Equivalent surface impedance. (a) $Z_{sh}(kp)$ for the horizontally polarized wave and (b) $Z_{sv}(kp)$ for the vertically polarized wave. The solid and broken curves indicate the real and imaginary parts, respectively. The Gaussian roughness spectrum (7) is assumed and κ is the correlation distance of the random surface.

$$V_{2h}(\lambda,\lambda') = -\frac{iF(\lambda)[k_z(\lambda')\mathbf{k}_0(\lambda')\mathbf{k}_0(\lambda+\lambda')A_1(\lambda') - kk_0(\lambda')\{\lambda\mathbf{e}_H[\lambda']\}B_1(\lambda')]}{2k_0(\lambda')k_0(\lambda+\lambda')}, \tag{42b}$$

$$B_2(\lambda,\lambda') = \left[\frac{k}{k_z(\lambda+\lambda') + kZ_{sv}(\mathbf{K}_0+\lambda+\lambda')}\right][V_{2v}(\lambda,\lambda') + V_{2v}(\lambda',\lambda)], \tag{43a}$$

$$V_{2v}(\lambda,\lambda') = -\frac{iF(\lambda)}{2k_0(\lambda+\lambda')}\left[k_z(\lambda')\{\lambda\mathbf{e}_H[\lambda']\}A_1(\lambda') + \frac{[k^2\mathbf{k}_0(\lambda')\mathbf{k}_0(\lambda+\lambda') - k_0^2(\lambda')\mathbf{k}_0^2(\lambda+\lambda')]B_1(\lambda')}{kk_0(\lambda)}\right]. \tag{43b}$$

The first factor in (42a) or in (43a) is a resonance factor that is an isotropic function depending only on $|\mathbf{K}_0+\lambda+\lambda'|$ by (19) and (40). A complex pole of the resonance factor determines a propagation constant of a guided surface wave propagating along the random surface, as will be discussed later. A numerical example of the resonance factor is illustrated in Fig. 3, in which a sharp peak suggests the existence of a guided surface wave. Solutions (42) and (43) may be represented by an equivalent network shown in Fig. 4(e) or Fig. 4(f); in Fig. 4(e), $V_{2h}(\lambda,\lambda') + V_{2h}(\lambda',\lambda)$ is the output voltage of the generator with internal impedance $Z_{sh}(\mathbf{K}_0+\lambda+\lambda')$ and $A_2(\lambda,\lambda')$ is the amplitude of the *voltage wave* traveling on the transmission line with impedance $k/k_z(\lambda+\lambda')$, whereas in Fig. 4(f), $V_{2v}(\lambda,\lambda') + V_{2v}(\lambda',\lambda)$ becomes the voltage of the generator with internal impedance $Z_{sv}(\mathbf{K}_0+\lambda+\lambda')$ and $B_2(\lambda,\lambda')$ is the amplitude of the *current wave* propagating on the transmission line with impedance $k_z(\lambda+\lambda')/k$.

Coefficient $A_2(\lambda,\lambda')$ remains finite without $Z_{sh}(\mathbf{K}_0+\lambda+\lambda')$. If we neglect $Z_{sv}(\mathbf{K}_0+\lambda+\lambda')$, however, $B_2(\lambda,\lambda')$ will diverge for $\mathbf{K}_0+\lambda+\lambda'$ such that

$$k_z(\lambda+\lambda') = [k^2 - (\mathbf{K}_0+\lambda+\lambda')^2]^{1/2} = 0, \quad |\mathbf{K}_0+\lambda+\lambda'| = k, \tag{44}$$

which is the so-called Rayleigh wave number. However, our solution (43) always remains finite and gives a finite value for the depolarized backscattering cross section, as will be shown below.

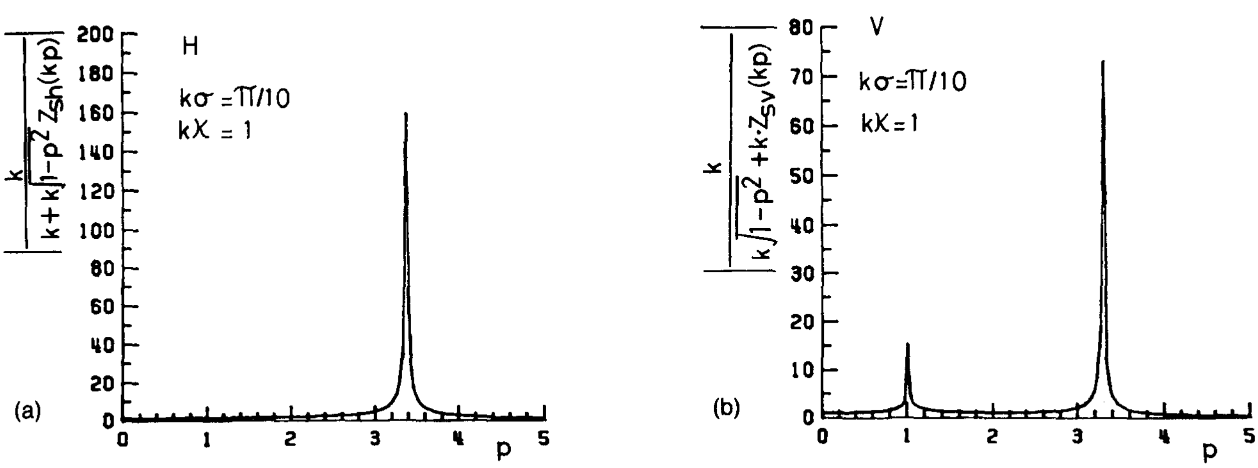

FIG. 3. Resonance factors in the solutions A_n and B_n, $n = 0,1,2$. (a) The resonance factor for the A_n's and (b) the resonance factor for the B_n's. A sharp peak suggests the existence of a guided surface wave propagating along the random surface.

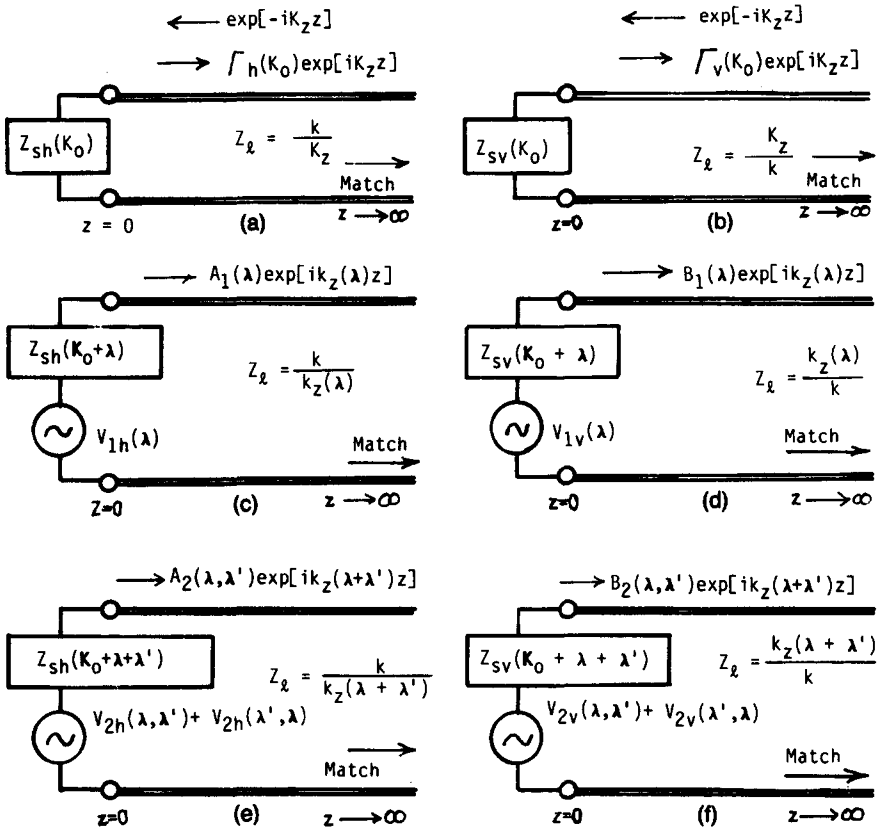

FIG. 4. Equivalent transmission line networks. The transmission line impedance is denoted by Z_l and the output voltage of the generator by V. The A_n's are the amplitude of the voltage wave, whereas the B_n's are the amplitude of the current wave. Γ_k is the voltage reflection coefficient, where Γ_v becomes the current reflection coefficient.

Next we look for A_1 and B_1. Neglecting the first term in (42a), the first term in (43a), $Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$, and $Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda}+\boldsymbol{\lambda}')$ as well, we insert (42) and (43) into (27). After calculation we find

$$A_1(\boldsymbol{\lambda})\left\{\mathbf{e}_H[\boldsymbol{\lambda}]\left[1+k_z(\boldsymbol{\lambda})\int k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')|F(\boldsymbol{\lambda}')|^2\,d\boldsymbol{\lambda}'\right]+k_z(\boldsymbol{\lambda})\int\frac{\boldsymbol{\lambda}'|F(\boldsymbol{\lambda}')|^2(\mathbf{e}_H[\boldsymbol{\lambda}]\boldsymbol{\lambda}')}{k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')}\,d\boldsymbol{\lambda}'\right\}$$
$$+B_1(\boldsymbol{\lambda})\left\{\mathbf{e}_{v0}[\boldsymbol{\lambda}]\left[1+k_z(\boldsymbol{\lambda})\int k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')|F(\boldsymbol{\lambda}')|^2\,d\boldsymbol{\lambda}'\right]+\int\frac{k\boldsymbol{\lambda}'|F(\boldsymbol{\lambda}')|^2[\mathbf{k}_0(\boldsymbol{\lambda})\boldsymbol{\lambda}']}{k_z(\boldsymbol{\lambda}+\boldsymbol{\lambda}')k_0(\boldsymbol{\lambda})}\,d\boldsymbol{\lambda}'\right\}$$
$$=iK_z(2+A_0)F(\boldsymbol{\lambda})\mathbf{e}_y+iB_0(K_0\boldsymbol{\lambda}-K_z^2\mathbf{e}_x)\,F(\boldsymbol{\lambda})/k\,, \tag{45}$$

where the limit of each integral is equal to R^2 again. Taking the $\mathbf{e}_H[\boldsymbol{\lambda}]$ and $\mathbf{e}_{v0}[\boldsymbol{\lambda}]$ components of this, we obtain linear equations for $A_1(\boldsymbol{\lambda})$ and $B_1(\boldsymbol{\lambda})$:

$$[k+k_z(\boldsymbol{\lambda})Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})]\,A_1(\boldsymbol{\lambda})+kQ(\mathbf{K}_0+\boldsymbol{\lambda})B_1(\boldsymbol{\lambda})=-\frac{ik\,[K_z(2+A_0)(K_0+\lambda_x)+B_0k\lambda_y]\,F(\boldsymbol{\lambda})}{k_0(\boldsymbol{\lambda})}, \tag{46}$$

$$k_z(\boldsymbol{\lambda})Q(\mathbf{K}_0+\boldsymbol{\lambda})A_1(\boldsymbol{\lambda})+[k_z(\boldsymbol{\lambda})+kZ_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})]B_1(\boldsymbol{\lambda})=iF(\boldsymbol{\lambda})\frac{K_z(2+A_0)\lambda_yk+B_0[K_0k_0^2(\boldsymbol{\lambda})-k^2(K_0+\lambda_x)]}{k_0(\boldsymbol{\lambda})}, \tag{47}$$

where $Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})$, $Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})$ and $Q(\mathbf{K}_0+\boldsymbol{\lambda})$ can be obtained from (35)–(37) by putting $\boldsymbol{\lambda}'=0$. Since $Q(\mathbf{K}_0+\boldsymbol{\lambda})=0$ by (41), we get

$$A_1(\boldsymbol{\lambda})=\left[\frac{k}{k+k_z(\boldsymbol{\lambda})Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})}\right]V_{1h}(\boldsymbol{\lambda})\,, \tag{48a}$$

$$V_{1h}(\boldsymbol{\lambda})=-\frac{i\,F(\boldsymbol{\lambda})\,[K_z(2+A_0)(K_0+\lambda_x)+\lambda_y\,kB_0]}{k_0(\boldsymbol{\lambda})}, \tag{48b}$$

$$B_1(\boldsymbol{\lambda})=\left[\frac{k}{k_z(\boldsymbol{\lambda})+kZ_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})}\right]V_{1v}(\boldsymbol{\lambda})\,, \tag{49a}$$

$$V_{1v}(\boldsymbol{\lambda})=\frac{iF(\boldsymbol{\lambda})\{K_z(2+A_0)\lambda_y\,k+B_0[K_0\,k_0^2(\boldsymbol{\lambda})-k^2(K_0+\lambda_x)]\}}{kk_0(\boldsymbol{\lambda})}, \tag{49b}$$

which also are represented by an equivalent transmission network shown in Figs. 4(c) and 4(d).

Next we obtain A_0 and B_0 to complete our second-order solution. Neglecting $Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})$ in (48a) and $Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})$ in (49a), we insert (48) and (49) into (26) to get an equation for A_0 and B_0:

$$\left[\mathbf{e}_y + K_z \int |F(\boldsymbol{\lambda})|^2 \left\{\mathbf{e}_y\, k_z(\boldsymbol{\lambda}) + \frac{\boldsymbol{\lambda}\lambda_y}{k_z(\boldsymbol{\lambda})}\right\} d\boldsymbol{\lambda}\right] A_0 - \left[\frac{K_z}{k}\mathbf{e}_x + \int \left\{\frac{K_z^2 k_z(\boldsymbol{\lambda})\mathbf{e}_x}{k} + \frac{\boldsymbol{\lambda}\lambda_x k}{k_z(\boldsymbol{\lambda})}\right\} |F(\boldsymbol{\lambda})|^2 d\boldsymbol{\lambda}\right] B_0$$

$$= -2K_z \int |F(\boldsymbol{\lambda})|^2 \left\{\mathbf{e}_y\, k_z(\boldsymbol{\lambda}) + \frac{\boldsymbol{\lambda}\lambda_y}{k_z(\boldsymbol{\lambda})}\right\} d\boldsymbol{\lambda}. \tag{50}$$

Dividing this equation into its x and y components, we have

$$[k + K_z Z_{sh}(\mathbf{K}_0)]\, A_0 - kQ(\mathbf{K}_0)B_0 = -2K_z Z_{sh}(\mathbf{K}_0), \tag{51a}$$

$$K_z Q(\mathbf{K}_0)\, A_0 - [K_z + kZ_{sv}(\mathbf{K}_0)]\, B_0 = -2K_z Q(\mathbf{K}_0). \tag{51b}$$

Here $Z_{sh}(\mathbf{K}_0)$ and $Z_{sv}(\mathbf{K}_0)$ are equivalent surface impedances given by (35) and (36), where $Q(\mathbf{K}_0)$ is the coupling coefficient given by (37). Since $Q(\mathbf{K}_0)=0$ for an isotropic random surface by (41), we find from (51) the solutions for A_0 and B_0,

$$A_0 = -2K_z Z_{sh}(\mathbf{K}_0)/[k + K_z Z_{sh}(\mathbf{K}_0)], \quad B_0 = 0, \tag{52}$$

where $B_0 = 0$ means no depolarization in the coherent scattering. We note, however, that depolarization in the coherent scattering may occur for an anisotropic random surface where $Q(\mathbf{K}_0)$ by (37a) may not vanish. Substituting (52) into (21), we have the coherent wave field

$$\langle \mathbf{E}(z,\mathbf{r},\omega)\rangle = e^{iK_0 x}[-\mathbf{e}_H(\mathbf{K}')e^{-iK_z z} - \Gamma_h(\mathbf{K}_0)\mathbf{e}_H(\mathbf{K})e^{iK_z z}], \tag{53}$$

where $\Gamma_h(\mathbf{K}_0)$ is the coherent reflection coefficient for the horizontal polarization given by

$$\Gamma_h(\mathbf{K}_0) = -(1+A_0) = [Z_{sh}(\mathbf{K}_0) - (k/K_z)]/[Z_{sh}(\mathbf{K}_0) + (k/K_z)]. \tag{54}$$

This equation means that the random surface works as a nonfluctuating flat surface with a surface impedance $Z_{sh}(\mathbf{K}_0)$ for the coherent scattering, while (k/K_z) is the wave impedance associated with the horizontal polarization. Therefore, (54) may be represented by an equivalent network shown in Fig. 4(a) also. We note that our solution A_n's and B_n's involve multiple scattering, because they have double scattering effects, Z_{sh} and Z_{sv}, in their denomonators.

Equation (52) reduces (48b) and (49b) to

$$V_{1h}(\boldsymbol{\lambda}) = -iF(\boldsymbol{\lambda})K_z(2+A_0)(K_0+\lambda_x)/k_0(\boldsymbol{\lambda}), \tag{55}$$

$$V_{1v}(\boldsymbol{\lambda}) = iF(\boldsymbol{\lambda})K_z(2+A_0)\lambda_y/k_0(\boldsymbol{\lambda}), \tag{56}$$

in terms of which we will calculate the scattering cross section later.

IV. BACKSCATTERING CROSS SECTION

Let us calculate the backscattering cross section in terms of the second-order solution obtained above. Inserting (42), (48), (52), and (55) into (23), we find

$$\sigma_{hh}^B(\theta) = \frac{16\pi k^4\cos^4(\theta)|F(-2k\sin\theta\,\mathbf{e}_x)|^2}{|1+\cos\theta\, Z_{sh}(k\sin\theta)|^4} + \frac{32\pi k^8\cos^4(\theta)}{|1+\cos\theta\, Z_{sh}(k\sin\theta)|^4}\int_0^{2\pi}\int_0^{\infty}|F(q^+)F(q^-)|^2$$
$$\times\left|\frac{(1-p^2)^{1/2}\cos^2(\alpha)}{1+(1-p^2)^{1/2}Z_{sh}(kp)} + \frac{\sin^2(\alpha)}{(1-p^2)^{1/2}+Z_{sv}(pk)}\right|^2 p\,dp\,d\alpha, \tag{57}$$

where we have

$$q^+ = k[(\sin\theta + p\cos\alpha)^2 + p^2\sin^2\alpha]^{1/2},$$
$$q^- = k[(\sin\theta - p\cos\alpha)^2 + p^2\sin^2\alpha]^{1/2}. \tag{58}$$

We substitute (43), (46), (52), and (56) into (24) to get the depolarized backscattering cross section

$$\sigma_{hv}^B(\theta) = \frac{32\pi k^8\cos^4(\theta)}{|1+\cos\theta\, Z_{sh}(k\sin\theta)|^2|\cos\theta + Z_{sv}(k\sin\theta)|^2}$$
$$\times\int_0^{2\pi}\int_0^{\infty}|F(q^+)F(q^-)|^2\cos^2(\alpha)\sin^2(\alpha)\left|\frac{(1-p^2)^{1/2}}{1+(1-p^2)^{1/2}Z_{sh}(kp)} - \frac{1}{(1-p^2)^{1/2}+Z_{sv}(pk)}\right|^2 p\,dp\,d\alpha, \tag{59}$$

where relation (58) has been used. Equation (59) involves only a second-order effect from A_2 and B_2, whereas Eq. (57)

includes both a first- and second-order effect. If we neglect $Z_{sh}(kp)$ and $Z_{sv}(kp)$, the scattering cross section by (57) or by (59) agrees with the result by the second-order perturbation, which, however, diverges unphysically. Since $Z_{sh}(kp)$ and $Z_{sv}(kp)$ are always complex, as is shown in Fig. 2, our solution yields a finite backscattering cross section. Assuming the Gaussian roughness spectrum (7), where the two-dimensional integral in (57) or in (59) is easily reduced to the one-dimensional one, we calculate the backscattering cross section, which is shown in Figs. 5 and 6.

V. VERTICAL POLARIZED INCIDENCE

We have considered the horizontally polarized case. In this section we will study a vertical polarized case. We write the primary wave over the nonfluctuating surface with $\sigma^2 = 0$ as

$$\mathbf{E}^0(z,\mathbf{r}) = e^{iK_0x}[\mathbf{e}_V(\mathbf{K}')e^{-iK_zz} + \mathbf{e}_V(\mathbf{K})e^{iK_zz}], \quad (60)$$

where $\mathbf{K}'$ and $\mathbf{K}$ are wave vectors of the incident wave and of the specularly scattered wave again. The perturbed electric field $\mathbf{E}^s(z,\mathbf{r},\omega)$ is also written by (18), the Wiener–Hermite expansion; the coefficients A_n and B_n are, of course, different from those in the horizontally polarized case and hence should be solved again from the boundary condition (8). From (10), (18), (60), and (A9) we have the coherent wave field

$$\langle \mathbf{E}(z,\mathbf{r},\omega)\rangle = e^{iK_0x}\{\mathbf{e}_V(\mathbf{K}')e^{-iK_zz} + [A_0\mathbf{e}_H(\mathbf{K}) + (1+B_0)\mathbf{e}_V(\mathbf{K})]e^{iK_zz}\}. \quad (61)$$

Since A_n and B_n describe the amplitude of the horizontally and vertically polarized partial waves, the backscattering cross section is given as

$$\sigma^B_{vv}(\theta) = 4\pi k^2\cos^2(\theta)\left\{|B_1(-2k\sin\theta\,\mathbf{e}_x)|^2 + 2!\int_{R^2}|B_2(-2k\sin\theta\,\mathbf{e}_x - \boldsymbol{\lambda}',\boldsymbol{\lambda}')|^2 d\boldsymbol{\lambda}' + \cdots\right\}, \quad (62)$$

$$\sigma^B_{vh}(\theta) = 4\pi k^2\cos^2(\theta)\left\{|A_1(-2k\sin\theta\,\mathbf{e}_x)|^2 + 2!\int_{R^2}|A_2(-2k\sin\theta\,\mathbf{e}_x - \boldsymbol{\lambda}',\boldsymbol{\lambda}')|^2 d\boldsymbol{\lambda}' + \cdots\right\}, \quad (63)$$

where subscripts vv and vh indicate the vertical transmission–vertical reception and vertical transmission–horizontal reception, respectively.

Substituting (4), (10), (18), and (60) into (25), the boundary condition, and using the orthogonality relation and the recurrence formula of $h^{(n)}$, we obtain another set of equations for A_n and B_n. The equations can be solved by a method similar to that described above. Omitting the details of the calculation, we only write the result for the isotropic random surface:

$$A_0 = 0, \quad B_0 = -2kZ_{sv}(\mathbf{K}_0)/[K_z + kZ_{sv}(\mathbf{K}_0)], \quad (64)$$

$$A_1(\boldsymbol{\lambda}) = [k/[k + k_z(\boldsymbol{\lambda})Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})]]\,V_{1h}(\boldsymbol{\lambda}), \quad (65a)$$

$$V_{1h}(\boldsymbol{\lambda}) = -ik(2+B_0)\lambda_y\,F(\boldsymbol{\lambda})/k_0(\boldsymbol{\lambda}), \quad (65b)$$

$$B_1(\boldsymbol{\lambda}) = [k/[k_z(\boldsymbol{\lambda}) + kZ_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})]]\,V_{1v}(\boldsymbol{\lambda}), \quad (66a)$$

$$V_{1v}(\boldsymbol{\lambda}) = \frac{i(2+B_0)[K_0k_0^2(\boldsymbol{\lambda}) - k^2(K_0+\lambda_x)]\,F(\boldsymbol{\lambda})}{kk_0(\boldsymbol{\lambda})}, \quad (66b)$$

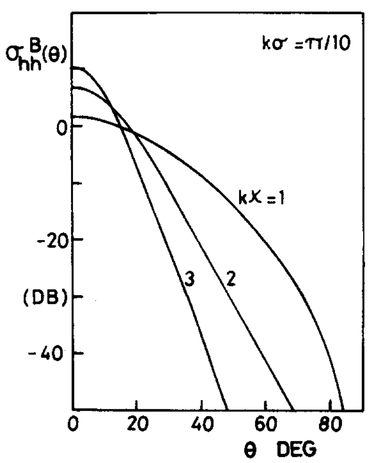

FIG. 5. Backscattering cross section per unit surface for horizontal transmission–horizontal reception. θ is a scattering angle, and the Gaussian roughness spectrum (7) is assumed. The correlation length of the random surface is denoted by κ.

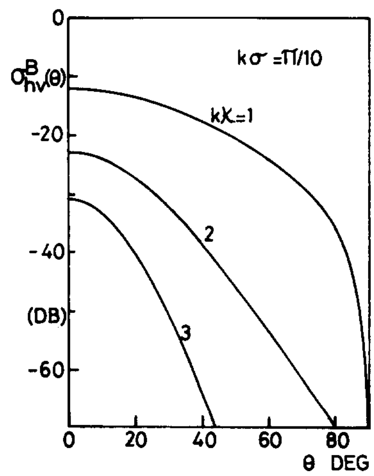

FIG. 6. Backscattering cross section with depolarization for the Gaussian roughness spectrum (7). The correlation length of the random surface is denoted by κ. By (72) it holds that $\sigma^B_{vh}(\theta) = \sigma^B_{hv}(\theta)$.

$$A_2(\lambda,\lambda') = \left[\frac{k}{k + k_z(\lambda+\lambda')Z_{sh}(\mathbf{K}_0+\lambda+\lambda')}\right][V_{2h}(\lambda,\lambda') + V_{2h}(\lambda',\lambda)], \tag{67a}$$

$$V_{2h}(\lambda,\lambda') = -\frac{iF(\lambda)[k_z(\lambda')\mathbf{k}_0(\lambda')\mathbf{k}_0(\lambda+\lambda')A_1(\lambda') - kk_0(\lambda')\{\lambda\mathbf{e}_H[\lambda']\}B_1(\lambda')]}{2k_0(\lambda')k_0(\lambda+\lambda')}, \tag{67b}$$

$$B_2(\lambda,\lambda') = \left[\frac{k}{k_z(\lambda+\lambda') + kZ_{sv}(\mathbf{K}_0+\lambda+\lambda')}\right][V_{2v}(\lambda,\lambda') + V_{2v}(\lambda',\lambda)], \tag{68a}$$

$$V_{2v}(\lambda,\lambda') = -\frac{iF(\lambda)}{2k_0(\lambda+\lambda')}\left[k_z(\lambda')\{\lambda\mathbf{e}_H[\lambda']\}A_1(\lambda') + \frac{[k^2\mathbf{k}_0(\lambda')\mathbf{k}_0(\lambda+\lambda') - k_0^2(\lambda')\mathbf{k}_0^2(\lambda+\lambda')]B_1(\lambda')}{kk_0(\lambda')}\right]. \tag{68b}$$

Here Z_{sh} and Z_{sv} are effective surface impedances given by (35) and (36), respectively. Therefore, the A_n's and the B_n's have the same resonance factors as the A_n's and the B_n's in the case of horizontally polarized incidence. Hence, they are also represented by equivalent networks shown in Fig. 4, but the output voltage V's take different values from those for the horizontal case. We note that (68) and (69) are exactly the same as (42) and (43), respectively; however, (65) and (66) should be substituted into (67) and (68) in the vertically polarized case.

Equation (64) physically means no depolarization in the coherent scattering. Substituting (64) into (61), we obtain the coherent wave field

$$\langle\mathbf{E}(z,\mathbf{r},\omega)\rangle = e^{iK_0x}[\mathbf{e}_V(\mathbf{K}')e^{-iK_zz} + \Gamma_v(\mathbf{K}_0)\mathbf{e}_V(\mathbf{K})e^{iK_zz}], \tag{69}$$

where $\Gamma_v(\mathbf{K}_0)$ is the coherent reflection coefficient for the vertically polarized wave

$$\Gamma_v(\mathbf{K}_0) = 1 + B_0 = [K_z - kZ_{sv}(\mathbf{K}_0)]/[K_z + kZ_{sv}(\mathbf{K}_0)], \tag{70}$$

which also is represented by an equivalent network shown in Fig. 4(b). If we expand (70) in a power series of $kZ_{sv}(\mathbf{K}_0)/K_z$ and keep the first two terms, we have the same result by the second-order perturbation,[5]

$$\Gamma_v(\mathbf{K}_0) \simeq 1 - (2k/K_z)Z_{sv}(\mathbf{K}_0),$$

which, however, diverges incorrectly for the Rayleigh wave number with $K_z = k\cos\theta_0 = 0$, where θ_0 is the angle of incidence of Fig. 1. Our solution (70), however, always remains finite for any real angle of incidence, but may diverge for a complex angle of incidence. Such a divergence suggests the existence of the guided complex wave.

Inserting (64)–(67) into (62) and (63), we find σ_{vv}^B:

$$\sigma_{vv}^B(\theta) = \frac{16\pi k^4\cos^4\theta(1+\sin^2\theta)^2\,|F(-2k\sin\theta)|^2}{|\cos\theta + Z_{sv}(k\sin\theta)|^4} + \frac{32\pi k^8\cos^4(\theta)}{|\cos\theta + Z_{sv}(k\sin\theta)|^4}\int_0^{2\pi}\int_0^\infty |F(q^+)F(q^-)|^2 \times\left|\frac{(1-p^2)^{1/2}\sin^2(\alpha)}{1+(1-p^2)^{1/2}Z_{sh}(kp)} + \frac{\cos^2(\alpha) - p^2\sin^2\theta}{(1-p^2)^{1/2} + Z_{sv}(pk)}\right|^2 p\,dp\,d\alpha, \tag{71}$$

where q^+ and q^- have been defined by (58). Numerical examples of (71) are illustrated in Fig. 7. We insert (64)–(68) into (63) to get $\sigma_{vh}^B(\theta)$. We then have the same equation as (59), namely,

$$\sigma_{vh}^B(\theta) = \sigma_{hv}^B(\theta), \tag{72}$$

which means that our second-order solution satisfies the reciprocal relation.

VI. DISCUSSIONS ON GUIDED SURFACE WAVES

As is well known, a corrugated periodic surface can support TM and TE guided surface waves.[33] A similar phenomenon may exist for a randomly corrugated surface and has been studied extensively[3–5,19,34]; the relation between the anomalous scattering and the guided wave was studied in Ref. 18. We first point out that a guided surface wave exists over a perfectly conductive surface without any roughness (see Fig. 8). In fact, one easily finds that a vertically polarized wave

$$\mathbf{E}_g(z,\mathbf{r}) = G\mathbf{e}_z\exp[i(\mathbf{K}_0+\lambda)\mathbf{r}],$$
$$|\mathbf{K}_0+\lambda| = k, \tag{73}$$

satisfies Maxwell's equations, the divergence-free condition ($\operatorname{div}\mathbf{E}_g = 0$), and the boundary condition $\mathbf{n}\times\mathbf{E}_g = \mathbf{e}_z\times\mathbf{E}_g = 0$ on the perfectly conductive flat surface for any constant G. Therefore, Eq. (73) may be a guided wave, which we call the unperturbed guided wave. In Rice's perturbation approach, the unperturbed guided wave is coupled and is excited by the surface roughness with the incident plane wave. Because the unperturbed guided wave has a real propagation constant equal to the so-called Rayleigh wave number (73), it has an infinite amplitude when it is so excited. As a result, many statistical properties of the scattering diverge unphysically; however, the scattering cross section by the first-order perturbation remains finite.

On the other hand, such a guided wave is treated more precisely in our theory, which takes into account multiple

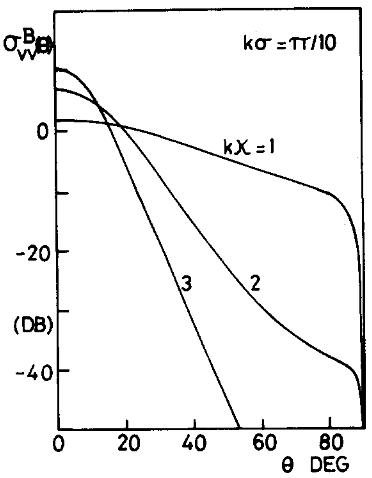

FIG. 7. Backscattering cross section for vertical transmission–vertical reception. The correlation length of the random surface is denoted by κ.

scattering. When the surface gets rough, such a guided wave is also scattered by the roughness, yielding outgoing waves scattered in any directions. Because of energy dissipation due to the scattering process, such a guided wave is physically expected to become a perturbed guided wave with a complex propagation constant. Our solution does include the perturbed guided waves; the complex propagation constant $\mathbf{K}_0+\boldsymbol{\lambda}$ of the perturbed guided wave is determined by a condition under which, for example, $B_1(\boldsymbol{\lambda})$ diverges. From (49) or (66), we have the condition[35]

$$k_z(\boldsymbol{\lambda})+kZ_{sv}(\mathbf{K}_0+\boldsymbol{\lambda}) = [k^2-(\mathbf{K}_0+\boldsymbol{\lambda})^2]^{1/2}+kZ_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})=0, \quad (74)$$

which is actually the transverse resonance condition (resonance condition in the z direction, see Fig. 4); a similar condition for $B_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$ or B_0 is obtained from (74) by replacing $\boldsymbol{\lambda}$ with $\boldsymbol{\lambda}+\boldsymbol{\lambda}'$ or by putting $\boldsymbol{\lambda}=0$. Since Z_{sv} is a quantity of the order of σ^2, one may find a complex propagation constant $\mathbf{K}_0+\boldsymbol{\lambda}$ by solving (73) by the small perturbation

$$k_z(\boldsymbol{\lambda}) = -kZ_{sv}(k) = -\int_{R^2}\frac{k\lambda_x^2}{\sqrt{k^2-(k+\lambda_x)^2-\lambda_y^2}}|F(\boldsymbol{\lambda})|^2\,d\boldsymbol{\lambda},$$
$$\mathbf{K}_0+\boldsymbol{\lambda}=k\,[1-Z_{sv}(k)^2/2](\cos\beta\mathbf{e}_x+\sin\beta\mathbf{e}_y), \quad (75)$$

where β is an arbitrary angle representing a direction of propagation, because of the isotropic surface assumed. Equation (75) corresponds to the first peak near $p=1$ in Fig. 3(b). We note that Eq. (75) is essentially the same as Rice's solution for the guided wave [cf. (6.18) in Ref. 5]. Thus, in our theory the incident plane wave couples with such a perturbed guided wave, i.e., where the amplitude becomes large but remains finite when it is excited by the incident plane wave, as is suggested by Fig. 3(b), whereas in Rice's theory the incident wave excites the unperturbed guided wave.

The number of guided waves propagating along a perfectly conductive slightly random surface has been considered to be 1 in the literature.[3,5] But we point out here that the other guided waves can exist.

As is seen from (36b) or in Fig. 2(b), however, $Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})$ behaves asymptotically for $|\mathbf{K}_0+\boldsymbol{\lambda}|\gg k$ as

$$Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})\sim -iC_v\sigma^2(\mathbf{K}_0+\boldsymbol{\lambda})^2, \quad (76)$$

where a small real part of $Z_{sv}(\mathbf{K}_0+\boldsymbol{\lambda})$ has been neglected and C_v is a constant depending on the correlation length of the surface. The transverse resonance condition (75) may have another complex zero when

$$\mathbf{K}_0+\boldsymbol{\lambda}\simeq(C_v\sigma^2)^{-1}(\cos\beta\mathbf{e}_x+\sin\beta\mathbf{e}_y), \quad (77)$$

which approximately gives the real part of the propagation constant of another guided wave with vertical polarization, but the imaginary part is neglected in (77), where β is an arbitrary angle again. Equation (77) corresponds to the second peak about $p=3.3$ in Fig. 3(b).

There is another guided complex wave with horizontal polarization. In the horizontally polarized case the transverse resonance condition is obtained from A_n, for example,

$$k+k_z(\boldsymbol{\lambda})Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})=0, \quad (78)$$

under which $A_1(\boldsymbol{\lambda})$ diverges. The transverse resonance condition for A_0 or $A_2(\boldsymbol{\lambda},\boldsymbol{\lambda}')$ may be obtained from (78) by putting $\boldsymbol{\lambda}=0$ or by replacing $\boldsymbol{\lambda}$ with $\boldsymbol{\lambda}+\boldsymbol{\lambda}'$. Since $Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})$ by (35b) behaves asymptotically for $|\mathbf{K}_0+\boldsymbol{\lambda}|\gg k$ as [see Fig. 2(a)]

$$Z_{sh}(\mathbf{K}_0+\boldsymbol{\lambda})\sim iC_h\sigma^2|\mathbf{K}_0+\boldsymbol{\lambda}| \quad (|\mathbf{K}_0+\boldsymbol{\lambda}|\gg k), \quad (79)$$

Eq. (78) may have a complex zero when

$$\mathbf{K}_0+\boldsymbol{\lambda}\simeq(C_h\sigma^2)^{1/2}(\cos\beta\mathbf{e}_x+\sin\beta\mathbf{e}_y), \quad (80)$$

which gives the real part of the propagation constant and corresponds to the peak in Fig. 3(a). However, the imaginary part of the propagation constant is neglected again in (80). We note that (80) is proportional to $(\sigma)^{-1}$, whereas Eq. (77) is proportional to $(\sigma)^{-2}$. Because these guided surface waves have propagation constant $\mathbf{K}_0+\boldsymbol{\lambda}$ much greater than k, they may not be excited by an incident plane wave for a slightly rough (small σ^2) and gently sloping surface. When the correlation length of the surface is relatively small ($k\kappa\simeq1$), however, the perturbed guided wave with (75) is strongly excited and is scattered again by the surface roughness to yield a relatively large depolarization in backscatter. For example, we see in Figs. 5–7 that, when $\theta=80$ (deg), $k\sigma=\pi/10$, and $k\kappa=1$, $\sigma^B_{hv}(\theta)$ becomes -36.1 dB, which is slightly larger than $\sigma^B_{hh}(\theta)=-41.5$ dB but is much smaller than $\sigma^B_{vv}(\theta)=-10.5$ dB.

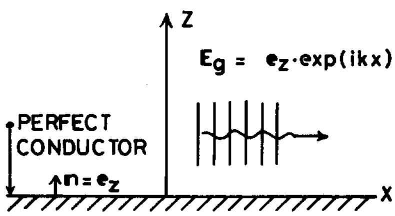

FIG. 8. Unperturbed guided wave propagating into the x direction along the perfectly conductive flat surface without any roughness. In this figure the guided wave has the wave vector equal to the so-called Rayleigh wave number $\mathbf{k}=(k_x,k_y,k_z)=(k,0,0)$.

VII. CONCLUSIONS

We have discussed the scattering of electromagnetic plane waves from a perfectly conductive slightly random surface by a probabilistic method. Assuming an isotropic random surface, we have given a definite solution for the backscattering cross section including the depolarization effect. Since the solution is written in a closed form, it may be readily used by the engineer. Introducing an equivalent network representation for coefficients of the Wiener–Hermite expansion of the scattered wave, we have found out that there exist two modes of the guided surface wave with vertical polarization and one mode with horizontal polarization. The theory presented in this paper is easily extended to a case of an anisotropic random surface. Such a case will be studied elsewhere.

ACKNOWLEDGMENT

The authors would like to thank H. Ogura of Kyoto Institute of Technology for his stimulating discussions.

APPENDIX: WIENER–HERMITE EXPANSION

Following Ogura's[25,28] notations and definitions, we summarize formulas concerning the Wiener–Hermite expansion, which is a slightly generalized version of Wiener's nonlinear theory of the Brownian motion process.[29] However, for the ergodic theory and detailed mathematical description of the expansion, see Itô[30] and Wiener.[29]

1. Gaussian random measure

Let us start with the basic hypothesis: there is a probability measure $P(\omega)$ on a Borel field of ω sets of the sample space Ω, ω being a probability parameter describing a sample point in Ω. We denote by $d\mathbf{r} = dx\,dy$ an infinitestimal area at $\mathbf{r} = (x, y)$ in the two-dimensional plane $R^2 = (-\infty < x, y < \infty)$. We introduce the Gaussian random measure $dB(\mathbf{r},\omega) = B(d\mathbf{r},\omega)$ on R^2, which is a Gaussian random variable with parameter $\mathbf{r}$ satisfying

$$\langle dB(\mathbf{r},\omega)\rangle = 0\,, \quad \langle dB(\mathbf{r},\omega)dB(\mathbf{r}',\omega)\rangle = \delta(\mathbf{r}-\mathbf{r}')d\mathbf{r}\,d\mathbf{r}'\,, \tag{A1}$$

Here $\langle\quad\rangle$ denotes the ensemble average, i.e., the average over the sample space, and the symbol $\delta(\mathbf{r}-\mathbf{r}')d\mathbf{r}\,d\mathbf{r}'$ stands for zero when $\mathbf{r}\neq\mathbf{r}'$ or $d\mathbf{r}$ when $\mathbf{r}=\mathbf{r}'$. We assume that $dB(\mathbf{r},\omega)$ is of a function space type (Doob[31]), where the sample space is assumed to be an infinite-dimensional Euclidian space R^{R^2} and a sample function $dB(\mathbf{r},\omega)$ is projected to an infinite-dimensional vector ω in $\Omega = R^{R^2}$ in such a manner that $\omega_\mathbf{r}$, the $\mathbf{r}$ component of ω, is given by

$$\omega_\mathbf{r} = dB(\mathbf{r},\omega)\,. \tag{A2}$$

Thus a function $g(\omega)$ of ω is always regarded as a functional of $dB(\mathbf{r},\omega)$, i.e., $g(\omega) = g[dB(\cdot,\omega)]$. Because of (A2), a shift of the sample function $dB(\mathbf{r},\omega)$ by $\mathbf{a}$ generates a shift of ω vector in the sample space. If we define the shift in the sample space by the relation

$$dB(\mathbf{r}+\mathbf{a},\omega) = dB(\mathbf{r},T^\mathbf{a}\omega)\,, \quad \mathbf{a}\in R^2\,, \tag{A3}$$

the shift $T^\mathbf{a}$ in the sample space becomes a group: $T^0 = 1$ (identity); $T^\mathbf{a}T^\mathbf{b} = T^{\mathbf{a}+\mathbf{b}}$. Furthermore the shift $T^\mathbf{a}$ becomes a measure-preserving transformation, because $dB(\mathbf{r},\omega)$ is a strictly homogeneous random function of which probability measure has the shift invariance $P(\omega) = P(T^\mathbf{a}\omega)$. This means that, if a sample function $dB(\mathbf{r},\omega)$ exists with some probability $P(\omega)$, the shifted sample function $dB(\mathbf{r}+\mathbf{a},\omega)$ always exists with the same probability $P(\omega)$. Once the measure-preserving transformation $T^\mathbf{a}$ is defined by (A3), $g(T^\mathbf{a}\omega)$ becomes a homogeneous random function for any random variable $g(\omega)$. Such a random function is often called a homogeneous random function generated by the shift. Note that a homogeneous random function generated by the shift is invariant under a translation in the product space $R^2\times\Omega$ that carries $(\mathbf{r},\omega)$ into $(\mathbf{r}+\mathbf{a},T^{-\mathbf{a}}\omega)$. This translation invariance is taken in order to determine a possible form of the scattered wave field.

2. Complex Gaussian random measure

We define the complex Gaussian random measure $dB(\boldsymbol{\lambda},\omega)$ by the Fourier transform of the Gaussian random measure $dB(\mathbf{r},\omega)$ [the same notation B is used for two different quantities, $B(\mathbf{r},\omega)$ and $B(\boldsymbol{\lambda},\omega)$, which are easily identified by their arguments without confusion]:

$$dB(\boldsymbol{\lambda},\omega) = \frac{d\boldsymbol{\lambda}}{2\pi}\int_{R^2} e^{-i\boldsymbol{\lambda}\mathbf{r}}\,dB(\mathbf{r},\omega)\,, \tag{A4}$$

which is a complex-valued random variable, such that its real and imaginary parts have an independent identical Gaussian distribution with

$$\langle dB(\boldsymbol{\lambda},\omega)\rangle = 0\,,$$
$$\langle dB(\boldsymbol{\lambda},\omega)dB(\boldsymbol{\lambda}',\omega)\rangle = \delta(\boldsymbol{\lambda}+\boldsymbol{\lambda}')\,d\boldsymbol{\lambda}\,d\boldsymbol{\lambda}'\,,$$
$$dB(\boldsymbol{\lambda},\omega) = dB^*(-\boldsymbol{\lambda}',\omega)\,, \tag{A5}$$

where the asterisk denotes the complex conjugate. Using (A3) and (A4), we find the translation property of $dB(\boldsymbol{\lambda},\omega)$ under $T^\mathbf{a}$,

$$dB(\boldsymbol{\lambda},T^\mathbf{a}\omega) = \exp(i\mathbf{a}\boldsymbol{\lambda})dB(\boldsymbol{\lambda},\omega)\,, \tag{A6}$$

which holds in means square sense (Itô[30]).

3. Wiener–Hermite differentials

Using the complex Gaussian random measure, we define the Wiener–Hermite differentials $h^{(n)}[dB(\boldsymbol{\lambda}_1),\ldots,dB(\boldsymbol{\lambda}_n)]$ $(n = 0,1,2,\ldots)$ by the relations (in what follows, however, we often drop the probability parameter ω)

$$h^{(0)} = 1\,, \quad h^{(1)}[dB(\boldsymbol{\lambda})] = dB(\boldsymbol{\lambda})\,,$$
$$h^{(2)}[dB(\boldsymbol{\lambda}_1), dB(\boldsymbol{\lambda}_2)]$$
$$= dB(\boldsymbol{\lambda}_1)dB(\boldsymbol{\lambda}_2) - \delta(\boldsymbol{\lambda}_1+\boldsymbol{\lambda}_2)\,d\boldsymbol{\lambda}_1\,d\boldsymbol{\lambda}_2\,, \tag{A7}$$

etc. Higher-order differentials may be obtained by the recurrence formula

$$dB(\boldsymbol{\lambda}_1)h^{(n-1)}[dB(\boldsymbol{\lambda}_2),\ldots,dB(\boldsymbol{\lambda}_n)]$$
$$= h^{(n)}[dB(\boldsymbol{\lambda}_1),\ldots,dB(\boldsymbol{\lambda}_n)]$$
$$+ \sum_{i=2}^{n} h^{(n-2)}[dB(\boldsymbol{\lambda}_2),\ldots,dB(\boldsymbol{\lambda}_{i-1}),dB(\boldsymbol{\lambda}_{i+1}),$$
$$\ldots,dB(\boldsymbol{\lambda}_n)]\,\delta(\boldsymbol{\lambda}_1+\boldsymbol{\lambda}_i)\,d\boldsymbol{\lambda}_1\,d\boldsymbol{\lambda}_i\,. \tag{A8}$$

These differentials satisfy the orthogonality relation

$$\langle h^{(n)}[dB(\lambda_{i1}),...,dB(\lambda_{in})]h^{*(m)}[dB(\lambda_{j1}),...,dB(\lambda_{jm})]\rangle = \delta_{nm}\delta_{ij}^{n} d\lambda_{i1}\ldots d\lambda_{in}\, d\lambda_{j1}\ldots d\lambda_{jm}, \quad (A9)$$

where δ_{ij}^n equals to the sum of all distinct products of n delta functions of the form $\delta(\lambda_{i\alpha}-\lambda_{j\mu})$, $i=(i1,i2,...,in)$, $j=(j1,j2,...,jm)$, all $i\alpha$ and $j\mu$ appearing just once in each product, for example,

$$\delta_{ij}^2 = \delta(\lambda_{i1}-\lambda_{j1})\delta(\lambda_{i2}-\lambda_{j2}) + \delta(\lambda_{i1}-\lambda_{j2})\delta(\lambda_{i2}-\lambda_{j1}). \quad (A10)$$

By (A5)–(A7) the Wiener–Hermite differential is translated under T^a as

$$h^{(n)}[dB(\lambda_1,T^a\omega),...,dB(\lambda_n,T^a\omega)] = \exp[ia(\lambda_1+\cdots+\lambda_n)] \times h^{(n)}[dB(\lambda_1,\omega),...,dB(\lambda_n,\omega)]. \quad (A11)$$

4. Orthogonal development of a stochastic functional

We denote by $g(\omega)$ a stochastic functional of the complex Gaussian random measure. If the functional $g(\omega)$ has a finite variance, it has the orthogonal development called the Wiener–Hermite expansion,

$$g(\omega) = \sum_{n=0}^{\infty} \int_{R^2}\cdots\int_{R^2} A_n(\lambda_1,...,\lambda_n) \times h^{(n)}[dB(\lambda_1),...,dB(\lambda_n)], \quad (A12)$$

where A_n is a symmetric function with respect to its argument. By (A9) these coefficients are given by the correlation

$$\langle g(\omega)h^{*(n)}[dB(\lambda_1),...,dB(\lambda_n)]\rangle = n!\,A_n(\lambda_1,...,\lambda_n)d\lambda_1\ldots d\lambda_n. \quad (A13)$$

By (A11) a homogeneous random function $g(T^r\omega)$ derived by the shift is represented as

$$g(T^r\omega) = \sum_{n=0}^{\infty} \int_{R^2}\cdots\int_{R^2} A_n(\lambda_1,...,\lambda_n) \times\exp[ir(\lambda_1+\cdots+\lambda_n)] \times h^{(n)}[dB(\lambda_1),...,dB(\lambda_n)], \quad (A14)$$

which holds as a function of ω in the mean square sense, namely,

$$\lim_{N\to\infty}\left\langle\left|g(T^r\omega) - \sum_{n=0}^{N}\int_{R^2}\cdots\int_{R^2} A_n(\lambda_1,...,\lambda_n) \times\exp[ir(\lambda_1+\cdots+\lambda_n)] \times h^{(n)}[dB(\lambda_1),...,dB(\lambda_n)]\right|^2\right\rangle = 0. \quad (A15)$$

The average and the correlation of $g(T^r\omega)$ are easily calculated by the orthogonality relation (A9) as

$$\langle g(T^r\omega)\rangle = A_0, \quad (A16)$$

$$\langle g(T^r\omega)g^*(\omega)\rangle = \sum_{n=0}^{\infty} n!\int_{R^2}\cdots\int_{R^2}|A_n(\lambda_1,...,\lambda_n)|^2 \times\exp[ir(\lambda_1+\cdots+\lambda_n)]d\lambda_1\cdots d\lambda_n. \quad (A17)$$

Thus the first term of the development in (A14) is equal to the average. We note that these statistical moments can be obtained by the space average over the two-dimensional plane R^2, because the measure-preserving transformation T^r is ergodic in case of the Gaussian random measure.[30]

[1]P. Beckmann and A. Sppichichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon, New York, 1963), pp. 70–98.

[2]A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 2, pp. 463–492.

[3]F. G. Bass and I. M. Fuks, *Wave Scattering from Statistically Rough Surfaces* (Pergamon, New York, 1981).

[4]*Surface Polaritons*, edited by V. M. Agranovich and D. L. Mills (North-Holland, Amsterdam, 1982).

[5]S. O. Rice, "Reflection of electromagnetic waves from slightly rough surfaces," Commun. Pure Appl. Math. **4**, 351 (1951).

[6]G. R. Valenzuela, "Depolarization of em waves by a slightly random surfaces," IEEE Trans. Antennas Propag. **AP-15**, 552 (1967).

[7]V. D. Freilikher and I. M. Fuks, "Green's function method for the Helmholtz equation with perturbed boundary condition," Izv. Vyssh Uchebn. Zaved. Radiopiz. **13**, 98 (1970).

[8]G. G. Zipfel and J. A. DeSanto, "Scattering of a scalar wave from a random rough surface: A diagrammatic approach," J. Math. Phys. **13**, 1903 (1972).

[9]J. A. DeSanto, "Green's function for electromagnetic wave scattering from a random rough surface," J. Math. Phys. **15**, 283 (1974).

[10]F. Toigo, A. Marvin, V. Celli, and N. R. Hill, "Optical properties of Rough surfaces: General theory and the small roughness limit," Phys. Rev. **15**, 5618 ((1977).

[11]M. Nieto-Vesperinas, "Depolarization of electromagnetic waves scattered from slightly random surfaces: a study by means of the extinction theorem," J. Opt. Soc. Am. **72**, 539 (1982).

[12]E. Bahar, "Scattering cross section for random rough surfaces: Full wave analysis," Radio Sci. **16**, 331 (1981).

[13]E. Bahar, "Full wave solution for the depolarization of the scattered radiation fields by rough surfaces of arbitrary slope," IEEE Trans. Antennas Propag. **AP-29**, 443 (1981).

[14]J. Nakayama, H. Ogura, and B. Matsumoto, "A probabilistic theory of scattering from a randomly rough surface," Radio Sci. **15**, 1049 (1980).

[15]J. Nakayama, H. Ogura, and M. Sakata, "Scattering of a scalar wave from a slightly random surface," J. Math. Phys. **22**, 471 (1981).

[16]J. Nakayama, H. Ogura, and M. Sakata, "A probabilistic theory of electromagnetic wave scattering from a slightly random surface: 1. Horizontal polarization," Radio Sci. **16**, 831 (1981).

[17]J. Nakayama, M. Sakata, and H. Ogura, "A probabilistic theory of electromagnetic wave scattering from a slightly random surface: 2. Vertical polarization," Radio Sci. **16**, 847 (1981).

[18]J. Nakayama, "Anomalous scattering from a slightly random surface," Radio Sci. **17**, 558 (1982).

[19]J. Nakayama, "A note on the guided complex wave supported by a slightly random surface," Inst. Electron. Commun. Eng. Jpn. **1983**, 202.

[20]J. Nakayama, K. Mitzutani, H. Ogura, and S. Hayashi, "Theory of light scattering from a random metal surface: Excitation of surface plasmons in a Ag film," J. Appl. Phys. **56**, 1462 (1984).

[21]S. Ito, "Analysis of scalar wave scattering from slightly rough surfaces: A multiple scattering theory," J. Radio Res. Lab. **31**, 1 (1984).

[22]G. S. Brown, "Back-scattering from a Gaussian-distributed perfectly conducting rough surfaces," IEEE Trans. Antennas Propag. **AP-28**, 788 (1978).

[23]G. S. Brown, "A stochastic Fourier transform approach to scattering from perfectly conducting randomly rough surfaces," IEEE Trans. Antennas Propag. **AP-30**, 1135–1144 (1982).

[24]J. G. Watson and J. B. Keller, "Rough surface scattering via the smoothing method," J. Acoust. Soc. Am. **75**, 1705 (1984).

[25]H. Ogura, "Theory of waves in a homogeneous random medium," Phys. Rev. A **11**, 942 (1975).

[26]H. Ogura and J. Nakayama, "Initial-value problem of the one-dimensional wave propagation in a homogeneous random medium," Phys. Rev. A **11**, 957 (1975).

[27]H. Ogura and Y. Yoshida, "Wave propagation in a slightly lossy random medium," Phys. Rev. A **14**, 796 (1976).

[28]H. Ogura, *Kakuritsu-Katei-Ron* (Corona, Tokyo, 1978), Chap. 5 (in Japanese).

[29]N. Wiener, *Nonlinear Problems in Random Theory* (M.I.T., Cambridge, MA, 1958).

[30]K. Itô, "Multiple Wiener integrals," J. Math. Soc. Jpn. **3**, 157 (1851); K. Itô, "Complex multiple Wiener integrals," Jpn. J. Math. **22**, 63 (1952).
[31]J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953), Chaps. X and XI.
[32]This is because $D^{\mathbf{a}}$ commutes with Maxwell's equations and the divergence-free condition. Therefore, if $\mathbf{E}^s(z,\mathbf{r},\omega)$ is a solution, $D^{\mathbf{a}}\mathbf{E}^s(z,\mathbf{r},\omega)$ becomes a solution for any vector $\mathbf{a}\in R^2$.
[33]R. E. Collin, *Field Theory of Guided Waves* (McGraw–Hill, New York, 1960), Chap. 11.
[34]D. E. Barrick, "Theory of HF and VHF propagation across the rough sea 1: the effective surface impedance for a slightly rough highly conductive medium at grazing incidence," Radio Sci. **6**, 517 (1971).
[35]The condition under which the coherent scattering diverges is usually considered as the transverse resonance condition determining the propagation constant of the guided surface wave. However, in our solution [(42), (43), (48), (49), (52), and (64)–(68)], a condition making A_n and B_n, $n = 1,2$, diverge also gives the transverse resonance condition. This is probably because of the approximation employed to get our solution.

The group-theoretical treatment of aberrating systems. I. Aligned lens systems in third aberration order

Miguel Navarro-Saad and Kurt Bernardo Wolf
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Apdo. Postal 20-726, 01000 México D.F., Mexico

(Received 8 February 1985; accepted for publication 13 November 1985)

The constituents of a lens system, i.e., slabs of homogeneous transparent material and the refracting surfaces between them, considered to third aberration order, are associated to elements of a nine-parameter *aberration* group. Three parameters correspond to Gaussian systems and six to group-classified aberrations. The group multiplication operation (through matrix-cum-vector algebra) corresponds to their concatenation, and the linear group action on an eight-dimensional homogeneous space corresponds to the nonlinear action of the system on the optical phase space. This leads to economical computation algorithms that may be extended to aberrating systems of higher order.

I. INTRODUCTION

A Lie-theoretical treatment of geometrical optics and optical aberrations has been given by Dragt[1] as an application of his work on magnetic optics.[2] Here we want to present a *group*-theoretical treatment that emphasizes the *finite* transformations that each optical element produces on the phase space of rays. These transformations are, in general, nonlinear and symplectic; they can be approximated to any desired degree by a product of exponentials of polynomials homogeneous in the phase space coordinates.[3] In this article we retain terms to third aberration order and work with the adjoint representation of the resulting quotient group. To first approximation, our *modus operandi* relates to that of Dragt and co-workers, as the 2×2 matrix representation of $Sp(2,R)$ elements relates to their parameters x_i in $\exp(\Sigma x_i J_i)$, where $\{J_i\}$ are the generators of the corresponding $sp(2,R)$ algebra. Conceptual and computational advantages may be found in considering both approaches.

Refracting surface transformations belong to a special factorizable type of canonical transformations, as we showed in Ref. 4, to which we may readily associate finite group elements. We would like to point out, moreover, the conceptual and calculational aspects that can be gained from a group-theoretical description of the system. This has been proven an invaluable aid in molecular, nuclear, and elementary-particle physics. Indeed, our group-theoretical "model" of third-order aberration aligned optics is the group $(T_5 \otimes T_1) ⓢ Sp(2,R)$. Here, $Sp(2,R)$ is the two-dimensional real symplectic group (i.e., the group of 2×2 unimodular matrices) of linear phase-space transformations accounting for Gaussian optics; $T_5 \otimes T_1 = T_6$ is a $(5+1=6)$-dimensional Abelian "translation" group, which transforms (here, in semidirect product ⓢ) as a symplectic quintuplet and a singlet under Gaussian transformations that produce cubic transformations of phase space, representing the six third-order aberrations. Having this nine-parameter group enables us to reduce calculations to matrix-cum-vector algebra through a minimal representation. This finite-parameter group is the quotient of the pseudogroup of all canonical transformations modulo higher-than-third-order transformations. No more than nine parameters can arise, and we are within third-order aberration optics in an essential way.

We are aware that group-theoretical methods are not yet standard in the study of optics, and we shall not assume the reader has detailed familiarity with them. The emphasis here is on the group at hand in matrix-cum-vector language. The general system including up-to-Nth-order compound aberrations will be published elsewhere.[5] It is the general case, however, which leads us to define the problem in the way we do, with the notation adapted to its future use. The particular case of third-order two-dimensional, aligned optical lens systems will be treated here, i.e., systems composed of free propagation through homogeneous media separated by axially symmetric refracting interfaces of up to fourth order.

The basic tools are presented in Sec. II: the optical Hamiltonian, Poisson brackets, evolution operators, and refracting-surface transformations of phase space. Section III organizes this information to bring out the Lie-algebraic structure of the system to third order, while in Secs. IV and V, we present our group-theoretic scheme and representation as cubic transformations of a phase space. The elements of a lens system are thereby described in Sec. VI and their aberrations in Sec. VII. In Sec. VIII we show how to concatenate these elements to a system through the group multiplication of their corresponding elements. The outlook offered in Sec. IX points to results on inhomogeneous optical media, whose treatment through traditional ray-tracing methods is sometimes problematic. They are easily incorporated in this group-theoretical treatment. Symplectic geometrization of classical mechanics is one of the gateways to quantum mechanics; the same path may be used to explore wave optics in aberrating systems.

II. THE HAMILTONIAN FORMULATION OF LENS OPTICS

Hamiltonian optics describes a light ray as a line in phase space $\mathbf{p}(z)$, $\mathbf{q}(z)$ along the optical axis z. In the physical case $\mathbf{q}$ is a two-vector with the *position* coordinates, q_1, q_2, of the ray at the $z=0$ reference plane. The canonically con-

 0022-2488/86/051449-09$02.50

jugate *momentum* $\mathbf{p}$ is also a two-vector, it is in the projection on the $z = 0$ plane of the direction of the ray, and has magnitude $p = n \sin \theta$, where n is the refraction index of the medium at $\mathbf{q}$ and θ is the angle between the ray direction and the z axis.

The Hamiltonian of an optical system is given by[1,6]

$$H(\mathbf{p},\mathbf{q};z) = -n\cos\theta = -\sqrt{n^2 - p^2}$$
$$= -n + \frac{1}{2n}p^2 + \frac{1}{8n^3}p^4 + \mathcal{O}(p^6), \quad (2.1)$$

where in general $n = n(\mathbf{p},\mathbf{q},z)$; this determines the z evolution of any observable function $f(\mathbf{p},\mathbf{q};z)$ of phase space through the differential equation

$$\frac{d\,\mathbf{f}(\mathbf{p},\mathbf{q};z)}{dz} = -\{H,f\}(\mathbf{p},\mathbf{q};z) = -\widehat{H} f(\mathbf{p},\mathbf{q};z). \quad (2.2)$$

Here $\{\cdot,\cdot\}$ is the usual Poisson bracket (Ref. 7, p. 252) with $\{q_i, p_j\} = \delta_{ij}$. For any continuously differentiable function $g(\mathbf{p},\mathbf{q})$ we define the first-order continuous operator $\hat{g}$, whose action on observables is

$$\hat{g} f := \{g,f\} = \sum_k \left(\frac{\partial g}{\partial q_k}\frac{\partial}{\partial p_k} - \frac{\partial g}{\partial p_k}\frac{\partial}{\partial q_k}\right) f. \quad (2.3)$$

A fundamental and well-known property[1] of the phase-space functions and their associated operators is that the latter's commutator is the operator associated to the Poisson bracket of the former, i.e., $[\hat{f},\hat{g}] = (\{f,g\})\hat{}$. [When $g(\mathbf{p},\mathbf{q})$ is displayed it is more convenient to denote $\hat{g}$ by delimiters such as $:g:$[1] or $\{g,\cdot\}$[8]; when need be, we may write $(g)\hat{}$.]

The z evolution of phase space produced by H is written, in general, as

$$f(\mathbf{p},\mathbf{q};z_0) \overset{z}{\mapsto} (\mathbb{G}(z)f)(\mathbf{p},\mathbf{q};z_0) = f(\mathbf{p},\mathbf{q};z_0 + z), \quad (2.4)$$

with the Green operator $\mathbb{G}(z)$ obtained through[2]

$$\dot{\mathbb{G}} = -\mathbb{G}\widehat{H}, \quad \mathbb{G}(0) = \mathbb{1} \quad \text{(identity operator)}. \quad (2.5)$$

When the Hamiltonian H is independent of z, as in free propagation in a medium where $n = n(\mathbf{q})$ only, then one may integrate this to

$$\mathbb{G}(z) = \exp(-z\widehat{H})$$
$$= \sum_{m=0}^{\infty} \frac{(-z)^m}{m!} \underbrace{\{H,\{H,\{\cdots\{H,\cdot\}\cdots\}\}\}}_{m}, \quad (2.6)$$

expressed in a series of m-fold Poisson brackets with H. For twice-differentiable H, the mapping (2.4) is *canonical*, i.e., $\{f,g\} = \{\mathbb{G}(z)f, \mathbb{G}(z)g\}$ in a finite neighborhood of $z = 0$. In particular, for the coordinates $f_k(\mathbf{p},\mathbf{q}) = q_k$, $g_k(\mathbf{p},\mathbf{q}) = p_k$, we denote $(\mathbb{G}(z)f_k)(\mathbf{p},\mathbf{q}) = q'_k$, $(\mathbb{G}(z)g_k)(\mathbf{p},\mathbf{q}) = p'_k$, $k = 1,2$.

Free propagation in a homogeneous medium (n constant) generated by (2.1) produces the following transformation of phase space:

$$\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} \overset{z}{\mapsto} \begin{pmatrix}\mathbf{p}'\\ \mathbf{q}'\end{pmatrix} = \mathbb{G}(z)\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p}\\ \mathbf{q} + z\,\mathbf{p}/\sqrt{n^2 - p^2}\end{pmatrix}$$
$$= \begin{pmatrix}\mathbf{p}\\ \mathbf{q} + (z/n)\mathbf{p} + (z/2n^3)p^2\mathbf{p} + \mathcal{O}(p^5)\end{pmatrix}. \quad (2.7)$$

The coefficient $z/2n^3$ of $p^2\mathbf{p}$ represents the third-order *spherical aberration* of simple propagation.

To treat *refracting surfaces*, a transformation operator $\mathbb{G}_L = \exp\widehat{L}$ exists along a one-parameter subgroup line $\mathbb{G}_L(z) = \exp(z\widehat{L})$ of direction $\widehat{L}$, so that $\mathbb{G}_L(0) =$ identity and $\mathbb{G}_L(1) = \mathbb{G}_L$, generating the surface canonical transformation. Although Lie canonical transformations of the type (2.7) and finite canonical transformations through "cross-variable" generating functions (Ref. 7, p. 240) are equivalent,[9,10] the connection is not direct. Reference 11 relates both to the eikonal equation and Feynman path integrals. To treat refracting surfaces here, we showed in Ref. 4 that one may characterize the effect of a refracting surface $z = \zeta(\mathbf{q})$ between two media n and n' on phase space $\mathbb{S}_\zeta : (\mathbf{p},\mathbf{q}) \mapsto (\mathbf{p}',\mathbf{q}')$, using simple geometry and Snell's law to write the pair of implicit equations

$$\bar{\mathbf{q}} := \mathbf{q} + \zeta(\bar{\mathbf{q}})\mathbf{p}/\sqrt{n^2 - p^2} = \mathbf{q}' + \zeta(\bar{\mathbf{q}})\mathbf{p}'/\sqrt{n'^2 - p'^2}, \quad (2.8a)$$

$$\bar{\mathbf{p}} := \mathbf{p} + \sqrt{n^2 - p^2}(\nabla\zeta)(\bar{\mathbf{q}}) = \mathbf{p}' + \sqrt{n'^2 - p'^2}(\nabla\zeta)(\bar{\mathbf{q}}). \quad (2.8b)$$

This factorizes the surface transformation $\mathbb{S}_\zeta$ into two "root" transformations $\mathbb{R}_{\zeta,n}$ given through the defining equality in (2.8) as $\mathbb{S}_\zeta = \mathbb{R}_{\zeta,n}\mathbb{R}_{\zeta,n'}^{-1}$. It is proven there[4] that $\mathbb{R}_{\zeta,n}$ is *canonical* at all its nonsingular points when the two media are homogeneous; singularity lines of the mapping must exist but may be kept far from the phase-space origin through judicious placing of the optical center of the surface and pupil collimation. Moreover, although the transformation $\mathbb{R}_{\zeta,n}$ is *implicitly* defined, self-replacement of $\bar{\mathbf{q}}$ into the right-hand-side members renders the transformation to any aberration order in homogeneous polynomial expansion. In Ref. 4 we reproduced Dragt's results[1] on the third-order aberration treatment of quartic axially symmetric aligned surfaces:

$$\zeta(\mathbf{q}) = \alpha q^2 + \beta(q^2)^2. \quad (2.9)$$

Iterating (2.8a) twice, replacing into (2.8b) and iterating three times, we find[4]

$$\mathbb{R}_{\zeta,n}\mathbf{q} = \bar{\mathbf{q}} = \mathbf{q} + (\alpha/n)q^2\mathbf{p} + \mathcal{O}((p,q)^5), \quad (2.10a)$$

$$\mathbb{R}_{\zeta,n}\mathbf{p} = \bar{\mathbf{p}} = \mathbf{p} + 2n\alpha\mathbf{q} - (\alpha/n)p^2\mathbf{q} + 2\alpha^2 q^2\mathbf{p}$$
$$+ 4n\beta q^2\mathbf{q} + \mathcal{O}((p,q)^5). \quad (2.10b)$$

We also find $\mathbb{R}_{\zeta,n'}^{-1}$ in the same way and concatenate with (2.10) to form the *surface transformation* given by

$$\mathbf{S}_\zeta \mathbf{q} = \mathbf{q}' = \mathbf{q} + \alpha\left(\frac{1}{n} - \frac{1}{n'}\right) q^2\mathbf{p} + 2\alpha^2\left(1 - \frac{n}{n'}\right) q^2\mathbf{q} + \mathscr{O}((p,q)^5), \tag{2.11a}$$

$$\mathbf{S}_\zeta \mathbf{p} = \mathbf{p}' = \mathbf{p} + 2\alpha(n-n')\mathbf{q} - \alpha\left(\frac{1}{n} - \frac{1}{n'}\right)p^2\mathbf{q} + 2\alpha\left(1 - \frac{n'}{n}\right)q^2\mathbf{p} - 4\alpha^2\left(1-\frac{n}{n'}\right)p\cdot q\mathbf{q} + 4(n-n')\left(\beta + \alpha^3\frac{n-n'}{n'}\right)q^2\mathbf{q} + \mathscr{O}((p,q)^5). \tag{2.11b}$$

In (2.7) and (2.11) there is a linear part and we have kept terms to third order, which are nonlinear (cubic) transformations of phase space, i.e., *aberrations*. If we test for canonicity in the truncated third-order expressions, we find that, in general, they are canonical through third order only. (However, keeping only *linear* terms, they are exactly canonical.)

It is important that the surface be *aligned*, i.e., that it be perpendicular to the optical axis of the system at its optical center, $\nabla\zeta(\mathbf{0}) = \mathbf{0}$, otherwise, *inhomogeneous* terms will appear as $\mathbf{p}' = \mathbf{p}_0 + \mathbf{p} + \cdots$, with constant $\mathbf{p}_0$. These will in turn compose with free propagation to shift the origin of phase space. In Ref. 12 we considered the quantum-mechanical counterpart of system with prism and anisotropy components, applicable only to Gaussian systems. If aberrations are to be present, however, misalignment leads these to compose to Gaussian correction terms, which at present we cannot treat coherently. We explicitly reduce ourselves here to free propagation and surfaces whose general form is (2.9).

III. THE THIRD-ORDER ABERRATION ALGEBRA a^3

Given the set of all monomials $\mathscr{L}^N = p_1^{n_1}p_2^{n_2}q_1^{m_1}q_2^{m_2}$ of degree $N = n_1 + n_2 + m_1 + m_2$ and their associated operators, we have a Lie algebra under the Poisson and commutator bracket. This Lie algebra a^∞ is infinite dimensional since N is not bounded. Two monomials $\mathscr{L}^{N_1}$ and $\mathscr{L}^{N_2}$ in a Poisson bracket yield monomials $\mathscr{L}^N$ of degree $N = N_1 + N_2 - 2$. Second-degree monomials $\mathscr{L}^2$ thus close into a six-dimensional algebra, which is sp$(4,R)$, while $N_1,N_2>4 \Rightarrow N>6$. If we intend to keep computations up to a given order N_a, the aberration order, we may build the *quotient* algebra of a^∞ modulo the (nilpotent) subalgebra generated by the $\mathscr{L}^N$, with $N > N_a + 2$. The N_ath-*order aberration* algebra a^{N_a} thus defined will be the algebra generated by all monomials $\mathscr{L}^2, \mathscr{L}^3, \ldots, \mathscr{L}^{N_a+1}$, with a Lie bracket that is zero if its order would exceed $N_a + 1$. We use the same Lie bracket symbol for the quotient algebra. (A *contraction* from a^∞ leads to the same object and is equivalent.)

If the system is axially symmetric around the optical axis, and also symmetric under reflections ρ across a plane containing the optical axis, then $\widehat{M} := (q_1p_2 - q_2p_1)\widehat{\ }$ (which generates parallel rotations in the $\mathbf{q}$ and $\mathbf{p}$ planes) and ρ are *symmetry* generators for the system, and all the latter's (*dynamical*) algebra generators must commute with them. The generators of a^{N_a} may thus only contain functions of p^2, $p\cdot q$, and q^2. To third order, we choose and label the following basis for a^3:

$$^2\chi_1^1 := p^2, \quad ^2\chi_0^1 := p\cdot q, \quad ^2\chi_{-1}^1 := q^2; \tag{3.1a}$$

$$^4\chi_2^2 := (p^2)^2, \quad ^4\chi_1^2 := p^2 p\cdot q, \quad ^4\chi_0^2 := \tfrac{1}{3}[p^2q^2 + 2(p\cdot q)^2], \quad ^4\chi_{-1}^2 := p\cdot q q^2, \quad ^4\chi_{-2}^2 := (q^2)^2; \tag{3.1b}$$

$$^4\chi_0^0 := \tfrac{1}{2}[p^2q^2 - (p\cdot q)^2]. \tag{3.1c}$$

(The *degree n* in $^n\chi_m^j$ need not be written in this article, since there will be no degeneracy in the group algebra, as there would be between $^4\chi_0^0$ and $^0\chi_0^0 = 1$.)

We first note that the three χ^1's form an algebra by themselves:

$$\{\chi_0^1, \chi_{\pm 1}^1\} = \pm 2\chi_{\pm 1}^1, \quad \{\chi_{+1}^1, \chi_{-1}^1\} = -4\chi_0^1. \tag{3.2}$$

This algebra is denoted by sp$(2,R)$, and we shall call it the *Gaussian* algebra γ. The operators $^n\hat{\chi}_m^j$ correspond to functions $^n\chi_m^j(\mathbf{p},\mathbf{q})$ of *degree n*, of "*spin*" j, and *weight m* under the Gaussian algebra. The weight m of χ_m^j is defined through the Lie bracket with the *weight operator* $\hat{\chi}_0^1$ as its eigenvalue:

$$\hat{\chi}_0^1\,\chi_m^j = \{\chi_0^1, \chi_m^j\} = 2m\chi_m^j \quad (m = j, j-1, \ldots, -j). \tag{3.3a}$$

The *raising and lowering operators* in γ are $\hat{\chi}_{\pm 1}^1$, with coefficients

$$\hat{\chi}_{\pm 1}^1\,\chi_m^j = \{\chi_{\pm 1}^1, \chi_m^j\} = 2(m \mp j)\chi_{m\pm 1}^j. \tag{3.3b}$$

The "*spin*" j is the maximum eigenvalue of $\hat{\chi}_0^1$ for the *multiplet* obtained by applying the raising operator repeatedly. Finally, the classical Casimir operator of the algebra γ is $(\widehat{M})^2$,

$$2\chi_0^0 = \chi_1^1\,\chi_{-1}^1 - (\chi_0^1)^2 = (q_1p_2 - q_2p_1)^2 = M^2, \tag{3.4}$$

and corresponds to the Petzval invariant.

In our case (3.3) contains (3.2) for $j = 1$, and is easily verified using (3.1) for $j = 2$, corresponding to aberration order 3. For N_ath aberration order in axially symmetric aligned systems (thus N_a odd), (3.3) extends to $j = (N_a + 1)/2$.

The *third-order* (axis-symmetric) *aberration algebra* (in two-dimensional systems), a^3, is generated by the nine functions (3.1) under a Poisson bracket [(3.2) and (3.3)]. The Lie bracket between two $^4\chi$'s is zero:

$$\{^4\chi_m^j, {}^4\chi_{m'}^{j'}\} = 0, \quad j,j' = 0,2, \tag{3.5}$$

since a^3 is a *quotient* algebra and any higher $^n\chi$, $n > 4$, is congruent with nought. The six generators $^4\chi_m^j$ thus constitute an Abelian "pure-aberration" subalgebra v. The Lie brackets (3.3) characterize a^3 as a *semidirect sum* of γ and v, $a^3 = v \oplus\!\!\!\!\supset \gamma$ (so that v is an Abelian normal ideal in a^3). Furthermore, of the six possible fourth-order homogeneous functions, $^4\chi_0^0$ is an *invariant* (i.e., a *singlet*) under γ. The other five $^4\chi^2$'s form a *quintuplet* under γ.

For higher aberration and homogeneity orders $n = 4,6,\ldots, N_a + 1$ the corresponding decomposition of the aberration ideal will be into multiplets $^n\chi^j$ with $j = n/2, n/2-2, \ldots, 1$ or 0. Then, the zero in (3.5) for $\{^4\chi, {}^4\chi\}$ will be replaced by ${}^6\chi$'s representing the compounding of aberrations to fifth order. The pure-aberration subalgebra will be, in general, nilpotent.

We are using the language of quantum-mechanical angular momentum theory, even though the sp(2,R) Lie bracket (3.2) differs from the true angular momentum algebra su(2) through the *minus sign* of $\{\chi^1_{+1},\chi^1_{-1}\}$ [propagating through the signs of (3.3b) and (3.4)]: our χ^j's transform as *finite-dimensional non-Hermitian* irreducible representations of the symplectic (and not rotation) algebras. In this article we need not be overly concerned with this fact, but we cannot, for example, introduce a conserved inner product leading to a standard normalization. The coefficients used in (3.1) are compatible with (3.3) and are chosen so as to make the best use of integers for easy manipulation.

IV. THE THIRD-ORDER ABERRATION GROUP A^3

We now exponentiate the nine-parameter algebra a^3 to its connected Lie group A^3. This will inherit the structure $v \oplus \gamma$ as a semidirect *product* $\Upsilon \otimes \Gamma$, Γ being the Gaussian linear canonical transformation group of phase space, Sp(2,R), generated by γ and Υ the $(5+1)$-dimensional Abelian group generated by v, leading to nonlinear aberrations.

We shall parametrize the A^3 group elements as $G\{\mathbf{v},w;\mathbf{M}\}$, $\mathbf{v}$ being a row vector with five components $\{v_m\}^2_{m=-2}$, w a scalar, and $\mathbf{M}$ a 2×2 unimodular (symplectic) matrix with three independent parameters $\{u_m\}^1_{m=-1}$. We do this by defining

$$G\{0,0;\mathbf{M}(\mathbf{u})\} := \exp(\mathbf{u}\cdot\hat{\chi}^1) = \exp((u_1 p^2 + u_0 p\cdot q + u_{-1}q^2)\hat{\ }), \quad (4.1a)$$

$$\mathbf{M}(\mathbf{u}) = \begin{pmatrix} \cosh\omega - u_0\omega^{-1}\sinh\omega & -2u_{-1}\omega^{-1}\sinh\omega \\ 2u_1\omega^{-1}\sinh\omega & \cosh\omega + u_0\omega^{-1}\sinh\omega \end{pmatrix}, \quad (4.1b)$$

$$\omega := \pm\sqrt{u_0^2 - 4u_1u_{-1}}, \quad (4.1c)$$

for the Gaussian part, and

$$G\{\mathbf{v},w;\mathbf{1}\} := \exp\left(\sum_{m=-2}^{2} v_m\hat{\chi}^2_m + w\hat{\chi}^0_0\right), \quad (4.2)$$

for the aberration part. Finally, we define

$$G\{\mathbf{v},w;\mathbf{M}\} := G\{\mathbf{v},w;\mathbf{1}\}G\{0,0;\mathbf{M}\}. \quad (4.3)$$

The subset of elements (4.1) forms the Sp(2,R) subgroup within A^3. [The way to find $\mathbf{M}(\mathbf{u})$ is through transforming the generic orbit in the algebra to $\omega\hat{\chi}^1_0$ in the coherent-state basis by Bargmann's transform and using one of the standard Pauli matrix representations of Sp(2,R).] The composition law is simply matrix multiplication. The set of elements (4.2) is the six-dimensional translation group $\Upsilon = T_5 \otimes T_1 = T_6$.

In order to give the general product of two A^3 elements we need the adjoint action of Γ on its $(5+1)$-dimensional ideal in the exponent of (4.2). This is, we noted, an sp(2,R) quintuplet and a singlet. Therefore (and independently of the degree n),

$$G\{0,0;\mathbf{M}\}\chi^j_m = \sum_{m'=-j}^{j} D^j_{mm'}(\mathbf{M}^{-1})\chi^j_{m'}, \quad (4.4)$$

where $\mathbf{D}^3(\mathbf{M})$ is a $(2j+1)$-dimensional *representation* of Sp(2,R) [i.e., a set of matrices following group composition: $\mathbf{D}^j(\mathbf{M})\mathbf{D}^j(\mathbf{N}) = \mathbf{D}^j(\mathbf{MN})$]. The composition law in A^3 can thus be ascertained to be

$$G\{\mathbf{v}_1,w_1;\mathbf{M}_1\}G\{\mathbf{v}_2,w_2;\mathbf{M}_2\} = G\{\mathbf{v}_1 + \mathbf{v}_2\mathbf{D}^2(\mathbf{M}_1^{-1}), w_1 + w_2; \mathbf{M}_1\mathbf{M}_2\}. \quad (4.5a)$$

The unit element is $G\{0,0;\mathbf{1}\}$ and the inverse is given by

$$G\{\mathbf{v},w;\mathbf{M}\}^{-1} = G\{-\mathbf{v}\mathbf{D}^2(\mathbf{M}), -w;\mathbf{M}^{-1}\}. \quad (4.5b)$$

For our $j=2$ case, we give the 5×5 matrix explicitly:

$$\mathbf{D}^2\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \begin{pmatrix} \alpha^4 & 4\alpha^3\beta & 6\alpha^2\beta^2 & 4\alpha\beta^3 & \beta^4 \\ \alpha^3\gamma & \alpha^2(\alpha\delta+3\beta\gamma) & 3\alpha\beta(\alpha\delta+\beta\gamma) & \beta^2(3\alpha\delta+\beta\gamma) & \beta^3\delta \\ \alpha^2\gamma^2 & 2\alpha\gamma(\alpha\delta+\beta\gamma) & \alpha^2\delta^2+4\alpha\beta\gamma\delta+\beta^2\gamma^2 & 2\beta\delta(\alpha\delta+\beta\gamma) & \beta^2\delta^2 \\ \alpha\gamma^3 & \gamma^2(3\alpha\delta+\beta\gamma) & 3\gamma\delta(\alpha\delta+\beta\gamma) & \delta^2(\alpha\delta+3\beta\gamma) & \beta\delta^3 \\ \gamma^4 & 4\gamma^3\delta & 6\gamma^2\delta^2 & 4\gamma\delta^3 & \delta^4 \end{pmatrix}. \quad (4.6)$$

For general integer (or half-integer)j, this is

$$D^j_{mm'}\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \sum_n \binom{j-m}{j+m'-n}\binom{j+m}{n}\alpha^n\beta^{j+m-n}\gamma^{j+m'-n}\delta^{n-m-m'}. \quad (4.7)$$

In particular, $\mathbf{D}^j(\mathbf{1}) = \mathbf{1}$, $D^0(\mathbf{M}) = 1$, and $\mathbf{D}^{1/2}(\mathbf{M}) = \mathbf{M}$. This matrix representation is, of course, not unitary. [The unitary representations of Sp(2,R) are infinite dimensional. In "continuous" bases, the group Sp(2,R) may be represented as a group of integral-canonical transforms describing Gaussian wave optics.]

V. NONLINEAR ACTION ON PHASE SPACE

Our aberration group A^3 acts as linear and cubic transformations of phase space, as we shall presently see.

When acting with the Gaussian subgroup Γ through (4.1a) on the coordinates of phase space $\mathbf{p},\mathbf{q}$, all terms in the exponential series of multiple Poisson brackets between the

χ^1_m return linear functions of $\mathbf{p},\mathbf{q}$, so the series must (and may) be summed. When acting with the "pure aberration" subgroup Υ, the first Poisson bracket term yields homogeneous functions of $(\mathbf{p},\mathbf{q})$ of degree 3, the second of order 5, etc. Moreover, since all transformation generators are invariant around the optical axis rotations (in the $\mathbf{q}$ plane and, correspondingly, in the $\mathbf{p}$ plane), there is no difference in the way the two components transform. Hence, the vector $\mathbf{q}$ will be transformed into a linear combination of $f(\mathbf{p},\mathbf{q})\mathbf{q}$ and $g(\mathbf{p},\mathbf{q})\mathbf{p}$, f and g being functions of the χ^1_m, which are invariant under rotations. Similarly for $\mathbf{p}$. We are thus led to define the *vector* functions

$$^1\chi^{1/2}_{1/2} := \mathbf{p}, \quad ^1\chi^{1/2}_{1/2} := \mathbf{q}; \tag{5.1a}$$

$$^3\chi^{3/2}_{3/2} := p^2\mathbf{p}, \quad ^3\chi^{3/2}_{1/2} := \tfrac{1}{3}(2p\cdot q\mathbf{p} + p^2\mathbf{q}),$$
$$^3\chi^{3/2}_{-1/2} := \tfrac{1}{3}(q^2\mathbf{p} + 2p\cdot q\mathbf{q}), \quad ^3\chi^{3/2}_{-3/2} = q^2\mathbf{q}; \tag{5.1b}$$

$$^3\chi^{1/2}_{1/2} := \tfrac{1}{2}(-p\cdot q\mathbf{p} + p^2\mathbf{q}),$$
$$^3\chi^{1/2}_{-1/2} := \tfrac{1}{2}(-q^2\mathbf{p} + p\cdot q\mathbf{q}). \tag{5.1c}$$

These we may obtain for general half-integer j by setting $^{2j}\chi^j_j := (p^2)^{j-1/2}\mathbf{p}$ and using $\hat{\chi}^1_{-1}$ to lower m down to $-j$. Now, already $^{2j}\chi^j_{j-1}$ will contain two monomials in a fixed linear combination. We then take an arbitrary linear combination of these monomials and ask that the raising operator annihilate this combination; this leads to a relation between the two monomial coefficients, which yields $^{2j}\chi^{j-1}_{j-1}$. Normalization in (5.1) was chosen for convenience in the expressions. This process leads down to $^{2j}\chi^{1/2}_{1/2}$. For (5.1b) which has $j=\frac{3}{2}$, we do this once to obtain $j=\frac{1}{2}$. Of course, there are other monomials of third order in $\mathscr{L}^3$, but only our six—four plus two in our decomposition—are obtained by acting with an axis-symmetric optical system on the general ray $\mathbf{p},\mathbf{q}$. The "initial conditions" (i.e., points in four-dimensional phase space) need not, of course, have the symmetry of the lens system.

We note that we never need the Poisson bracket of two half-integer χ^j's: they are elements of the *homogeneous space* for the algebra and group action, *not* elements of any group we wish to consider. The difference must be borne in mind when building aberration algebras and groups for non-axially-symmetric optical elements and for higher-order aberrations.

The validity of (3.3) for the half-integer-j two-vectors given in (5.1) assures us that under Γ (and independently of n),

$$G\{0,0;\mathbf{M}\}\chi^j_m = \sum_{m'=-j}^{j} D^j_{mm'}(\mathbf{M}^{-1})\chi^j_{m'}. \tag{5.2}$$

To third aberration order, it fits our needs to write this as

$$G\{0,0;\mathbf{M}\}\begin{pmatrix} ^1\chi^{1/2} \\ ^3\chi^{3/2} \\ ^2\chi^{1/2}\end{pmatrix} = \begin{pmatrix} \mathbf{M}^{-1} & 0 & 0 \\ 0 & \mathbf{D}^{3/2}(\mathbf{M}) & 0 \\ 0 & 0 & \mathbf{M}^{-1}\end{pmatrix}\begin{pmatrix} ^1\chi^{1/2} \\ ^3\chi^{3/2} \\ ^3\chi^{1/2}\end{pmatrix}. \tag{5.3}$$

This $2+4+2=$ eight-dimensional space of linear plus cubic (vector) functions of phase space thus transforms in block-diagonal form as a direct sum of two symplectic doublets ($^1\chi^{1/2}$ and $^3\chi^{1/2}$) and a quadruplet ($^3\chi^{3/2}$).

One last construction is essential if we are to have a consistent *representation* for the full third-order aberration group on phase space. We extend the quotient construction to the latter modulo all $^n\chi$, $n=5,7,\ldots$. In other words, all half-integer monomials *but* those in (5.1) are put congruent to nought. This leaves the eight-dimensional (vector) components (5.3) as our *full* representation space. The action of the pure aberration part of on this is the block-triangular matrix

$$G\{\mathbf{v},w;\mathbf{1}\}\begin{pmatrix} ^1\chi^{1/2} \\ ^3\chi^{3/2} \\ ^3\chi^{1/2}\end{pmatrix} = \begin{pmatrix} \mathbf{1} & \begin{pmatrix} v_1 & 2v_0 & 3v_{-1} & 4v_{-2} \\ -4v_2 & -3v_1 & -2v_0 & -v_{-1}\end{pmatrix} & 2w\mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}\end{pmatrix} \times \begin{pmatrix} ^1\chi^{1/2} \\ ^3\chi^{3/2} \\ ^3\chi^{1/2}\end{pmatrix}, \tag{5.4}$$

with the blocks arranged as before.

The action of the general aberration group element (4.3) is thus obtained acting *first* with $G\{0,0;\mathbf{M}\}$ to produce (5.3), and *second* with $G\{\mathbf{v},w;\mathbf{1}\}$, which, being a differential operator in $\mathbf{p},\mathbf{q}$, acts on the χ's jumping over the scalar coefficients of the $\mathbf{M}$-block matrix (5.3), puttng the 8×8 matrix in (5.4) to its *right*. The write-out of the first two rows has the form

$$G\{\mathbf{v},w;\mathbf{M}\}\,^1\chi^{1/2} = \mathbf{M}^{-1}\,^1\chi^{1/2} + \mathbf{M}^{-1}\mathbf{V}\,^3\chi^{3/2} + 2w\mathbf{M}^{-1}\,^3\chi^{1/2}, \tag{5.5a}$$

$$G\{\mathbf{v},w;\mathbf{M}\}\begin{pmatrix}\mathbf{p} \\ \mathbf{q}\end{pmatrix} = \mathbf{M}^{-1}\left[\begin{pmatrix}\mathbf{p} \\ \mathbf{q}\end{pmatrix} + \begin{pmatrix} v_1 & 2v_0 & 3v_{-1} & 4v_{-2} \\ -4v_2 & -3v_1 & -2v_0 & -v_{-1}\end{pmatrix} \times \begin{pmatrix} ^3\chi^{3/2}_{3/2} \\ ^3\chi^{3/2}_{1/2} \\ ^3\chi^{3/2}_{-1/2} \\ ^3\chi^{3/2}_{-3/2}\end{pmatrix} + 2w\begin{pmatrix} ^3\chi^{1/2}_{1/2} \\ ^3\chi^{1/2}_{-1/2}\end{pmatrix}\right]. \tag{5.5b}$$

This is the general A^3 action on phase-space, modulo terms of order 5. [In (5.5a), $\mathbf{V}$ stands for the 2×4 matrix in (5.5b), which will not be needed further here; higher aberration orders will require variously dimensioned matrices of this type.] The last six rows of (5.3) and (5.4) are not independently needed, so we shall only use $\mathbf{D}^2(\mathbf{M})$ denoting it by $\mathbf{D}(\mathbf{M})$.

Of the six independent simple monomials in $^3\chi$, three are directed along $\mathbf{p}$ and three along $\mathbf{q}$. The distinguishing feature of our symplectic-adapted basis is that we are thereby reducing the Cartesian 6×6 matrix action in (5.3) to a 4×4 plus a 2×2 action given by (5.5). For *meridional* rays (i.e., rays that lie on a plane with the optical axis), $\mathbf{p}$ and $\mathbf{q}$ are collinear and ${}^3\chi^{1/2}=\mathbf{0}$, and the symplectic quadruplet gives all image position aberrations ($\mathbf{q}'$ does not depend on v_{-2}) or image momentum aberrations ($\mathbf{p}'$ does not depend on v_2). For *skew* rays, where $\mathbf{p}$ and $\mathbf{q}$ make an angle δ in a $z=$ const plane, $\chi^{3/2}_{3/2}$ and $\chi^{3/2}_{-3/2}$ still lie in the directions of $\mathbf{p}$ and $\mathbf{q}$, respectively; the others are linear combinations and distribute in the following way: $\chi^{3/2}_{1/2}$ makes an angle ϵ with the former equal to the angle between $\chi^{3/2}_{-1/2}$ and the latter, and $\tan\epsilon=\frac{1}{3}\tan\delta$, so the quadruplet vectors lie between $\mathbf{p}$ and $\mathbf{q}$. For the ${}^3\chi^{1/2}$ doublet, ${}^3\chi^{1/2}_{1/2}$ is perpendicular to $\mathbf{p}$ and ${}^3\chi^{1/2}_{-1/2}$ to $\mathbf{q}$.

VI. LENS SYSTEM ELEMENTS

We now examine the group elements in A^3 corresponding to the contituents of the most general system. These consist of homogeneous slabs where free propagation (2.7) occurs, separated by refracting surfaces where (2.11) applies.

Free propagation is a z-parametrized subgroup generated by the Hamiltonian (2.1) kept to three terms. We may use then (4.1) and (4.2) to find $\exp(-z\hat{H})$ due to the circumstance that $\hat{\chi}^1_1$ and $\hat{\chi}^2_2$ commute (for optical fibers we need the exponential of the general algebra element, as we shall show in part II of this series of articles). We thus have *free propagation* over length z:

$$\mathbb{G}(z)=G\left\{\left(-\frac{z}{8n^3},0,0,0,0\right),0;\begin{pmatrix}1 & 0\\ -z/n & 1\end{pmatrix}\right\}. \quad (6.1)$$

These group elements produce the geometric result (2.7) on phase space through (5.5), and could alternatively be obtained from the phase-space transformation, since (5.5) contains already *all* the group parameters of A^3. It is the latter method we must use for refracting surfaces.

Writing out (2.10) so as to fit it to the form (5.5) we obtain the "*root*" *surface transformation* for (2.9)

$$\mathbb{R}_{\zeta,n}=G\left\{\left(0,0,-\frac{\alpha}{2n},0,n\beta\right),-\frac{2\alpha}{3n};\begin{pmatrix}1 & -2n\alpha\\ 0 & 1\end{pmatrix}\right\}. \quad (6.2)$$

The concatenation $\mathbb{R}_{\zeta,n}\mathbb{R}^{-1}_{\zeta,n'}$ is, from (4.4) and (4.5),

$$\begin{aligned}\mathbb{S}_\zeta&=\mathbb{R}_{\zeta,n}\mathbb{R}^{-1}_{\zeta,n'}=G\{\mathbf{r},\rho;\mathbf{M}\}G\{\mathbf{r}',\rho';\mathbf{M}'\}^{-1}\\&=G\{\mathbf{r},\rho,\mathbf{M}\}G\{-\mathbf{r}'\mathbf{D}(\mathbf{M}'),-\rho';\mathbf{M}'^{-1}\}\\&=G\{\mathbf{r}-\mathbf{r}'\mathbf{D}([\mathbf{M}\mathbf{M}'^{-1}]^{-1}),\rho-\rho';\mathbf{M}\mathbf{M}'^{-1}\}.\end{aligned} \quad (6.3a)$$

The 2×2 matrix $\mathbf{M}\mathbf{M}'^{-1}$ is upper triangular, with 1-2 element $P=-2(n-n')\alpha$ giving the *Gaussian power* of the refracting surface. The aberration singlet coefficient w simply subtracts, and the aberration quintuplet is given by the vector

$$\begin{aligned}&\mathbf{r}-\mathbf{r}'\mathbf{D}([\mathbf{M}\mathbf{M}'^{-1}]^{-1})\\&\quad=(0,0,-\alpha/2n,0,n\beta)-(0,0,-\alpha/2n',0,n'\beta)\\&\quad\times\begin{pmatrix}\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot&\cdot&\cdot\\ 0&0&1&-2P&P^2\\ \cdot&\cdot&\cdot&\cdot&\cdot\\ 0&0&0&0&1\end{pmatrix}.\end{aligned} \quad (6.3b)$$

In the $\mathbf{M}$-matrix we have replaced by dots all elements that are irrelevant due to the zeros in the components of $\mathbf{r}'$.

In this way [or by fitting (2.11) to the form (5.5)] we find the *refracting surface transformation* for (2.9):

$$\begin{aligned}\mathbb{S}_\zeta=G\Bigg\{\Bigg(&0,0,\frac{\alpha}{2}\left[\frac{1}{n'}-\frac{1}{n}\right],\frac{2\alpha^2}{n'}[n-n'],\\&\frac{2\alpha^3}{n'}[n-n']^2+\beta[n-n']\Bigg),\\&\frac{2\alpha}{3}\left[\frac{1}{n'}-\frac{1}{n}\right];\begin{pmatrix}1 & -2\alpha[n-n']\\ 0 & 1\end{pmatrix}\Bigg\}.\end{aligned} \quad (6.4)$$

The geometrical meaning of each of the aberration coefficients will be seen in the next section. Here we wish to remark that the *order* of aberrations (left) and Gaussian transformations (right) in $G\{\mathbf{v},w;\mathbf{M}\}$ means that the aberration matrix (5.4) acts *first*, aberrating the object coordinates, and the Gaussian part of the group element acts *second* producing a focused (if possible), magnified image. The relation of aberration size/object size is thus not changed by the Gaussian transformation. [One may transform to the opposite order,

$$\widetilde{G}\{\mathbf{M};\mathbf{v},w\}:=G\{\mathbf{0},0;\mathbf{M}\}\times G\{\mathbf{v},w;\mathbf{1}\}=G\{\mathbf{v}\mathbf{D}(\mathbf{M}^{-1}),w;\mathbf{M}\}$$

to match the results of Dragt,[1] as was done in Ref. 4.]

We point out that in parametrizing the aberration group, we may have the linear combination

$$v_0\chi^2_0+w\chi^0_0=Cp^2q^2+A(p\cdot q)^2, \quad (6.5a)$$

$$\begin{pmatrix}C\\ A\end{pmatrix}=\begin{pmatrix}\frac{1}{3} & \frac{1}{2}\\ \frac{2}{3} & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}v_0\\ w\end{pmatrix},\quad \begin{pmatrix}v_0\\ w\end{pmatrix}=\begin{pmatrix}1 & 1\\ \frac{4}{3} & -\frac{2}{3}\end{pmatrix}\begin{pmatrix}C\\ A\end{pmatrix}. \quad (6.5b)$$

Since the surface transformation (6.4) has a fixed relation $4v_0=3w$, this implies that $A=0$ in (6.5) and $C=\frac{1}{2}\alpha\times(1/n'-1/n)$. The term with coefficient C effects $\mathbf{p}\mapsto 2p^2\mathbf{q}$, $\mathbf{q}\mapsto -2q^2\mathbf{p}$, i.e., crosses the $\mathbf{p}$ and $\mathbf{q}$ directions. The term absent from the surface transformation, $A(p\cdot q)^2$, preserves the directions of $\mathbf{p}$ and $\mathbf{q}$. These effects make C correspond to the *curvature of field* parameter and A to the *astigmatism* parameter. Under Gaussian transformations these two aberrations mix, whereas the singlet aberration generated by ${}^4\chi^0_0$ stays *separate* from the rest. In this $\mathrm{Sp}(2,R)$-classified scheme, 5×5 and 1×1 matrices replace the Cartesian-basis 6×6 ones. Similarly, 2×4 rectangular matrices (5.5) replace the 2×6 ones needed in the latter.

VII. THIRD-ORDER ABERRATIONS OF PHASE SPACE

We shall now examine the effect on (four-dimensional) phase space (at the $z = 0$ plane) of each of the aberration coefficients $(\mathbf{v}, w)$ classified into a quintuplet and a singlet. These will relate to the traditional third-order Seidel aberration coefficients.[1,13,14] We write the χ's out in terms of $\mathbf{p}$ and $\mathbf{q}$ through (5.1).

Spherical aberration ($j=2$, $m=2$):

$$G\{(v_2,0,0,0,0),0;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p}\\ \mathbf{q} - 4v_2 p^2\mathbf{p}\end{pmatrix}. \tag{7.1}$$

This aberration produces no change in the ray direction ($\mathbf{p}' = \mathbf{p}$), but *unfocuses* $\mathbf{q}$, the object position. Rays on a cone (with axis parallel to the optical axis) of half angle θ are made to fall on a circle of radius $4v_2(n \sin\theta)^3$, preserving the direction of $\mathbf{p}$, regardless of the image position in the object plane.

Coma ($j=2$, $m=0$):

$$G\{0,v_1,0,0,0),0;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p} + v_1 p^2\mathbf{p}\\ \mathbf{q} - v_1[2p\cdot q\mathbf{p} + p^2\mathbf{q}]\end{pmatrix}. \tag{7.2}$$

Here the image position aberration vector $\mathbf{q}' - \mathbf{q}$ vanishes on the optical axis $\mathbf{q} = \mathbf{0}$, and for rays parallel to the axis $\mathbf{p} = \mathbf{0}$. Off the axis ($\mathbf{q} \neq \mathbf{0}$), rays on a cone ($|\mathbf{p}|$ constant) issuing from an object point $\mathbf{q}$ fall on a circle with center at $\mathbf{q}_c = \mathbf{q}(1 - 2v_1 p^2)$ and radius $p = v_1 p^2|\mathbf{q}|$. This gives rise to the familiar 60° "comet" image of points. The two meridional rays on the cone fall on the *same* point in the direction of $\mathbf{q}$, and the two sagittal rays (perpendicular to the former) on the diametrically opposed point, also in the direction of $\mathbf{q}$. One turn around the cone maps onto two turns around the image circle. Momentum space $\mathbf{p}$ is also *distorted* [i.e., $\mathbf{p} \mapsto \mathbf{p}(1 + v_1 p^2)$] but *not* unfocused ($\mathbf{p}'$ is independent of $\mathbf{q}$).

"Curvatism" ($j=2$, $m=0$):

$$G\{(0,0,v_0,0,0),0;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p} + v_0[\frac{4}{3}p\cdot q\mathbf{p} + \frac{2}{3}p^2\mathbf{q}]\\ \mathbf{q} - v_0[\frac{2}{3}q^2\mathbf{p} + \frac{4}{3}p\cdot q\mathbf{q}]\end{pmatrix}. \tag{7.3}$$

Here also, the aberration vectors $\mathbf{q}' - \mathbf{q}$ and $\mathbf{p}' - \mathbf{p}$ vanish on the axis and for rays parallel to it. Rays on a cone $|\mathbf{p}|$ issuing from $\mathbf{q}$ fall on an ellipse centered on $\mathbf{q}$ in a one-to-one map. The major half axis of the ellipse is $a = 2v_0 q^2|\mathbf{p}|$ along the direction of $\mathbf{q}$, and the minor axis is $b = a/3$. The momentum aberration vector of curvatism is proportional (by $-\frac{2}{3}$) to the position aberration vector for coma. Meridional rays can be made to fall into focus (to third order) when we free-propagate to a *paraboloid* $z = 2v_0 q^2$, as can be seen applying (6.1) to (8.3). Skew rays cannot. Finally, curvatism is invariant under the Fourier conjugation ($\mathbf{p} \mapsto \mathbf{q}$, $\mathbf{q} \mapsto -\mathbf{p}$) of phase space represented by

$$G\left\{0,0;\begin{pmatrix}0 & -1\\ +1 & 0\end{pmatrix}\right\}.$$

Distortion ($j=2$, $m=-1$):

$$G\{(0,0,0,v_{-1},0),0;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix}$$

$$= \begin{pmatrix}\mathbf{p} + v_{-1}[q^2\mathbf{p} + 2p\cdot q\mathbf{q}]\\ \mathbf{q} - v_{-1}q^2\mathbf{q}\end{pmatrix}. \tag{7.4}$$

This is the Fourier conjugate of coma. Position two-space is distorted (for $v_{-1} > 0$ it is of the "barrel" type), but remains in focus ($\mathbf{q}'$ is independent of $\mathbf{p}$), while in momentum space $\mathbf{p}$ the comatic phenomenon appears.

"Pocus" ($j=2$, $m=-2$):

$$G\{(0,0,0,0,v_{-2}),0;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p} + 4v_{-2}q^2\mathbf{q}\\ \mathbf{q}\end{pmatrix}. \tag{7.5}$$

This is a $\mathbf{p}$-*unfocusing* aberration, Fourier conjugate to spherical aberration. Position space is neither distorted nor unfocused, so this has not been usually counted among the Seidel aberrations of a system. For $v_{-2} > 0$, ray slopes increase cubically with $|\mathbf{q}|$ in the direction of $\mathbf{q}$ and thus depth of field decreases away from the origin. Quartic-surface Schmidt-type correction plate surfaces [Eq. (6.4) with $\alpha = 0$, $\beta \neq 0$] produce pure pocus aberration; preceded by free propagation, this contributes to all quintuplet aberrations.

"Astigmature" ($j=0$, $m=0$):

$$G\{\mathbf{0},w;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p} + w[-p\cdot q\mathbf{p} + p^2\mathbf{q}]\\ \mathbf{q} + w[-q^2\mathbf{p} + p\cdot q\mathbf{q}]\end{pmatrix}. \tag{7.6}$$

This aberration is the symplectic singlet and thus invariant under Gaussian optics; it is present in quadratic refracting surfaces, but has no effect on meridional rays, which always remain in focus. Rays on a cone $|\mathbf{p}|$ issuing from $\mathbf{q}$ will map on a straight segment perpendicular to $\mathbf{q}$, and of half width $wq^2|\mathbf{p}|$.

The more familiar linear combinations of the ${}^4\chi_0^2 - {}^4\chi_0^0$ degeneracy are curvature of field and astigmatism; we obtain them from (6.5).

Curvature:

$$G\{(0,0,C,0,0),\tfrac{4}{3}C;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p} + 2Cp^2\mathbf{q}\\ \mathbf{q} - 2Cq^2\mathbf{p}\end{pmatrix}. \tag{7.7}$$

Positions are unfocused as well as momenta off the phase-space origin. Rays on a cone map onto a circle with center $\mathbf{q}$ and radius $2Cq^2|\mathbf{p}|$ (equal to the major half axis of the curvatism ellipse with the same coefficient). They fall into focus in $\mathbf{q}$ on a paraboloid $z = 2Cq^2$.

Astigmatism:

$$G\{(0,0,A,0,0),-\tfrac{2}{3}A;\mathbf{1}\}\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \begin{pmatrix}\mathbf{p} + 2Ap\cdot q\mathbf{p}\\ \mathbf{q} - 2Ap\cdot q\mathbf{q}\end{pmatrix}. \tag{7.8}$$

Due to the common $p\cdot q$ coefficient, sagittal rays are unaffected but meridional ones are. Rays on a cone map onto straight segments, centered and directed along $\mathbf{q}$, of half length $2Aq^2|\mathbf{p}|$.

VIII. COMPOUND LENS SYSTEMS

To calculate the effect of an optical system on phase space in third aberration order, we follow the usual plot in optics, which makes the z axis extend towards the right. There are three steps to the process: (i) corresponding to each element [free space (6.1) or refracting surface (6.4)] we write its group element $G\{\mathbf{v},v;\mathbf{M}\}$ in the same order (left to right); (ii) the product of these group elements through (4.4) yields the group element that describes the concatenated system; and (iii) the action on phase space is obtained through (5.5).

For the purpose of illustration, consider a system with three elements: two homogeneous media n,n' of lengths l,l', separated by a quartic axisymmetric interface $\zeta(\mathbf{q})$. The group elements corresponding to the optical elements are $\hat{G}(l)$, $\hat{S}_\zeta$, and $\hat{G}(l')$ in (6.1) and (6.4). Their concatenation is found now from (4.4) applied twice (multiplication is associative):

$$G\{\mathbf{f},\phi;\mathbf{F}\}G\{\mathbf{v},w;\mathbf{M}\}G\{\mathbf{f}',\phi';\mathbf{F}'\}$$
$$= G\{\mathbf{f} + \mathbf{v}\mathbf{D}(\mathbf{F}^{-1}) + \mathbf{f}'\mathbf{D}([\mathbf{FM}]^{-1}),\phi + w + \phi';\mathbf{FMF}'\}; \quad (8.1\text{a})$$

$$\mathbf{F} := \begin{pmatrix} 1 & 0 \\ -l/n & 1 \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} 1 & -2\alpha(n-n') \\ 0 & 1 \end{pmatrix},$$
$$\mathbf{F}' := \begin{pmatrix} 1 & 0 \\ -l'/n' & 1 \end{pmatrix}, \quad (8.1\text{b})$$

$$\mathbf{f} = (-l/8n^3,0,0,0,0), \quad \phi = 0,$$
$$\mathbf{f}' = (-l'/8n'^3,0,0,0,0), \quad \phi' = 0, \quad (8.1\text{c})$$
$$\mathbf{v} = (0,0,v_0,v_{-1},v_{-2}), \quad w \text{ given by } (6.4). \quad (8.1\text{d})$$

Regarding the Gaussian part, an optical system is *in focus* if it maps object points onto image points, i.e., $\mathbf{q}\mapsto\mathbf{q}' = \mu\mathbf{q}$, and then it has *magnification* μ. The Gaussian 2×2 matrix for the system should be upper triangular

$$\begin{pmatrix} \mu^{-1} & \varphi^{-1} \\ 0 & \mu \end{pmatrix} = \mathbf{N};$$

points in momentum space are then $\mathbf{p}$-unfocused as $\mathbf{p}\mapsto\mathbf{p}' = \mu^{-1}\mathbf{p} + (1/\varphi)\mathbf{q}$ with φ the focal *length* of the system. For the system's symplectic-quintuplet aberration part, we see from (8.1a) that it is the sum of the first (spherical) aberration $\mathbf{f}$ plus the surface aberration $\mathbf{v}$ acted by the preceding Gaussian free propagation $\mathbf{F}$, plus the last aberration $\mathbf{f}'$ acted upon by the concatenation of all *preceding* Gaussian elements $\mathbf{FM} = \mathbf{NF}'^{-1}$. The symplectic-singlet parts add and only surfaces contribute to it.

For the computational task, we should note that since $\mathbf{D}(\mathbf{F}^{-1})$ is lower-triangular with diagonal 1's, and $\mathbf{v}$ has only three nonzero components, for $\mathbf{v}\mathbf{D}(\mathbf{F}^{-1})$ we need only seven sums and nine products. Any final free-space propagation $\mathbf{f}'$ requires only the first row of $\mathbf{D}(\mathbf{FM})$, since it has only the first nonvanishing element. The concatenation of two general elements $G\{\mathbf{v};w;\mathbf{M}\}$ costs four sums and eight products for the Gaussian part, ten and around 24 of each to build $\mathbf{D}(\mathbf{M})$ through (4.6), and 26 and 25, respectively, to calculate the aberration part. An image point $(\mathbf{p},\mathbf{q})$ costs ten sums and 24 products to find all ${}^3\chi$'s through (5.1); applying the optical system through (5.5) requires 24 of each.

Lens systems also yield to analytic expressions in terms of the constituent parameters if we write out (8.1) fully. The expressions start to be lengthy beyond four elements, however. The symbolic computation algorithms are well defined, however, and involve only polynomial expressions up to a finite order with no truncation needed and essential third-order phase-space deformation kept. For numerical and symbolic computation, the process of creation of elements and their concatenation has been automatized into an interactive language based on FORTRAN and REDUCE.[15]

IX. OUTLOOK

Group theory has been proved a sharp tool for theoretical physics, where it is usually associated with quantum systems. It could have been introduced first for classical mechanics, which it also describes, but for the cost of abstraction. Our eventual aim is to be able to treat wave optics in aberration, while providing competitive tools for geometric optics. So, we have made A^3 into a group-theoretical model for third-order aberrations. Group-theoretical answers to the question of the "most general" third-order aberrating systems include that free propagation and refracting surfaces indeed generate the most general such system. Minimal numbers of these for generic and nongeneric orbits are of interest. There is the "inverse lens" problem: can it be built with *positive* free propagations? The design of corrector plates to globally or selectively eliminate aberrations leads to an optimization problem with five plus one variables.

In the next article we shall detail the consideration of nonhomogeneous optical elements such as media with $\mathbf{q}$-dependent refraction indices $n(\mathbf{q}) = n_0 - vq^2 - \rho q^4$ (Ref. 16, Sec. 6.4), modeling optical fibers, quadratic interfaces between them, and z-dependent versions of them[12] modeling optical funnels, for example.

Aberration orders higher than 3 require groups larger than A^3. These groups include parameters for a symplectic aberration septuplet and a triplet for order 5, a nonuplet, quintuplet, and singlet for order 7, etc. Correspondingly there is an enlarged quotient phase space where the group acts linearly. Fifth aberration order requires a (vector) sextuplet, quadruplet, and doublet, and so on. Non-axially-symmetric systems will require a larger $\mathrm{Sp}(4,R)$ Gaussian group as well as even-ordered aberrations. These are the problems with magnetic lenses.[2]

The question of a wave-optical group of integral transforms describing aberrations is, presently, not purely technical. Gaussian optics (including prisms and birefringent media, and z-inhomogeneous elements) has been found[17] to yield to well-developed group-theoretical representations: Canonical integral transforms (see Refs. 17, 18, and 19, part IV). This extends to results with coherent states, Gaussian beams, and the Wigner distribution function (the radar autoambiguity function). (See Refs. 12 and 20–22.) A similar if approximate aberration group of integral transforms should exhibit the pattern of diffraction in aberration in the kernel, if such can be formalized.

[1]A. J. Dragt, "Lie algebraic theory of geometrical optics and optical aberrations," J. Opt. Soc. Am. 72, 372 (1982).
[2]A. J. Dragt, *Lectures on Nonlinear Orbit Dynamics, AIP Conference Proceedings*, Vol. 87 (AIP, New York, 1982).
[3]A. J. Dragt and J. M. Finn, "Lie series and invariant functions for analytic symplectic maps," J. Math. Phys. 17, 2215 (1976).
[4]M. Navarro-Saad and K. B. Wolf, "Factorization of the phase-space transformation produced by an arbitrary refracting surface," J. Opt. Soc. Am. (to appear).
[5]A. J. Dragt, E. Forest, and K. B. Wolf, "Foundations of a Lie algebraic theory of geometrical optics," in *Lie Methods in Optics, Lecture Notes in Physics* (Springer, Berlin, 1986).
[6]V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U. P., London, 1984).
[7]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1959).
[8]S. Steinberg, "Factored product expansions of solutions of nonlinear dif-

ferential equations," SIAM J. Math. Anal. **15**, 108 (1984).
[9]F. J. Testa, "Finite and infinitesimal canonical transformations. I," J. Math. Phys. **11**, 2698 (1970).
[10]E. J. Testa, "Finite and infinitesimal canonical transformations. II," J. Math. Phys. **14**, 1097 (1973).
[11]F. J. Testa, "Finite and infinitesimal canonical transformations in classical field theory," Physica **69**, 579 (1973).
[12]K. B. Wolf, "On time-dependent quadratic quantum Hamiltonians," SIAM J. Appl. Math. **40**, 419 (1981).
[13]H. A. Buchdahl, *Optical Aberration Coefficients* (Oxford U. P., London, 1954).
[14]F. D. Cruickshank and G. A. Hills, "Use of optical aberration coefficients in optical design," J. Opt. Soc. Am. **50**, 379 (1960).
[15]M. Navarro-Saad, thesis, Facultad de Ciencias, UNAM, 1985. See also: M. Navarro-Saad and K. B. Wolf, "Applications of a factorization theorem for ninth-order aberration optics," J. Symbolic Comput. **1**, 235 (1985).
[16]K. B. Wolf, "Approximate canonical transformations and the treatment of aberrations. I. One-dimensional simple Nth order aberrations in optical systems," Communicaciones Técnicas IIMAS preprint, 352/1983.
[17]M. Nazarathy and J. Shamir, "Fourier optics decribed by operator algebra," J. Opt. Soc. Am. **70**, 150 (1980); "First-order optics—a canonical operator representation: lossless systems," J. Opt. Soc. Am. **72**, 356 (1982).
[18]M. Moshinsky and C. Quesne, "Oscillator systems," in *Proceedings of the 15th Solvay Conference in Physics (1970)* (Gordon and Breach, New York, 1974).
[19]K. B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979).
[20]M. Nazarathy and J. W. Goodman, "Space-time Fourier optics in dispersive media," ISL, Stanford preprint, 1983; "Diffraction transforms in homogeneous birefringent media," ISL, Stanford preprint, 1984.
[21]W. Schempp, "On the role of the Heisenberg group in information theory," preprint, Universität Siegen, 118/1984; "Radar ambiguity functions, nilpotent harmonic analysis, and holomorphic theta series," preprint, Universität Siegen, 117/1984.
[22]G. García Calderón and M. Moshinsky, "Wigner distribution functions and the representation of canonical transformations in quantum mechanics," J. Phys. A **13**, L185 (1980); J. Ojeda Castañeda, "Focus-error operator and related special functions," J. Opt. Soc. Am. **73**, 1042 (1983); M. Nazarathy and J. W. Goodman, "Ambiguity function processors using linear birefringent dispersive materials," ISL, Stanford preprint, 1984.

The group-theoretical treatment of aberrating systems. II. Axis-symmetric inhomogeneous systems and fiber optics in third aberration order

Kurt Bernardo Wolf
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Apdo. Postal 20-726, 01000 México D.F., Mexico

(Received 8 February 1985; accepted for publication 13 November 1985)

The group-theoretical description of geometrical optical systems whose elements are, in general, inhomogeneous but axis symmetric, in third aberration order, is presented. Closed algebraic expressions are given for fiber elements, i.e., media homogeneous under translations along the optical axis, and refracting interfaces between them.

I. INTRODUCTION

In the first article of this series[1] we introduced a nine-parameter group to describe aligned lens systems to third-order aberration. The structure of this group, denoted A^3, was that of the two-dimensional real symplectic group $\Gamma = \mathrm{Sp}(2,R)$ (isomorphic to the group of 2×2 real unimodular matrices) accounting for Gaussian optics, in semidirect product with a six-parameter Abelian translation group, Υ, having for parameters the six third-order aberration coefficients. In this article we want to use this group-theoretical model to describe optical systems with elements that are inhomogeneous slabs of refracting material, which include optical fibers of possibly varying "diameter" and interfaces between such elements, whose only restriction is that they be axis symmetric, i.e., invariant under rotations around a common optical axis. Aligned lens systems are thus a particular case of the systems to be described here. Basically, group theory provides the framework to represent each optical element of a system by a group element, the concatenation of the former corresponding to the product of the latter. The algebra reduces to matrices and vectors, which is computationally economic.

The Lie-algebraic and group-theoretical preliminaries were given in Ref. 1, to which we shall refer as I. For legibility, the main results of I are summarized in Sec. II, and to recall the notation into which we incorporate the more general results offered here. Inhomogeneous axis-symmetric optical media are described by a refraction index n, which may depend on the distance $q = |\mathbf{q}|$ to the optical axis, and/or the distance z along the axis, i.e., $n = n(q,z)$. Section III tackles the general problem: arbitrary (continuous) z dependence leading to differential equations that find the representing line of evolution operators in the group manifold. They are a set of sequential linear differential equations, of which the first is of second order and the rest of first order. The calculation of the Γ and Υ parts is independent.

A solution for the most general case cannot be written in closed form and numerical integration is to be called for. If the system is z independent, the representing line in the group is a one-parameter subgroup and the general solution in A^3 is given algebraically. We give the result relating any Lie algebra element with its group exponential. The derivation uses the coherent-state basis and defines the "logarithm" of any optical system. A model for optical fibers is described in Sec. IV, essentially as harmonic plus quartic oscillators to third order in phase-space transformations; interfaces between fibers are also addressed using the methods developed in I.

Once a third-order optical system is associated to a group element (or line of group elements) in A^3, its action on the whole of optical phase space is determined. Group-theoretical image-formation algorithms are thus quite distinct from ray tracing ones, which follow individual rays. Third aberration order is kept intrinsically throughout. Some remarks regarding directions of research are presented in the concluding section.

II. THE GROUP-THEORETICAL DESCRIPTION OF OPTICS WITH ABERRATION

In the Hamiltonian formulation,[2,3] an optical system maps an "object" phase space $(\mathbf{p},\mathbf{q})$ of light rays on an "image" phase space $(\mathbf{p}',\mathbf{q}')$, the former at a plane $z = z_0$ perpendicular to the optical axis and the latter at a second $z = z_1$ such plane. Each ray has (at a z plane) its *position* two-vector $\mathbf{q} = (q_1,q_2)^T$ and its conjugate *momentum* two-vector $\mathbf{p} = (p_1,p_2)^T$; the latter is the two-vector projection of the ray on the z plane and of magnitude $p = n \sin\theta$, $n(q,z)$ being the refraction index of the medium at the point and θ the angle between the ray and the z axis. (We do not consider here the case of anisotropic media where n may depend on $\mathbf{p}$.) If $n(0,0) = n_0$, we may write $n(\mathbf{q},z) = n_0 - \bar{n}(q,z)$; the optical Hamiltonian[2] is then

$$\begin{aligned} H(\mathbf{p},\mathbf{q};z) &= -\sqrt{n^2 - p^2} \\ &= -n_0 + (1/2n_0)(p^2 + 2n_0\bar{n} - \bar{n}^2) \\ &\quad + (1/8n_0^3)(p^2 + 2n_0\bar{n} - \bar{n}^2)^2 + \cdots. \end{aligned} \tag{2.1}$$

The z evolution of functions $f(\mathbf{p},\mathbf{q};z)$ of phase space is then given by the Hamilton equation of motion

$$\frac{d}{dz} f(\mathbf{p},\mathbf{q};z) = -\{H,f\}(\mathbf{p},\mathbf{q};z) = : -\widehat{H} f(\mathbf{p},\mathbf{q};z), \tag{2.2}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket and $\widehat{H}$ the phase-space operator (linear in ∇_p and ∇_q) associated to $H(\mathbf{p},\mathbf{q};z)$, which is given by the last equality for arbitrary f.

In I we define the following Γ basis of phase-space functions ${}^n\chi^j_m$, of *order n*, *spin j*, and *weight m*:

$${}^2\chi^1_1 = p^2, \quad {}^2\chi^1_0 = p\cdot q, \quad {}^2\chi^1_{-1} = q^2; \tag{2.3a}$$

$${}^4\chi^2_2 = (p^2)^2, \quad {}^4\chi^2_1 = p^2 p\cdot q,$$

$${}^4\chi^2_0 = \tfrac{1}{3}[p^2q^2 + 2(p\cdot q)^2], \tag{2.3b}$$

$$ {}^4\chi^2_{-1} = p\cdot q q^2, \quad {}^4\chi^2_{-2} = (q^2)^2; $$
$$ {}^4\chi^0_0 = \tfrac{1}{2}[p^2q^2 - (p\cdot q)^2]. \tag{2.3c} $$

We noted there that the three quadratic functions ${}^2\chi^1_m$ in (2.3a) close under Poisson bracket into a $\gamma = \mathrm{sp}(2,R)$ Lie algebra generating the $\Gamma = \mathrm{Sp}(2,R)$ group of Gaussian, linear canonical transformations of phase space. The six functions ${}^4\chi$ in (2.3b) and (2.3c) are all the quartic axis-symmetric functions of phase space. They constitute an ideal v under $\mathrm{sp}(2,R)$, i.e., $\{{}^2\chi, {}^4\chi\} = {}^4\chi$; the five ${}^4\chi^2$'s form a symplectic quintuplet and ${}^4\chi^0$ a singlet under Γ. We consider the quotient algebra of all (axis-symmetric) polynomials modulo those of order 5 or larger, and called it a^3. The latter's Lie bracket (denoted with the same symbol $\{\cdot,\cdot\}$ as the Poisson bracket) is given by

$$ \{{}^n\chi^j_m, {}^{n'}\chi^{j'}_{m'}\} = \begin{cases} 2(jm' - j'm)^{n+n'-2}\chi^{j+j'-1}_{m+m'}, \\ \quad j + j' - 1 < 2 \text{ and } n + n' - 2 < 4, \\ 0, \quad \text{otherwise.} \end{cases} \tag{2.4} $$

The structure of the *aberration algebra* a^3 with basis (2.3) is thus that of a semidirect sum between the *Gaussian* algebra γ and a $5 + 1 =$ six-dimensional Abelian *pure-aberration* algebra v with basis ${}^4\chi$, i.e., $a^3 = v \oplus \gamma$. It is not necessary to write here the order n of ${}^n\chi^j_m, j = 0, 1, 2$.

The third-order aberration *group* A^3 was defined in I as the connected nine-parameter Lie group generated by a^3, with the specific parametrization where the elements $G \in A^3$ are given by [(I 4.1)–(I 4.3)]

$$ G\{\mathbf{v}, w; \mathbf{M}(\mathbf{u})\} := \exp\left(\sum_{m=-2}^{2} v_m \hat{\chi}^2_m + w\hat{\chi}^0_0\right)\exp\left(\sum_{m=-1}^{1} u_m \hat{\chi}^1_m\right). \tag{2.5a} $$

Here, all parameters range over the real line; $\mathbf{v} = (v_2 \;\; v_1 \;\; v_0 \;\; v_{-1} \;\; v_{-2})$ is a five-dimensional row vector, w a scalar, and $\mathbf{M}(\mathbf{u}) \in \mathrm{Sp}(2,R)$ is the 2×2 matrix

$$ \mathbf{M}(\mathbf{u}) = \begin{pmatrix} \cosh\omega - u_0\omega^{-1}\sinh\omega & -2u_{-1}\omega^{-1}\sinh\omega \\ 2u_1\omega^{-1}\sinh\omega & \cosh\omega + u_0\omega^{-1}\sinh\omega \end{pmatrix}, \tag{2.5b} $$

$$ \omega := \pm\sqrt{u_0^2 - 4u_1u_{-1}}. \tag{2.5c} $$

The group composition is given by

$$ G\{\mathbf{v}_1, w_1; \mathbf{M}_1\}G\{\mathbf{v}_2, w_2; \mathbf{M}_2\} = G\{\mathbf{v}_1 + \mathbf{v}_2\mathbf{D}^2(\mathbf{M}_1^{-1}), w_1 + w_2; \mathbf{M}_1\mathbf{M}_2\}, \tag{2.6} $$

where $\mathbf{D}^2$ us a 5×5 *representation* of Γ to be detailed below. The A^3 unit element is $G\{0,0;0\}$ and the inverse $G\{\mathbf{v},w;\mathbf{M}\}^{-1} = G\{-\mathbf{v}\mathbf{D}^2(\mathbf{M}), -w; \mathbf{M}^{-1}\}$. The structure of A^3 is thus revealed to be $A^3 = \Upsilon \otimes \Gamma$, where $\Gamma = \mathrm{Sp}(2,R)$ is the subgroup of Gaussian optics, in semidirect product with the "pure-aberration" six-dimensional Abelian (translation) group $\Upsilon = T_5 \otimes T_1$, consisting of a quintuplet $\mathbf{v}$ and a singlet w under the former, as

$$ G\{0,0;\mathbf{M}\}\hat{\chi}^j_m G\{0,0;\mathbf{M}\}^{-1} = \sum_{m'=-j}^{j} D^j_{mm'}(\mathbf{M}^{-1})\hat{\chi}^j_{m'}. \tag{2.7a} $$

The $(2j+1)\times(2j+1)$ representation matrices D^j of $\mathbf{M} \in \mathrm{Sp}(2,R)$ are given by the elements

$$ D^j_{mm'}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \sum_n \binom{j-m}{j+m'-n}\binom{j+m}{n} \times \alpha^n \beta^{j+m'-n}\gamma^{j+m'-n}\delta^{n-m-m'}. \tag{2.7b} $$

These are finite-dimensional (nonunitary) representations of $\mathrm{Sp}(2,R)$, i.e., $\mathbf{D}^j(\mathbf{M}_1)\mathbf{D}^j(\mathbf{M}_2) = \mathbf{D}^j(\mathbf{M}_1\mathbf{M}_2)$ and $\mathbf{D}^j(\mathbf{1}) = \mathbf{1}$. The 5×5 case was written out explicitly in I.

The homogeneous space $\mathscr{X}^3$ for A^3 group action includes the four-dimensional phase space $(\mathbf{p},\mathbf{q})$. This is mapped onto itself by Γ, but Υ maps $\mathscr{X}^3$ onto a 12-dimensional space of *cubic* functions of p_i and q_j. We are thus led to define the vector functions

$$ {}^1\boldsymbol{\chi}^{1/2}_{1/2} = \mathbf{p}, \quad {}^1\boldsymbol{\chi}^{1/2}_{-1/2} = \mathbf{q}; \tag{2.8a} $$

$$ {}^3\boldsymbol{\chi}^{3/2}_{3/2} = p^2\mathbf{p}, \quad {}^3\boldsymbol{\chi}^{3/2}_{1/2} = \tfrac{1}{3}(2p\cdot q\mathbf{p} + p^2\mathbf{q}), $$
$$ {}^3\boldsymbol{\chi}^{3/2}_{-1/2} = \tfrac{1}{3}(q^2\mathbf{p} + 2p\cdot q\mathbf{q}), \quad {}^3\boldsymbol{\chi}^{3/2}_{-3/2} = q^2\mathbf{q}; \tag{2.8b} $$

$$ {}^3\boldsymbol{\chi}^{1/2}_{1/2} = \tfrac{1}{2}(-p\cdot q\mathbf{p} + p^2\mathbf{q}), $$
$$ {}^3\boldsymbol{\chi}^{1/2}_{-1/2} = \tfrac{1}{2}(-q^2\mathbf{p} + p\cdot p\mathbf{q}). \tag{2.8c} $$

The action of A^3 on $\mathscr{X}^3$ is effective (but not transitive) and is found to be thus indecomposable with the block form

$$ G\{\mathbf{v},w;\mathbf{M}\}\begin{pmatrix} {}^1\boldsymbol{\chi}^{1/2} \\ {}^3\boldsymbol{\chi}^{3/2} \\ {}^3\boldsymbol{\chi}^{1/2} \end{pmatrix} = \begin{pmatrix} \mathbf{M}^{-1} & \mathbf{M}^{-1}\mathbf{V} & 2w\mathbf{M}^{-1} \\ 0 & \mathbf{D}^{3/2}(\mathbf{M}^{-1}) & 0 \\ 0 & 0 & \mathbf{M}^{-1} \end{pmatrix}\begin{pmatrix} {}^1\boldsymbol{\chi}^{1/2} \\ {}^3\boldsymbol{\chi}^{3/2} \\ {}^3\boldsymbol{\chi}^{1/2} \end{pmatrix}, \tag{2.9a} $$

$$ \mathbf{V}(\mathbf{v}) = \begin{pmatrix} v_1 & 2v_0 & 3v_{-1} & 4v_{-2} \\ -4v_2 & -3v_1 & -2v_0 & -v_{-1} \end{pmatrix}. \tag{2.9b} $$

Explicitly, the first two rows yield the linear-plus-cubic transformation of the phase space coordinates

$$ G\left\{(v_2v_1v_0v_{-1}v_{-2}), w; \begin{pmatrix} a & b \\ c & d \end{pmatrix}\right\}\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} [1 + v_1p^2 + (\tfrac{2}{3}v_0 - w)p\cdot q + v_{-1}q^2]\mathbf{p} + [(\tfrac{2}{3}v_0 + w)p^2 + 2v_{-1}p\cdot q + 4v_{-2}q^2]\mathbf{q} \\ [-4v_2p^2 - 2v_1p\cdot q - (\tfrac{2}{3}v_0 + w)q^2]\mathbf{p} + [1 - v_1p^2 - (\tfrac{2}{3}v_0 - w)p\cdot q - v_{-1}q^2]\mathbf{q} \end{pmatrix}. \tag{2.9c} $$

For meridional rays, ${}^3\chi_m^{1/2} = 0$, so one-dimensional optics requires only the first 2×2 indecomposable block matrix in (2.9a). In particular, and corresponding to (2.7a),

$$G\{0,0;M\}\chi_m^j = \sum_{m'=-j}^{j} D^j_{mm'}(M^{-1})\chi^j_{m'} \tag{2.9d}$$

is valid for all j's (integer or half integer), independent of the order n.

Aligned lens systems have their optical elements represented by the following group elements.

Free propagation over length z in a homogeneous medium of index n_0 is represented by [(I 6.1)]

$$\mathbf{G}_H(z) := \exp(-z\hat{H}) = G\left\{\left(-\frac{z}{8n_0^3},0,0,0,0\right),0;\begin{pmatrix}1 & 0\\ -z/n_0 & 1\end{pmatrix}\right\}. \tag{2.10}$$

An up-to-quartic *refracting interface*

$$z = \zeta(\mathbf{q}) = \alpha q^2 + \beta(q^2)^2, \tag{2.11a}$$

where rays pass from a homogeneous medium n to another n', is represented (at the $z=0$ plane) by [(I 6.4)]

$$\mathbf{S}_\zeta = G\left\{\left(0,0,\frac{\alpha}{2}\left[\frac{1}{n'}-\frac{1}{n}\right],\frac{2\alpha^2}{n'}[n-n'],\ \frac{2\alpha^3}{n'}[n-n']^2+\beta[n-n']\right),\ \frac{2\alpha}{3}\left[\frac{1}{n'}-\frac{1}{n}\right];\begin{pmatrix}1 & -2\alpha[n-n']\\ 0 & 1\end{pmatrix}\right\}. \tag{2.11b}$$

On the basis of their separate action on phase space through (2.9c), the parameters $\mathbf{v}$, w were named or identified with the Seidel aberration[2] coefficients in I as follows. For the symplectic quintuplet, v_2 is the spherical aberration, v_1 is the coma and v_0 is the "curvatism" = curvature of field + astigmatism, v_{-1} is the distortion, and v_{-2} is the "pocus." For the symplectic singlet, w is the "astigmature" = $\frac{4}{3}$ curvature of field $-\frac{2}{3}$ astigmatism. The 2×2 matrix part $\mathbf{M}$ of $G\{\mathbf{v},w;\mathbf{M}\}$ gives rise to the familiar matrix formulation of Gaussian optics.

Compound optical systems (with a common optical axis extending to the *right*) are described through (i) writing for each optical element its group element (left to right); (ii) performing the product of group elements through (2.6), thus concatenating their optical counterparts, to build thus the system element; and (iii) calculating the action of a compound system on object light rays through (2.9c). In this article we shall work basically with the first point in the process for general inhomogeneous elements. Their compounding and action on phase space follows exactly as for lens systems.

III. PROPAGATION IN INHOMOGENEOUS MEDIA TO THIRD ABERRATION ORDER

We consider the general Hamiltonian $H(\mathbf{p},\mathbf{q};z)$ in (2.1) and pose the following question: given a medium described by a refraction index $n(q,z)$, what line of A^3 group elements $G\{\mathbf{v}(z),w(z);\mathbf{M}(z)\}$ will describe the z evolution of the system? If H were z independent, this line would be $\exp(-z\hat{H})$, a one-parameter subgroup with a fixed tangent vector $H(\mathbf{p},\mathbf{q})\in a^3$. If the tangent vector depends on z, the line twists through the group manifold and is not a subgroup. We shall turn the table on this question and ask the following: *Given* a line $G\{\mathbf{v}(z),w(z);\mathbf{M}(z)\}$ in A^3, what is the tangent vector? Once we have found it in terms of the line parameter functions and their derivatives, inverting these equations will give the group line generated by a particular Hamiltonian.

We address first the Gaussian part of the problem,[4] noting that for one-parameter groups generated by $x(z)$, we have $d/dz \exp\hat{x} = \widehat{(dx/dz)} \exp\hat{x}$, $\widehat{(dx/dz)} = d\hat{x}/dz$. Using (2.5b) for z-dependent matrix elements and indicating differentiation by a dot,

$$\frac{d}{dz}G\left\{0,0;\begin{pmatrix}1 & 0\\ c & 1\end{pmatrix}\right\} = \frac{1}{2}\dot{c}\hat{\chi}_1^1 G\left\{0,0;\begin{pmatrix}1 & 0\\ c & 1\end{pmatrix}\right\}, \tag{3.1a}$$

$$\frac{d}{dz}G\left\{0,0;\begin{pmatrix}a & 0\\ 0 & a^{-1}\end{pmatrix}\right\} = -\frac{\dot{a}}{a}\hat{\chi}_0^1 G\left\{0,0;\begin{pmatrix}a & 0\\ 0 & a^{-1}\end{pmatrix}\right\}, \tag{3.1b}$$

$$\frac{d}{dz}G\left\{0,0;\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix}\right\} = -\frac{1}{2}\dot{b}\hat{\chi}_{-1}^1 G\left\{0,0;\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix}\right\}. \tag{3.1c}$$

Next, since 2×2 unimodular matrices may be generically decomposed into a product of the three above, we may use on them the Leibnitz rule, (3.1) to differentiate, and (2.7a) to shift all generators to the left. Thus

$$\frac{d}{dz}G\left\{0,0;\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\right\} = [A\hat{\chi}_1^1 + B\hat{\chi}_0^1 + C\chi_{-1}^1]G\left\{0,0;\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\right\}, \tag{3.2a}$$

$$A = (\alpha\dot{\gamma} + \dot{\alpha}\gamma + \gamma^2[\dot{\alpha}\beta - \alpha\dot{\beta}])/2\alpha^2,\quad B = \dot{\beta}\gamma - \dot{\alpha}\delta,\quad C = \tfrac{1}{2}(\dot{\alpha}\beta - \alpha\dot{\beta}). \tag{3.2b}$$

The differentiation of a line $G\{\mathbf{v}(z),w(z);\mathbf{1}\}$ in the pure aberration subgroup Υ is trivial since the group is Abelian and the operator $\dot{\mathbf{v}}\cdot\chi^2 + \dot{w}\chi_0^0$ may be extracted on either side. Finally, we use the Leibnitz rule (2.5a) for the general group element $G\{\mathbf{v},w;\mathbf{M}\} = G\{\mathbf{v},w;\mathbf{1}\}G\{0,0;\mathbf{M}\}$ to obtain

$$\frac{d}{dz}G\{\mathbf{v},w;\mathbf{M}\} = \hat{L}(z)G\{\mathbf{v},w;\mathbf{M}\}, \tag{3.3}$$

where $L(z)$ is written

$$L(z) = A\chi_1^1 + B\chi_0^1 + C\chi_{-1}^1 + \sum_{m=-2}^{2}\lambda_m\chi_m^2 + \mu\chi_0^0, \tag{3.4}$$

with the Gaussian coefficients A, B, and C given by (3.2b), while for the aberration part

$$\boldsymbol{\lambda} = \dot{\mathbf{v}} + \mathbf{v}\mathbf{F},\quad \mu = \dot{w}, \tag{3.5a}$$

$$\mathbf{F} := \begin{pmatrix} -4B & -8C & 0 & 0 & 0\\ 2A & -2B & -6C & 0 & 0\\ 0 & 4A & 0 & -4C & 0\\ 0 & 0 & 6A & 2B & -2C\\ 0 & 0 & 0 & 8A & 4B \end{pmatrix} = \dot{\mathbf{D}}^2(\mathbf{M})\mathbf{D}^2(\mathbf{M}^{-1}). \tag{3.5b}$$

The additional term $\mathbf{vF}$ in $\boldsymbol{\lambda}$ results from the passage of the Gaussian generators through the Υ factor. The last expression for $\mathbf{F}$ comes from the equivalent process in differentiating $G\{\mathbf{v},w;\mathbf{M}\} = G\{\mathbf{0},0;\mathbf{M}\}G\{\mathbf{vD}(\mathbf{M}),w;\mathbf{1}\}$ with respect to z.

The one-parameter line integration representing evolution of phase space under a given inhomogeneous medium Hamiltonian $H(\mathbf{p},\mathbf{q};z)$ may be performed now using the result (3.3)–(3.5) and reasoning as follows. Let $f(\mathbf{p},\mathbf{q};0)$ be an object-plane function of phase space (at $z=0$), transformed by the optical system $\mathbb{G}_H(z)$ to an image-plane function at z:

$$f(\mathbf{p},\mathbf{q};0) \rightarrow f(\mathbf{p},\mathbf{q};z) = \mathbb{G}_H(z)f(\mathbf{p},\mathbf{q};0) = f(\mathbf{p}',\mathbf{q}';0). \tag{3.6a}$$

On the phase-space coordinates themselves, $\mathbf{p}\rightarrow\mathbf{p}'(\mathbf{p},\mathbf{q};z)$, $\mathbf{q}\rightarrow\mathbf{q}'(\mathbf{p},\mathbf{q};z)$. Then the z derivative of this and the Hamiltonian equations of motion (2.2) lead to[5]

$$\begin{aligned}\mathbb{G}_H(z)f(\mathbf{p},\mathbf{q};0) &= \frac{d}{dz}f(\mathbf{p},\mathbf{q};z) = -\widehat{H}(\mathbf{p},\mathbf{q};z)f(\mathbf{p},\mathbf{q};z)\\ &= -\widehat{H}(\mathbf{p},\mathbf{q};z)\mathbb{G}_H(z)f(\mathbf{p},\mathbf{q};0)\end{aligned} \tag{3.6b}$$

for arbitrary f. Hence, (3.3) should hold for $H = -L$ and the task is to find $\mathbb{G}_H(z)$ from a given $\widehat{L}(z)$. Yet, (3.2b) and (3.5) yield the parameters in the latter in terms of derivatives of those in the former. It is perhaps surprising that one may invert the Gaussian relations (3.2b) to sequential differential equations[4] for the parameters of G_H with coefficients in L, thus

$$\begin{cases} C=0: & \ddot{\alpha} + \alpha(\dot{B} - B^2) = 0,\\ & \alpha(0) = 1, \quad \dot{\alpha}(0) = -B(0),\\ C\neq 0: & \ddot{\alpha} - \dot{\alpha}\dot{C}/C + \alpha([4AC - B^2] \\ & \quad + \dot{B} - B\dot{C}/C) = 0,\end{cases} \tag{3.7a}$$

$$\dot{\beta} - \beta\dot{\alpha}/\alpha + 2C/\alpha = 0, \quad \beta(0) = 0, \tag{3.7b}$$

$$\begin{cases} C=0: & \dot{\gamma} + \gamma\dot{\alpha}/\alpha - 2\alpha A = 0, \quad \gamma(0) = 0,\\ C\neq 0: & \gamma = -(\dot{\alpha} + \alpha B)/2C.\end{cases} \tag{3.7c}$$

Lastly, δ is obtained from the unimodularity condition $\alpha\delta - \beta\gamma = 1$.

The aberration parameters in $\mathbb{G}_H(z)$ are obtained in terms of those in $L(z)$ through solving (3.5) with appropriate boundary conditions:

$$\dot{\mathbf{v}} + \mathbf{vF} - \boldsymbol{\lambda} = \mathbf{0}, \quad \mathbf{v}(0) = \mathbf{0}, \tag{3.7d}$$

$$\dot{w} - \mu = 0, \quad w(0) = 0. \tag{3.7e}$$

One application of the Gaussian-parameter differential equations above is to find the matrix $\mathbf{M}(\mathbf{u})$ representing $\exp(\Sigma_m u_m \chi^1_m)$, for $u_1 = Az$, $u_0 = Bz$, $u_1 = Cz$ (A, B, C constant, so all their derivative terms are zero). For $z=1$, we obtain $\mathbf{M}(\mathbf{u})$ as given in (2.5).

Another simple application of (3.7) is the description of graded-index media that are inhomogeneous only in the z-direction, i.e., $n = n_0(z)$. Then, the ruling Hamiltonian is $H = p^2/2n_0(z) + p^4/8n_0(z)^3 = -L$. We set in (3.7), $A = -1/2n_0(z)$, $\lambda_2 = -1/8n_0(z)^3$, and all other coefficients zero. Immediately we obtain the third-order aberration evolution operator

$$\mathbb{G}_H(z) = G\left\{\left(-\int_0^z \frac{dz'}{8n_0(z')^2},0,0,0,0\right),0;\ \begin{pmatrix} 1 & 0\\ -\int_0^z \frac{dz'}{n_0(z')} & 1\end{pmatrix}\right\}, \tag{3.8}$$

of which (2.10) is the particular case $n_0(z) = \text{const}$.

Axis-symmetrical optical fibers may be modeled as media where the refraction index n is a function of q^2. If the resulting Hamiltonian is restricted to a^3, then the most general form n can have is

$$n = n_0 - \nu q^2 - \rho q^4, \tag{3.9}$$

with possibly z-dependent parameters. If $\nu, \rho > 0$, the medium at the optical axis ($\mathbf{q} = \mathbf{0}$) will be densest and light rays will propagate oscillating around it. The corresponding "fiber" Hamiltonian in a^3 will then be

$$H(\mathbf{p},\mathbf{q}) = \frac{1}{2n_0}p^2 + \nu q^2 + \frac{1}{8n_0^3}p^4 + \frac{\nu}{2n_0^2}p^2q^2 + \rho q^4. \tag{3.10}$$

The foregoing integration algorithm (3.7) applies, thus, using

$$A = -1/2n_0, \quad B = 0, \quad C = -\nu; \tag{3.11a}$$

$$\begin{aligned}&\lambda_2 = -1/8n_0^3, \quad \lambda_1 = 0, \quad \lambda_0 = -\nu/2n_0^2 = \tfrac{3}{4}\mu,\\ &\lambda_{-1} = 0, \quad \lambda_{-2} = -\rho.\end{aligned} \tag{3.11b}$$

If the fiber has varying "diameter" (i.e., z-dependent ν and ρ), little simplification of (3.7a) is achieved by B being null, and an explicit solution for $\mathbb{G}_H(z)$ cannot be written down. Yet this is a well-defined numerical integration task that may be performed by computer. We note that we must calculate independently the Γ-projection Gaussian parameters in $\mathbf{M}(z)$ and the Υ-projection aberration parameters in $\mathbf{v}$ and w.

Since our inclination is towards explicit algebraic solutions,[6] we shall examine those cases where (3.7) can be solved in a closed form. Indeed, we shall approach now[7] the expression for $G\{\mathbf{v},w;\mathbf{M}\} = \exp(z\widehat{L})$, $\widehat{L}$ being a general element (4.3) of the algebra a^3, with constant coefficients. This will solve quartic z-homogeneous fibers in third aberration order, in the next section.

The Gaussian part of the solution has been given in (2.5b). We now introduce the *coherent-state* description of the aberration part. The gist of the matter is to diagonalize the Gaussian part $\mathbf{M}(\mathbf{u})$ so that the equations leading to the aberration coefficients (3.7d) and (3.7e) *uncouple*.

When in (3.4) $B = \omega$ and $A = 0 = C$, let us call L by $L^d = \boldsymbol{\lambda}^d\cdot\boldsymbol{\chi}^2 + \mu\chi^0_0 + \omega\chi^1_0$ (the index d for "diagonal"). Equation (3.3) then solves easily for $G\{\mathbf{v}^d,w;\mathbf{M}^d\} = \exp(z\widehat{L}^d)$ to yield $\mathbf{M}^d(\omega z) = \text{diag}\,(e^{-\omega z}, e^{\omega z})$ and, since $\mathbf{F}$ in (3.5b) is now $\mathbf{F}^d = \text{diag}(-4\omega, -2\omega, 0, 2\omega, 4\omega)$, the aberration part uncouples to $v^d_m = \lambda^d_m(e^{2m\omega z} - 1)/2m\omega$, $m = 2,1,-1,-2$, $v^d_0 = z\lambda^d_0$, and $w = z\mu$. This is the case where we may write

$$\begin{aligned}&\mathbf{v}^d = z\boldsymbol{\lambda}^d\mathbf{E}^d(\omega z),\\ &E^{d(j)}_{mm'}(x) = \delta_{mm'}\frac{e^{2mx} - 1}{2mx}, \quad m = j, j-1,\ldots,-j,\end{aligned} \tag{3.12}$$

formally including v_0^d as v_m^d, $m\to 0$. For third aberration order $j=2$. To achieve this diagonalization, we write the matrix in (2.5) as

$$\mathbf{M}(z\mathbf{u}) = B(\mathbf{u})\mathbf{M}^d(\omega z)B(\mathbf{u})^{-1},$$
$$\mathbf{B}(\mathbf{u}) = \begin{pmatrix} (\omega+u_0)/\omega\sqrt{2} & (\omega-u_0)/2u_1\sqrt{2} \\ -u_1\sqrt{2}/\omega & 1/\sqrt{2}\end{pmatrix}, \tag{3.13}$$

where $\mathbf{B}$ is the *Bargmann* matrix transforming to (the hyperbolic stratum element) $\omega\chi_0^1$; when $\omega^2<0$, as in the case of an "ordinary" fiber (acting as a harmonic oscillator and in the elliptic stratum), $\mathbf{B}$ is *complex* [i.e., an automorphism of sp(2,R) *outer* to Sp(2,R), *inner* to SL(2,C)] but well defined. Hence a similarity transformation by $G\{0,0;\mathbf{B}^{-1}\}$ sends Eq. (3.3) to the equation determining $G\{\mathbf{v}^d,w;\mathbf{M}^d\}$ above, the transformation back yields the general case

$$\begin{aligned}\exp(z[\mathbf{u}\cdot\hat{\chi}^1 + \boldsymbol{\lambda}\cdot\hat{\chi}^2 + \mu\hat{\chi}_0^0]) &= G\{0,0;\mathbf{B}\}\exp(z[\omega\hat{\chi}_0^1 + \boldsymbol{\lambda}\mathbf{D}^2(\mathbf{B})\cdot\hat{\chi}^2 + \mu\hat{\chi}_0^0])\\ &\quad\times G\{0,0;\mathbf{B}\}^{-1}\\ &= G\{0,0;\mathbf{B}\}G\{z\boldsymbol{\lambda}\mathbf{D}^2(\mathbf{B})\mathbf{E}^d(\omega z),\mu z;\mathbf{M}^d(\omega z)\}\\ &\quad\times G\{0,0;\mathbf{B}\}^{-1}\\ &= G\{z\boldsymbol{\lambda}\dot{\mathbf{E}}(z\mathbf{u}),\mu z;\mathbf{M}(z\mathbf{u})\},\end{aligned} \tag{3.14a}$$

$$\mathbf{E}(z\mathbf{u}) = \mathbf{D}^2(\mathbf{B}(\mathbf{u}))\mathbf{E}^d(z\omega)\mathbf{D}^2(\mathbf{B}(\mathbf{u})^{-1}),\quad \mathbf{M} = \mathbf{B}\mathbf{M}^d\mathbf{B}^{-1}. \tag{3.14b}$$

We remind the reader that $\mathbf{D}^2(\mathbf{B})$ is found using (2.7) or, explicitly, (I 4.6) for $j=2$. (The $\mathbf{E}$-matrices do *not*[6] constitute a group representation.) This is the general relation between the a^3 algebra elements $L(\boldsymbol{\lambda},\mu,\mathbf{u})$ and the A^3 group elements. Since we only have matrix-cum-vector algebra, the relation just obtained solves—once and for all—*every* Baker-Campbell-Hausdorff-Zassenhaus formula in $\exp a^3$ in a closed, sometimes lengthy, but algebraic formula.

We shall close this section with the explicit computation of the $\mathbf{E}$-matrix in (3.14b) for upper-triangular $\mathbf{M}$ corresponding to $u_1=0$ (so $\omega=u_0$) and, then, the (parabolic-stratum) limit $u_0\to 0$. The purpose is to provide the logarithm of the refracting surface transformation (2.11) reported before [Ref. 7, Eq. (7.10)].

For $\mathbf{u}=(0,u_0,u_{-1})$, we have, from (2.5) and an appropriate limit in (3.13),

$$\mathbf{M}(\mathbf{u}) = \begin{pmatrix} e^{-u_0} & -2u_{-1}u_0^{-1}\sinh u_0 \\ 0 & e^{u_0}\end{pmatrix},$$
$$\mathbf{B}(\mathbf{u}) = \begin{pmatrix}\sqrt{2} & -u_{-1}/u_0\sqrt{2} \\ 0 & 1/\sqrt{2}\end{pmatrix}, \tag{3.15a}$$

$$D^j_{mm'}(\mathbf{M}(\mathbf{u})) = \binom{j+m}{j+m'} e^{-u_0(m+m')}(-2u_{-1}u_0^{-1}\sinh u_0)^{m-m'}, \tag{3.15b}$$

$$E^j_{mm'}(0,u_0,u_{-1}) = \frac{(j+m)!}{(j+m')!}\left(\frac{u_{-1}}{u_0}\right)^{m-m'}\sum_{m''=m'}^{m} \times \frac{(-1)^{m-m''}}{(m-m'')!(m''-m')!}\left(\frac{e^{2m''u_0}-1}{2m''u_0}\right). \tag{3.15c}$$

Recalling that rows and columns are numbered by $m,m'=j,j-1,\dots,-j$, we see that both the $\mathbf{D}$ and $\mathbf{E}$ matrices are upper triangular (i.e., nonzero for $m\geqslant m'$).

Now we may let $u_0\to 0$; in the limit, $\mathbf{M}(\mathbf{u})$ is no longer diagonalizable, but $\mathbf{E}(0,0,u_{-1})$ exists. It is

$$E^j_{mm'}(0,0,u_{-1}) = \binom{j+m}{j+m'}\frac{(2u_{-1})^{m-m'}}{m-m'+1}. \tag{3.16}$$

The proof[7] proceeds through Taylor-expanding the exponential in (3.15c) and grouping powers of u_0^n; the coefficients are recognized as the $(m-m')$th finite difference of the displaced nth power function, which is zero for $n<m-m'$. It may also be proved directly through the Schwinger series of the exponential group action to $2j$ terms. The net result of the above expressions is, for $j=2$, the proof[9] of

$$\begin{aligned}\exp(u\hat{\chi}^1_{-1} + \boldsymbol{\lambda}\cdot\hat{\chi}^2 + \mu\hat{\chi}_0^0) &= \exp(uq^2 + \lambda_2 p^4 + \lambda_1 p^2 p\cdot q + [\tfrac{1}{3}\lambda_0 + \tfrac{1}{2}\mu]p^2q^2\\ &\quad + [\tfrac{2}{3}\lambda_0 - \tfrac{1}{2}\mu](p\cdot q)^2 + \lambda_{-1}p\cdot qq^2 + \lambda_{-2}q^4)^\wedge\\ &= G\Big\{(\lambda_2,\lambda_1 + 4u\lambda_2,\lambda_0 + 3u\lambda_1 + 8u^2\lambda_2,\\ &\quad \lambda_{-1} + 2u\lambda_0 + 4u^2\lambda_1 + 8u^3\lambda_2,\\ &\quad \lambda_{-2} + 4\lambda_{-1} + \tfrac{4}{3}u^2\lambda_0 + 2u^3\lambda_1 + \tfrac{16}{5}u^4\lambda_2),\\ &\quad \mu;\begin{pmatrix}1 & -2u\\ 0 & 1\end{pmatrix}\Big\}.\end{aligned} \tag{3.17}$$

It is instructive to verify the validity of (3.14) and (3.17) for $\exp(z_1\hat{L})\exp(z_2\hat{L}) = \exp([z_1+z_2]\hat{L})$. Solving for u, $\boldsymbol{\lambda}$, and μ in (3.17) one may find the logarithm of the surface transformation (2.11) to be

$$\mathbf{S}_\zeta = \exp\Big([n-n']\Big[\alpha q^2 + \frac{\alpha}{2nn'}p^2q^2 + \frac{\alpha^2(n+n')}{nn'}p\cdot qq^2 + \Big\{\beta + \frac{\alpha^3(n-n')^2}{3nn'}\Big\}q^4\Big]\Big)^\wedge. \tag{3.18}$$

There is otherwise no simple one-parameter subgroup description of the refracting surface transformation. Nevertheless, as shown in Ref. 8, one may *factorize* $\mathbf{S}_z$ into two different, but similar and simpler, "root" transformations

$$\mathbf{S}_\zeta = \mathbf{R}_{\zeta,n}\mathbf{R}^{-1}_{\zeta n'},$$
$$\mathbf{R}_{\zeta,n} = \exp(n\alpha q^2 - (\alpha/n)p^2q^2 + 2\alpha p\cdot qq^2 + n[\beta - \tfrac{2}{3}\alpha^3]q^4)^\wedge. \tag{3.19}$$

IV. UP-TO-QUARTIC FIBER OPTICS IN THIRD ABERRATION ORDER

We may now construct the one-parameter subgroup of A^3 elements generated by the Hamiltonian (3.10), modeling an up-to-quartic fiber of length z in third aberration order. It is

$$\mathbb{G}_H(z) = G\{\mathbf{v}(z),w(z);\mathbf{M}(z)\} = \exp(-z\hat{H}). \tag{4.1}$$

For the Gaussian part Γ, from (2.5) it follows that

$$\mathbf{M}(z) = \begin{pmatrix}\cos\kappa z & \kappa n_0\sin\kappa z\\ -(\kappa/2\nu)\sin\kappa z & \cos\kappa z\end{pmatrix},\quad \kappa = \pm\sqrt{2\nu/n_0}. \tag{4.2}$$

When $\nu > 0$, κ is real [and $\mathbf{M}(z)$ belongs to the elliptic stratum]. The motion is oscillatory with "Gaussian" period $\lambda_G = 2\pi/\kappa = \pi\sqrt{2n_0/\nu}$.

For the aberration part we use (3.14). The symplectic singlet ("astigmature") is

$$w(z) = -\tfrac{2}{3}(\nu/n_0^2)z, \tag{4.3}$$

increasing linearly with time. Finally, to evaluate the symplectic quintuplet of aberrations, the Bargmann transform matrix (3.13) is calculated to be

$$\mathbf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & ib \\ i/b & 1 \end{pmatrix}, \quad b = \kappa n_0 = \sqrt{2n_0\nu}. \tag{4.4}$$

This is z independent since (4.1) is a subgroup. From here one proceeds to find $\mathbf{D}^2(\mathbf{B})$ and $D^2(B^{-1}) = D^2(B)^*$ to calculate $\mathbf{E}(z)$ in (3.14b). It is convenient to abbreviate

$$S_m(z) := \frac{\sin m\kappa z}{m\kappa}, \quad C_m(z) := \frac{1-\cos m\kappa z}{m\kappa}, \tag{4.5}$$

so that we may write

$$z\mathbf{E} = \begin{pmatrix} \tfrac{1}{8}[S_4+4S_2+3z] & \tfrac{1}{2}b[-C_4-2C_2] & \tfrac{3}{4}b^2[-S_4+z] & \tfrac{1}{2}b^3[C_4-2C_2] & \tfrac{1}{8}b^4[S_4-4S_2+3z] \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \tfrac{1}{8}b^{-2}[-S_4+z] & \tfrac{1}{2}b^{-1}C_4 & \tfrac{3}{4}[S_4+\tfrac{1}{3}z] & -\tfrac{1}{2}bC_4 & \tfrac{1}{8}b^2[-S_4+z] \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \tfrac{1}{8}b^{-4}[S_4-4S_2+3z] & \tfrac{1}{2}b^{-3}[-C_4+2C_2] & \tfrac{3}{4}b^{-2}[-S_4+z] & \tfrac{1}{2}b^{-1}[C_4+2C_2] & \tfrac{1}{8}[S_4+4S_2+3z] \end{pmatrix}. \tag{4.6a}$$

The matrix $\mathbf{E}$ will act on the row vector $\boldsymbol{\lambda}z$ whose components for the fiber Hamiltonian are given in (3.11b), so only the first, third, and fifth rows are actually needed; the two others[7] are replaced by dots. If we define the fiber parameters

$$\eta := \frac{\nu}{4n_0^2}\left(1 - \frac{2n_0\rho}{\nu^2}\right), \quad \zeta = \frac{\nu}{4n_0^2}\left(5 + \frac{6n_0\rho}{\nu^2}\right), \tag{4.6b}$$

we may write the values of the aberration coefficients after length z as

spherical aberration $\quad v_2(z) = \tfrac{1}{8}(2n_0\nu)^{-1}[\eta(S_4-4S_2)-\zeta z]$, (4.7a)

coma $\quad v_1(z) = \tfrac{1}{2}(2n_0\nu)^{-1/2}\eta[-C_4+2C_2]$, (4.7b)

"curvatism" $\quad v_0(z) = -\tfrac{3}{4}[\eta S_4 + \tfrac{1}{3}\zeta z]$, (4.7c)

distortion $\quad v_{-1}(z) = \tfrac{1}{2}(2n_0\nu)^{1/2}\eta[C_4+2C_2]$, (4.7d)

"pocus" $\quad v_{-2}(z) = \tfrac{1}{8}(2n_0\nu)[\eta(S_4+4S_2)-\zeta z]$. (4.7e)

The quartic fiber evolution element (4.1) is thus given by the Gaussian part (4.2), singlet aberration (4.3), and quintuplet aberration (4.7).

The coefficient η gives the relative size of oscillating aberration terms with period $\tfrac{1}{4}\lambda_G$ and $\tfrac{1}{2}\lambda_G$, while ζ multiplies the linear drift terms. One may verify that as ρ/ν^2, $\nu\to 0$, one regains the free propagation group line (2.10); $\nu\to 0$ but ρ finite leads to the Fourier conjugate of the corresponding formula (3.17).

Some properties of aberration in fibers are evident from the above explicit expressions. Since for $z = \tfrac{1}{2}N\lambda_G = N\pi\sqrt{n_0/2\nu}$, N integer, both S_m and C_m vanish, for those lengths the fiber will have no coma nor distortion, and will only present the drift term ($\sim z$) for the other four aberrations. The oscillation of all quintuplet aberrations may be eliminated by putting η to zero through building the fiber with $\rho = \nu^2/2n_0$. The drift coefficient ζ in the quintuplet may be set to zero when $\rho = -5\nu^2/6n_0$. Through length (z) and gradation n-shape ($\rho{:}\nu$) we may thus eliminate the five quintuplet aberration coefficients. This will not cure the aberration singlet, however (recall I, Sec. VII), which will grow linearly with z, unfocusing object ray cones into images which grow a segment perpendicular to the meridional direction.

We may trace the path of a light ray in an optical fiber specified by the Taylor-expansion parameters of $n(q)$ in five-dimensional space $(\mathbf{p},\mathbf{q};z)$. It is analytically given by (2.9c) through the replacement of the parameters in (4.2), (4.3), and (4.7), with the general form

$$\mathbf{q}' = \mathbf{Q}(\mathbf{p},\mathbf{q};z;n), \quad \mathbf{p}' = \mathbf{P}(\mathbf{p},\mathbf{q};z;n), \tag{4.8}$$

and the following properties: (i) Q and P contain terms that are linear and cubic in the initial phase-space coordinates, and linear combinations of (2.8), (ii) $\mathbf{Q}(\mathbf{p},\mathbf{q};0;n) = \mathbf{q}$ and $\mathbf{P}(\mathbf{p},\mathbf{q};0;n) = \mathbf{p}$, and (iii) the inverse functions are obtained replacing z by $-z$. [We note the special case $n = n_0$, which corresponds to a homogeneous medium propagation (2.10).]

Suppose two fiberlike media, the first defined by the Taylor parameters of n and the second by those of n', are joined at an interface of the form (2.11a), which thus becomes a refracting interface between the two fibers. If the two media are homogeneous, this is the ordinary surface of an up-to-quartic lens. The effect of this interface may be associated to an A^3 element following the argument presented in Ref. 8, replacing free propagation by the fiber propagation (4.8).

A ray $(\mathbf{p},\mathbf{q})$ crossing the $z=0$ plane will hit the surface ζ at a point $\bar{\mathbf{q}}$ with momentum $\mathbf{p}_\zeta$, a distance $z = \zeta(\bar{\mathbf{q}})$ along the optical axis. The point $\bar{\mathbf{q}}$ is defined by the *implicit* first equality in

$$\mathbf{Q}(\mathbf{p},\mathbf{q};\zeta(\bar{\mathbf{q}});n) = \bar{\mathbf{q}} = \mathbf{Q}(\mathbf{p}',\mathbf{q}';\zeta(\bar{\mathbf{q}});n'). \tag{4.9}$$

The second equality, also implicit, states that $\bar{\mathbf{q}}$ also may be reached from a $z=0$ point $(\mathbf{p}',\mathbf{q}')$ (to be determined) through propagation as if it were in the environment of the *second* medium. Since $\zeta(\bar{\mathbf{q}})$ is at least quadratic in $\bar{\mathbf{q}}$ and can be most cubic within the A^3 model, repeated self-replacement of $\bar{\mathbf{q}}$ in the first $\mathbf{Q}$ function will be terminated after the first step. The result will be an explicit expression $\bar{\mathbf{q}}(\mathbf{p},\mathbf{q})$ as a linear-plus-cubic transformation depending on the parameters n_0, ν, ρ. Similarly, the second equality yields an expression for $\bar{q}(\mathbf{p}',\mathbf{q}')$ depending on the parameters n_0', ν', ρ'. These relations do not yet fully determine an A^3 element mapping $(\mathbf{p},\mathbf{q})$ to $(\mathbf{p}',\mathbf{q}')$, since there is as yet no relation between the value and direction of $\mathbf{p}$ propagated to the surface, $\mathbf{p}_\zeta$, $|\mathbf{p}_\zeta| = n(\bar{q})\sin\theta$, and $\mathbf{p}'$ propagated to the same surface, $\mathbf{p}_\zeta'$, $|\mathbf{p}_\zeta'| = n'(\bar{q})\sin\theta'$. This is provided by Snell's law, which states that, if ϕ and ϕ' are the angles between the ray three-dimensional vectors $\mathbf{p}_\zeta,\mathbf{p}_\zeta'$ and the vector $\nabla\zeta(\bar{q})$ normal to ζ at $\bar{q}$, then

$$n(\bar{\mathbf{q}})\sin\phi = n'(\bar{\mathbf{q}})\sin\phi', \tag{4.10a}$$

and that the three vectors are coplanar. Now, if ψ is the angle between $\nabla\zeta(\bar{q})$ and the $z=0$ plane, then $\phi=\psi+\theta$ and $\phi'=\psi+\theta'$. We replace above and expand the trigonometric functions in ϕ, ϕ' in terms of θ, θ', and $\tan\psi = |\nabla\zeta(\bar{\mathbf{q}})|$. Recognizing that since this takes place on any plane containing the three vectors, the equation is satisfied vectorially,

$$p_\zeta + \sqrt{n(\bar{q})^2 - p_\zeta^2}\,(\nabla\zeta)(\bar{q}) =: \bar{p} = p_\zeta' + \sqrt{n'(\bar{q})^2 - p_\zeta'^2}\,(\nabla\zeta)(\bar{q}), \tag{4.10b}$$

where we have divided both sides by $\cos\psi$ and defined $\bar{p}$ by the second equality.

The refracting surface transformation we are looking for thus maps (i) $(\mathbf{p},\mathbf{q})$ to $(\mathbf{p}_\zeta,\bar{\mathbf{q}})$ through propagation in the first medium, (ii) $(\mathbf{p}_\zeta,\bar{\mathbf{q}})$ to $(\mathbf{p}_\zeta',\bar{\mathbf{q}})$ through obeyance to Snell's law (4.10), and (iii) $(\mathbf{p}_\zeta',\bar{\mathbf{q}})$ to $(\mathbf{p}',\mathbf{q}')$ through inverse free propagation in the second medium. Note that step (ii), as written in (4.10b), may be broken into two steps: (ii_1) $(\mathbf{p}_\zeta,\bar{\mathbf{q}})$ to $(\bar{\mathbf{p}},\bar{\mathbf{q}})$, and (ii_2) $(\bar{\mathbf{p}},\bar{\mathbf{q}})$ to $(\mathbf{p}_\zeta',\bar{\mathbf{q}})$. The concatenation of (i) and (ii_1) is the "root" transformation in the first medium (examined in Ref. 8), of which the concatenation of (ii_2) and (iii) is the inverse, in the second medium. The statements are independent of the aberration order.

We now implement the transformations explicitly to third order for the surface (2.11a), setting $z=\zeta(\bar{q}) = \alpha\bar{q}^2 + \beta\bar{q}^4$ in (4.2), (4.3), and (4.7). We note that all aberration coefficients will multiply cubic terms and thus only the Gaussian part need be calculated for step (i). It yields[9]

$$\bar{\mathbf{q}} = \mathbf{q} + (\alpha/n_0)q^2\mathbf{p}, \tag{4.11}$$

$$\mathbf{p}_\zeta = \mathbf{p} - 2\nu\alpha q^2\mathbf{q}. \tag{4.12a}$$

Step (ii_1) yields

$$\bar{\mathbf{p}} = \mathbf{p}_\zeta + 2n_0\alpha\bar{\mathbf{q}} - (\alpha/n_0)p_\zeta^2\bar{\mathbf{q}} + (4n_0\beta - 2\nu\alpha)\bar{q}^2\bar{\mathbf{q}}. \tag{4.12b}$$

The concatenation of the preceding equations yields the root transformation as (4.11) and

$$\bar{\mathbf{p}} = \mathbf{p} + 2n_0\alpha\mathbf{q} - (\alpha/n_0)p^2\mathbf{q} + 2\alpha^2q^2\mathbf{p} + 4(n_0\beta - \nu\alpha)q^2\mathbf{q}. \tag{4.13}$$

We note that this transformation does not depend on the quartic fiber parameter ρ, and that it is the same form as the root transformations by inhomogeneous media [Ref. 8, Eq. (7.2) and (I 2.10)]. The only difference lies in the $q^2\mathbf{q}$ term added to $\bar{p}$, which is the last summand above, it replaces the $n\beta$ term of homogeneous media interfaces by $n_0\beta - \nu\alpha$ above; only there does the quadratic fiber parameter ν appear.[10] The concatenation of the two root transformations (4.11a)–(4.13) may be now performed through $\mathbf{S}_{\zeta,n,n'} = \mathbb{R}_{\zeta,n}\mathbb{R}_{\zeta,n'}^{-1}$ as shown in Ref. 8, yielding (2.11b) with the same replacement for $n\beta$ and $n'\beta$.

V. OUTLOOK

The group-theoretical description of geometric optics seems to lead presently to development in the following directions: (1) classification of higher-order aberrations as multiplets under the Gaussian symplectic group, be this $Sp(2,R)$ for axis-symmetric systems, or, otherwise, (2) $Sp(4,R)$. In nonsymmetric magnetic optics with chromatic dispersion[2], $Sp(6,R)$ is required.

The symplectic groups possess integral transform representations that provide a kind of wave "quantization" of the system. This is well known and has been applied to Gaussian wave optics of lens systems,[11] so wave optical systems with *aberration* should follow the same line of development through embedding A^3 as a *subgroup* within the Heisenberg–Weyl *ring*.[12] Some progress has been made in this direction in collaboration with W. Lassner, NTZ Karl-Marx University, Leipzig, and W. Schempp, University of Siegen.

ACKNOWLEDGMENTS

I would like to thank Dr. Doz. Wolfgang Lassner for the hospitality extended at the Naturwissenschaftlich-Theoretisches Zentum, Karl-Marx University, Leipzig, where this paper was written, giving a modest tribute to Sophus Lie, who also worked at the Leipziger Universität.

[1] M. Navarro-Saad and K. B. Wolf, "The group-theoretical treatment of aberrating systems. I. Aligned lens systems in third aberration order," J. Math. Phys. **27**, 1449 (1986).

[2] A. J. Dragt, *Lectures on Nonlinear Orbit Dynamics A.I.P. Conference Proceedings*, No. 87 (A.I.P., New York, 1982).

[3] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U.P., Cambridge, 1984).

[4] K. B. Wolf, "On time-dependent quadratic quantum Hamiltonians," SIAM J. Appl. Math. **40**, 419 (1981).

[5] Compare with A. J. Dragt and E. Forest, "Computation of nonlinear behavior of Hamiltonian systems using Lie algebraic methods," J. Math. Phys. **24**, 2734 (1983), Sec. 4, Eq. (4.7), we note that $\mathbb{G}_H(z) \times \hat{L}(\mathbf{p},\mathbf{q})\mathbb{G}_H(z)^{-1} = \hat{L}(\mathbf{p}',\mathbf{q}')$. Note in this regard the work by S. Steinberg, "Factored product expansions of solutions of nonlinear differential equations," SIAM J. Math. Anal. **15**, 108 (1984), which may be used for higher aberration orders.

[6] Compare A. J. Dragt, "Lie algebraic theory of geometrical optics and optical aberrations," J. Opt. Soc. Am. **72**, 372 (1982), Eqs. (5.19)–(5.24).

[7] K. B. Wolf, "Approximate canonical transformations and the treatment of aberrations. I. One-dimensional simple Nth-order aberrations in optical systems," Comunicaciones Técnicas IIMAS No. 352, 1983.

[8] M. Navarro-Saad and K. B. Wolf, "Factorization of the phase-space

transformation produced by an arbitrary refracting surface," preprint CINVESTAV, March 1984.
[9]K. B. Wolf, "A group-theoretical model for Gaussian optics and third-order aberrations," in *Proceedings of the XII Colloquium on Group-theoretical Methods in Physics, Lecture Notes in Physics*, Trieste, 1983 (Springer, Berlin, 1984).
[10]I am indebted to Prof. Alex Dragt and Dr. Etienne Forest for pointing out an error in the original calculation.
[11]M. Nazarathy and J. Shamir, "First-order optics—a canonical operator representation: lossless systems," J. Opt. Soc. Am. **72**, 356 (1982).
[12]K. B. Wolf, "The Heisenberg-Weyl ring in quantum mechanics," in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. 3.

On steady transverse magnetogas flow

S. N. Singh, K. R. Choubey, and B. P. Singh
Department of Mathematics, Faculty of Science, Banaras Hindu University, Varanasi 221005, India

(Received 26 September 1985; accepted for publication 8 January 1986)

The geometry and solutions for steady transverse magnetogas dynamic flow with arbitrary equation of state when magnitude of the velocity is constant along each individual streamline are investigated.

I. INTRODUCTION

In spite of the general difficulty of integrating the momentum equation, there exists a large body of special exact solutions. Much of the literature on the subject is continued in the encyclopedic article of Berker.[1] Prim and Nemenyi[2] studied the geometry of plane, steady rotational flow of an inviscid perfect gas with the velocity magnitude constant along each individual streamline. This result was later extended by Prim[3] for gases with a product equation of state. Martin[4] modified Prim's analysis and generalized his result to any fluid for which the density is a function of the pressure along a streamline. Ansari and Babu[5] also discussed the case for transverse hydromagnetic flow. An excellent survey of this method, with applications to numerous nonlinear problems, has been given by Ames.[6]

Another kinematic class that has received a considerable amount of attention consists of circulation-preserving solutions. Such solutions are dynamically possible in any compressible magneto or classical fluids, provided that the boundary of the flow region be subjected to appropriate fractions. Hence, the circulation-preserving solutions are called universal solutions of hydromagnetic fluids. In the case of plane flows, Kampé De Feriet[7] has determined the class of motions that not only satisfy the integrability condition but also annul the only nonlinear term in that condition. His work was significant because it followed the tradition of hydrodynamics and sought to determine all the solutions within a well-defined kinematic class. This work was the first step towards the systematic study of all the exact solutions of hydrodynamic flows. Recently Singh and Singh[8] and Choubey and Singh[9] obtained the necessary and sufficient condition for the existence of circulation-preserving flow in hydromagnetic fluids. In the case of plane MFD flows, the circulation preserving property implies that the nonlinear term in the integrability of equation of motion vanishes. This condition yields the previously mentioned class of solutions of De Feriet[7] and Choubey and Singh,[9] although their brief papers contain only an outline of the analysis. The later work of Gortler and Weighardt,[10] though frequently cited in the literature, contains in fact no new solution.

In this paper, the results from differential geometry are used systematically to investigate plane solutions of MFD flows. Although the scope of the present work is limited to circulation preserving MFD flows, the method can be applied to more general cases. In Sec. IV, we prove the only possible plane MFD flows are (i) motions having constant vorticity magnitude and (ii) motions whose streamlines are concentric circles. In Sec. III, by applying Legendre transform technique to the problem at hand, we achieve the geometry of flows and also obtain solutions for such flows. In Sec. V, we prove that the analysis of Sec. III is applicable to sonic flows also.

II. FLOW EQUATIONS

The motion of an adiabatic, steady compressible hydromagnetic fluid is given by[11]

$$\operatorname{div}(\rho \mathbf{v}) = 0, \tag{2.1}$$

$$\rho[\nabla(\tfrac{1}{2}V^2) - \mathbf{v}\times(\nabla\times\mathbf{v})] = -\nabla p + \mu \operatorname{curl}\mathbf{H}\times\mathbf{H}, \tag{2.2}$$

$$\operatorname{curl}(\mathbf{v}\times\mathbf{H}) = \mathbf{0} \tag{2.3}$$

$$\mathbf{v}\cdot\operatorname{grad} S = 0, \tag{2.4}$$

$$p = f(\rho, S), \tag{2.5}$$

and

$$\operatorname{div}\mathbf{H} = 0, \tag{2.6}$$

where $\mathbf{v}$ is the velocity vector, $\mathbf{H}$ is the magnetic field vector, p is the pressure function, S the specific entropy function, μ the magnetic permeability, and ρ the density function. The flows under consideration have velocity magnitude constant along each individual streamline and therefore the velocity vector field satisfies the equation

$$\mathbf{v}\cdot\operatorname{grad}(\tfrac{1}{2}V^2) = 0. \tag{2.7}$$

In the case of plane transverse flow in the (x,y) plane, we have $\mathbf{H} = (0,0,H)$ and $\partial/\partial z \equiv 0$. For such a flow, Eqs. (2.2) and (2.3) reduce, respectively, to

$$\rho[\nabla(\tfrac{1}{2}V^2) - \mathbf{v}\times(\nabla\times\mathbf{v})] = -\nabla(p + \tfrac{1}{2}\mu H^2), \tag{2.8}$$

$$\operatorname{div}(H\mathbf{v}) = 0, \tag{2.9}$$

where $V = |\mathbf{v}|$ and $H = |\mathbf{H}|$. From Eqs. (2.1) and (2.9) we observe that

$$H = \alpha, \tag{2.10}$$

where α is a scalar function satisfying condition

$$\mathbf{v}\cdot\operatorname{grad}\alpha = 0. \tag{2.11}$$

Taking the dot product of the linear momentum equation with $\mathbf{v}$ and using (2.7) and (2.11), we obtain

$$\mathbf{v}\cdot\operatorname{grad} p = 0, \tag{2.12}$$

which implies that the pressure function is constant on each individual streamline. From (2.5) and (2.12) it follows that

$$\mathbf{v}\cdot\operatorname{grad}\rho = 0. \tag{2.13}$$

Using (2.13), the continuity equation (2.1) reduces to

$$\operatorname{div} \mathbf{v} = 0. \qquad (2.14)$$

Taking the curl of the linear momentum equation and using (2.12) and (2.13), we get

$$\operatorname{curl}(\mathbf{v}\times\mathbf{w}) = 0, \quad \mathbf{w} = \operatorname{curl} \mathbf{v}. \qquad (2.15)$$

From (2.15), we have the following theorem.

Theorem 1: The steady transverse MFD flow is circulation preserving if the velocity magnitude is constant along each streamline.

The proof of this theorem is obvious because equation (2.15) is itself necessary and sufficient condition for the flow to be circulation preserving.[12] Now the equations satisfied by the velocity field for the given motion, therefore, are (2.11), (2.14), and (2.15).

III. ANALYSIS FOR PLANE FLOWS

Considering plane flows where $\mathbf{v} = \mathbf{v}(x,y) = \{u(x,y),v(x,y)\}$ in the physical (x,y) plane, u and v satisfy

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \qquad (3.1)$$

$$u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} = 0, \quad w = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \qquad (3.2)$$

$$u^2\frac{\partial u}{\partial x} + v^2\frac{\partial v}{\partial y} + uv\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) = 0. \qquad (3.3)$$

We define the Jacobian for transformation of the above equations from the physical plane to hodrograph plane as $J = \{\partial(u,v)/\partial(x,y)\}$.

We shall now consider the following cases: (1) $J \neq 0$ in the entire region of the flow, (2) $J = 0$ in the entire region of the flow, and (3) $J = 0$ in the part of the region and $J \neq 0$ in the remaining region of the flow.

A. Case (1) ($J\neq 0$)

In this case, the system of equations (3.1)–(3.3) can be transformed into an equivalent linear system by interchanging the roles of dependent and independent variables. For a solution $u(x,y)$, $v(x,y)$, we may consider x and y as functions of u and v. Employing the transformations

$$\frac{\partial u}{\partial x} = J\frac{\partial y}{\partial v}, \quad \frac{\partial u}{\partial y} = -J\frac{\partial x}{\partial v},$$

$$\frac{\partial v}{\partial x} = -J\frac{\partial y}{\partial u}, \quad \frac{\partial v}{\partial y} = J\frac{\partial x}{\partial u},$$

in Eqs. (3.1)–(3.3) we see that $x(u,v)$ and $y(u,v)$ satisfy the following linear differential equations

$$\frac{\partial x}{\partial u} + \frac{\partial y}{\partial v} = 0, \qquad (3.4)$$

$$u\frac{\partial(w,y)}{\partial(u,v)} + v\frac{\partial(x,w)}{\partial(u,v)} = 0, \qquad (3.5)$$

$$u^2\frac{\partial y}{\partial v} + v^2\frac{\partial x}{\partial u} - uv\left(\frac{\partial y}{\partial u} + \frac{\partial x}{\partial v}\right) = 0. \qquad (3.6)$$

Equation (3.4) implies the existence of a Legendre transfer function $L(u,v)$ such that

$$dL = x\,dv - y\,du, \quad \frac{\partial L}{\partial u} = -y, \quad \frac{\partial L}{\partial v} = x. \qquad (3.7)$$

Employing (3.7) in (3.4)–(3.6), Eq. (34) is identically satisfied and Eqs. (3.5) and (3.6) take the form

$$v\frac{\partial((\partial L/\partial v),w)}{\partial(u,v)} - u\frac{\partial(w,(\partial L/\partial u))}{\partial(u,v)} = 0, \qquad (3.8)$$

$$\frac{\partial^2 L}{\partial u\,\partial v}(v^2 - u^2) - \left(\frac{\partial^2 L}{\partial u^2} - \frac{\partial^2 L}{\partial v^{23}}\right)uv = 0, \qquad (3.9)$$

where

$$w = J\left(\frac{\partial^2 L}{\partial u^2} + \frac{\partial^2 L}{\partial v^2}\right),$$

$$J = \left[\frac{\partial^2 L}{\partial u^2}\frac{\partial^2 L}{\partial v^2} - \left(\frac{\partial^2 L}{\partial u\,\partial v}\right)^2\right]^{-1}.$$

Defining polar coordinates (q,θ) in the (u,v) plane given by $q = (u^2+v^2)^{1/2}$ and $\theta = \tan^{-1}(v/u)$, the above equations are reduced to

$$\frac{\partial w}{\partial\theta}\frac{\partial^2 L}{\partial q^2} - \frac{\partial w}{\partial q}\left(\frac{\partial^2 L}{\partial q\,\partial\theta} - \frac{1}{q}\frac{\partial L}{\partial\theta}\right) = 0, \qquad (3.10)$$

$$\frac{\partial^2 L}{\partial\theta\,\partial q} - \frac{1}{q}\frac{\partial L}{\partial\theta} = 0, \qquad (3.11)$$

where

$$w = J\left[q^4\left(\frac{\partial^2 L}{\partial q^2} + \frac{1}{q^2}\frac{\partial^2 L}{\partial\theta^2} + \frac{1}{q}\frac{\partial L}{\partial q}\right)\right] \times\left[q^2\frac{\partial^2 L}{\partial q^2}\left(q\frac{\partial L}{\partial q} + \frac{\partial^2 L}{\partial\theta^2}\right) - \left(\frac{\partial L}{\partial\theta} - q\frac{\partial^2 L}{\partial q\,\partial\theta}\right)^2\right]^{-1}. \qquad (3.12)$$

The general solution of (3.11) is found to be

$$L(q,\theta) = q\phi(\theta) + \chi(q), \qquad (3.13)$$

where $\phi(\theta)$ and $\chi(q)$ are arbitrary functions of their arguments. From (3.13) and (3.10), we have

$$\frac{\partial}{\partial\theta}\left[\frac{\phi(\theta) + \phi''(\theta) + q\chi''(q) + \chi'(q)}{\chi(q)\{\phi(\theta) + \phi''(\theta) + \chi'(q)\}}\right]\chi''(q) = 0,$$
$$\chi''(q)\neq 0. \qquad (3.14)$$

Simplifying this equation, we find that $\phi(\theta)$ satisfies

$$\phi'(\theta) + \phi''(\theta) = 0. \qquad (3.15)$$

The general solution of (3.14) may be shown to be

$$\phi(\theta) = c_1 + c_2\cos\theta + c_3\sin\theta, \qquad (3.16)$$

where c_1, c_2, and c_3 are the arbitrary constants.

Employing (3.15) in (3.14), the general solution of the system of equations (3.10) and (3.11) is given by

$$L(q,\theta) = q(c_1 + c_2\cos\theta + c_3\sin\theta) + \chi(q). \qquad (3.17)$$

Using this solution in (3.7), in (q,θ) coordinates, we have

$$x - c_3 = [c_1 + \chi'(q)]\sin\theta,$$
$$y + c_2 = -[c_1 + \chi'(q)]\cos\theta. \qquad (3.18)$$

Squaring and adding, we obtain

$$(x-c_3)^2 + (y+c_2)^2 = [c_1 + \chi'(q)]^2. \qquad (3.19)$$

From (3.19), we observe that curves along the velocity magnitude are constant, i.e., *the streamlines are concentric circles with center* $(c_3, -c_2)$.

Defining the distance of any point (x, y) from the center $(c_3, -c_2)$ by R, we obtain from (3.19) that $R = c_1 + \chi'(q)$. Differentiating this result with respect to q, we obtain $dR/dq = \chi''(q)$. From (3.14) it is required that $\chi''(q) \neq 0$, and therefore we have the previous result

$$q = q(R).$$

Using this result in (3.18), the velocity field is given by

$$u = -\frac{(y+c_2)q(R)}{\chi'[q(R)]+c_1} = -\frac{(y+c_2)q(R)}{R},$$
$$v = \frac{(x-c_3)q(R)}{\chi'[q(R)]+c_1} = \frac{(x-c_3)q(R)}{R}. \quad (3.20)$$

B. Case (2) ($J = 0$)

The Jacobian being zero implies that

$$v = \phi(u), \quad (3.21)$$

where ϕ is an arbitrary function of u. Using (3.21) in (3.30) we find that $u(x,y)$ must satisfy

$$[u + \phi(u)\phi'(v)]\left[u\frac{\partial u}{\partial x} + \phi(u)\frac{\partial u}{\partial y}\right] = 0. \quad (3.22)$$

From the above it follows that either the first term or the second term or both the terms are zero. This in conjunction with (3.1) and (3.2), gives rise to systems of equations as

$$u\frac{\partial u}{\partial x} + \phi(u)\frac{\partial u}{\partial y} = 0,$$
$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \quad (3.23)$$
$$u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} = 0,$$

$$u + \phi(u)\phi'(u) = 0,$$
$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \quad (3.24)$$
$$u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} = 0.$$

For a flow field given by $u(x,y)$, $v(x,y) = \phi(u)$ any one of the above systems of equation must be satisfied.

Using the first two equations of the system of equations (3.23), we observe that

$$[u\phi'(u) - \phi(u)]\frac{\partial u}{\partial y} = 0. \quad (3.25)$$

Equation (3.26) is satisfied either by putting

$$\frac{\partial u}{\partial y} = 0, \quad \text{or} \quad [u\phi'(u) - \phi(u)] = 0.$$

The first possibility leads to $\partial u/\partial x = 0$, and therefore $u = u_0$ and $v = \phi(u_0) = v_0$, where u_0 and v_0 are two arbitrary real numbers. *The flow is irrotational and the streamlines are parallel straight lines.* On integrating, the second possiblity leads to $\phi(u) = k_1 u$. Substituting this result in the second equation of the system, we find that $u(x,y)$ satisfied

$$\frac{\partial u}{\partial x} + k_1\frac{\partial u}{\partial y} = 0,$$

and therefore

$$u(x,y) = f(k_1 x - y),$$

and (3.26)

$$v(x,y) = k_1 u = k_1 f(k_1 x - y),$$

where f is an arbitrary function. Using the above result, the vorticity w is given by

$$w = (k_1^2 + 1)f'(k_1 x - y). \quad (3.27)$$

The third equation of the system is identically satisfied by u, v, and w.

Considering the system of equations (3.24), and integrating the first equation with respect to u, we obtain

$$u^2 + v^2 = k_2, \quad (3.28)$$

where $k_2 > 0$ is an arbitrary constant.

Introducing the stream function $\psi(x,y)$, which is given by $d\psi = u\,dy - v\,dx$, the second equation of this system is identically satisfied and (3.28) becomes

$$\left(\frac{\partial\psi}{\partial y}\right)^2 + \left(\frac{\partial\psi}{\partial x}\right)^2 = k_2. \quad (3.29)$$

The complete integral of (3.29) is given by

$$\psi(x,y) = \sqrt{(k_2)}\,[x\cos\alpha + y\sin\alpha + \beta], \quad (3.30)$$

where α and β are two arbitrary parameters.

The general integral of (3.29) is obtained by taking $\beta = \beta(\alpha)$ in the complete integral and eliminating α between

$$\psi = \sqrt{k_2}[x\cos + y\sin\alpha + \beta(\alpha)], \quad (3.31a)$$
$$0 = -x\sin\alpha + y\cos\alpha + \beta'(\alpha). \quad (3.31b)$$

Differentiating (3.31a) with respect to x and y and using (3.31b), we get

$$\frac{\partial\psi}{\partial x} = p = \sqrt{k_2}\cos\alpha, \quad \frac{\partial\psi}{\partial y} = q = \sqrt{k_2}\sin\alpha. \quad (3.32)$$

Again differentiating (3.32) with respect to x and y, we have

$$\frac{\partial^2\psi}{\partial x^2} = r = -\sqrt{k_2}\sin\frac{\partial\alpha}{\partial x},$$
$$\frac{\partial^2\psi}{\partial y^2} = t = \sqrt{k_2}\cos\alpha\frac{\partial\alpha}{\partial y}, \quad (3.33)$$

$$\frac{\partial^2\psi}{\partial x\,\partial y} = s = -\sqrt{k_2}\sin\alpha\frac{\partial\alpha}{\partial y} = \sqrt{k_2}\cos\alpha\frac{\partial\alpha}{\partial x}, \quad (3.34)$$

Using (3.32)–(3.34) in (3.2), we obtain

$$\sin\alpha\frac{\partial}{\partial\alpha}\left[-\sin\alpha\frac{\partial\alpha}{\partial x} + \cos\alpha\frac{\partial\alpha}{\partial y}\right]$$
$$= \cos\alpha\frac{\partial}{\partial y}\left[-\sin\alpha\frac{\partial\alpha}{\partial x} + \cos\alpha\frac{\partial\alpha}{\partial y}\right].$$

Employing (3.34) in this relation, we obtain

$$\frac{\partial^2\alpha}{\partial x^2} + \frac{\partial^2\alpha}{\partial y^2} = 0. \quad (3.35)$$

Differentiating (3.31b) twice with respect to x and twice with respect to y and then adding, we get

$$\{\beta'(\alpha) + \beta'''(\alpha)\}\left\{\left(\frac{\partial\alpha}{\partial x}\right)^2 + \left(\frac{\partial\alpha}{\partial y}\right)^2\right\} = 0. \quad (3.36)$$

Equation (3.36) implies either $(\partial\alpha/\partial x)^2 + (\partial\alpha/\partial y)^2 = 0$, or $\beta'(\alpha) + \beta'''(\alpha) = 0$.

The first possibility gives $\alpha = \text{const}$ and, therefore, *the streamlines are a family of parallel straight lines.* The second possibility gives

$$\beta(\alpha) = C_4 + C_5 \sin\alpha - C_6 \cos\alpha,$$

where C_4, C_5, and C_6 are arbitrary constants. Using this expression for $\beta(\alpha)$ in (3.31), we have

$$(1/\sqrt{k_2})\psi - c_1 = (x - c_3)\cos\alpha + (y + c_2)\sin\alpha,$$
$$0 = -(x - c_3)\sin\alpha + (y - c_2)\cos\alpha.$$

Squaring and adding, we eliminate and obtain

$$((1/\sqrt{k_2})\psi - c_1) = \pm\sqrt{\{(x - c_3)^2 + (y - c_2)^2\}},$$

which implies that *the streamlines are concentric circles* with

$$u = \pm\frac{\sqrt{k_2(y + c_2)}}{\sqrt{\{(x - c_3)^2 + (y + c_2)^2\}}},$$
$$v = \mp\frac{\sqrt{k_2(x - c_3)}}{\sqrt{\{(x - c_3)^2 + (y + c_2)^2\}}}. \qquad (3.37)$$

C. Case (3) ($J = 0$ in part of the region and $J \neq 0$ in the remaining part of the flow)

From the above analysis, it is apparent that the common streamline pattern for the cases $J = 0$ and $J \neq 0$ are concentric circles. If we restrict our attention to any one streamline, parts of which lie in each region, we find from the solutions (3.20) and (3.27) that we encounter discontinuity in the velocity components as we cross from one region to another. Therefore, such flows cannot exist.

IV. ANALYSIS FOR CIRCULATION-PRESERVING FLOWS

More generally, we may consider the problem of universal motions of an incompressible viscous hydromagnetic fluids. Motion of steady incompressible viscous hydromagnetic fluids are given by

$$\text{curl}(\mathbf{w}\times\mathbf{v}) = \nu\,\text{curl curl}\,\mathbf{w},$$

where ν is the kinematic viscosity. Universal motions are motions in which the velocity field is same for all hydromagnetic fluids. It is, therefore, independent of viscosity. At the same time, the stress producing the motion may depend on the viscosity and hence the universal MFD flows are determined by the condition

$$\text{curl}(\mathbf{w}\times\mathbf{v}) = \mathbf{0}, \qquad (4.1)$$
$$\text{curl curl}\,\mathbf{w} = 0. \qquad (4.2)$$

Now we establish following result.

Theorem 2: The circulation preserving hydromagnetic motions of steady vorticity are (i) motions with constant vorticity on which may be superposed a coplanar isochoric irrotational motion, which may be unsteady, and (ii) motions whose streamlines are concentric circles, for which the velocity magnitude is

$$v = Ar\log r + Br + (C/r),$$

where A, B, and C are constants.

To prove this proposition we let $\mathbf{s}$ and $\mathbf{n}$ be the tangent and normal to the streamlines. The constant unit vector perpendicular to the plane of the motion is denoted by $\mathbf{b}$. The velocity is then given by

$$\mathbf{v} = v\mathbf{s}.$$

Taking the curl of the velocity vector, we have

$$\mathbf{w} = \text{curl}\,\mathbf{v} = w\mathbf{b}$$
$$= \left(k_s v - \frac{\delta v}{\delta n}\right)\mathbf{b}. \qquad (4.3)$$

From the equation of continuity we obtain

$$\frac{\delta v}{\delta s} + k_n v = 0, \qquad (4.4)$$

where the scalars k_s and k_n are identified as the curvatures of the vector lines of $\mathbf{s}$ and $\mathbf{n}$, respectively. Now from (4.3) we have

$$\text{curl}(\mathbf{w}\times\mathbf{v}) = \text{grad}(vw)\times\mathbf{n} + \mathbf{v}\times\text{curl}\,\mathbf{n}$$
$$= v\frac{\delta w}{\delta s}\mathbf{b} = 0. \qquad (4.5)$$

By (3.1), we obtain

$$\text{curl}\,\mathbf{w} = \frac{\delta w}{\delta n}\mathbf{s} - \frac{\delta w}{\delta s}\mathbf{n}. \qquad (4.6)$$

From (4.5) and (4.6), we observe that

$$\text{curl}\,\mathbf{w} = \frac{\delta w}{\delta n}\mathbf{s}. \qquad (4.7)$$

The vector lines of curl $\mathbf{w}$ are now streamlines. The condition (4.2) now gives

$$\frac{\delta^2 w}{\delta n^2} - k_s\frac{\delta w}{\delta n} = 0. \qquad (4.8)$$

Again since div curl $\mathbf{w} = 0$, we have

$$\frac{\delta^2 w}{\delta s\,\delta n} + k_n\frac{\delta w}{\delta n} = 0. \qquad (4.9)$$

Since $\delta w/\delta n$ is nonvanishing we may write (4.8) and (4.9) in the forms

$$k_s = \frac{\delta h}{\delta n} \quad\text{and}\quad k_n = -\frac{\delta h}{\delta s}, \qquad (4.10)$$

where

$$h = \log\left|\frac{\delta w}{\delta n}\right|.$$

Applying the commutation formulas on (4.10), we obtain

$$\frac{\delta k_s}{\delta s} + \frac{\delta k_n}{\delta n} = 0. \qquad (4.11)$$

By (4.3)–(4.5), we have

$$k_n = 0, \qquad (4.12)$$

so that, for rotational motion, $k_n = 0$ and the vector lines of $\mathbf{n}$ are straight lines. By (4.11), $\delta k_s/\delta s = 0$, *so the vector lines of* $\mathbf{s}$ *are concentric circles.* Here we write $k_s = 1/r$. By (4.5) and (4.6), v and w are functions of r only. Equation (4.5) becomes

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial w}{\partial r}\right) = 0, \qquad (4.13)$$

where, by (4.4)

$$w = \frac{1}{r}\frac{\partial}{\partial r}(rv). \tag{4.14}$$

By direct integration of (4.13) and (4.14), we have

$$v = Ar \log r + Br + c/r.$$

This proves Theorem 2.

V. ANALYSIS FOR SONIC FLOWS

Taking the dot product of the linear momentum, Eq. (2.8) with unit tangent vector **t** to the streamline, we have

$$\mathbf{t}\cdot\nabla(\tfrac{1}{2}v^2) + \mathbf{t}\cdot\nabla(p + \tfrac{1}{2}\mu H^2) = 0. \tag{5.1}$$

Using the state equation, adiabatic condition, and (2.11) in (5.1), we have

$$\mathbf{t}\cdot\nabla\left(\frac{1}{2}v^2\right) + \frac{1}{\rho}\frac{\partial p}{\partial \rho}\,t\cdot\nabla\rho = 0. \tag{5.2}$$

Using the assumption that the flows are sonic, that is $v^2 = c^2 = \partial p/\partial \rho$, and employing the adiabatic condition again, we find

$$\left(\frac{\partial^2 p}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial p}{\partial \rho}\right)\mathbf{t}\cdot\nabla\rho = 0. \tag{5.3}$$

Since $(\partial^2 p/\partial \rho^2) + (1/\rho)(\partial p/\partial \rho)$ is strictly positive, it follows that $\mathbf{t}\cdot\nabla\rho = 0$, and, therefore, $\mathbf{t}\cdot\nabla(\frac{1}{2}v^2) = 0$ from (5.2). Therefore, for sonic flows, the velocity magnitude is constant on each individual streamline and the results of Sec. III hold for sonic flows as well.

ACKNOWLEDGMENTS

The authors are grateful to Professor K. Lal for suggestions.

One of the authors (K.R.C.) is also thankful to the Council of Scientific and Industrial Research, New Delhi, for financial provision.

[1]R. Berker, "Integration des equations du mouvement d'un fluid visqueux incompressible," in *Handbuck der Physik*, Vol. VIII/2 (Springer, Berlin, 1963), pp. 341 and 346.

[2]R. Prim and P. Nemenyi, J. Math. Phys. **27**, 130 (1948).

[3]R. Prim, J. Rat. Mech. Anal. **1952**, 425.

[4]A. M. H. Martin, Quart. Appl. Math. **8**, 137 (1959).

[5]M. A. Ansari and R. Babu, Bull. Math. Soc. Sci. Math. R. S. Roumanie **25**, 339 (1981).

[6]W. F. Ames, *Non-linear Partial Differential Equations in Engineering* (Academic, New York, 1965).

[7]J. Kampé De Feriet, *4th International Congress of Mathematicians*, Zurich, 1932, Vol. 2 (Kraus Reprint Ltd., Nendeln/Liechtensten, 1967), p. 298.

[8]S. N. Singh and H. P. Singh, Acta Mech. **54**, 181 (1985).

[9]K. R. Choubey and S. N. Singh, Nuovo Cimento, in press.

[10]H. Gortler and K. Weighardt, Math. Z. **48**, 247 (1942).

[11]S. Lundquist, Ark. Fys. **5**, 297 (1952).

[12]C. Truesdell, *Kinematics of Vorticity* (Indiana U. P., Bloomington, 1954).

The 32 minimal general generator sets of 230 double space groups

Shoon K. Kim
Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122

(Received 24 July 1985; accepted for publication 11 December 1985)

The 32 minimal general generator sets (MGGS) for the three-dimensional double space groups are provided; each MGGS has a very limited number of parameters that describe the translational parts of the space groups belonging to a crystal class. The method of construction is based on the algebraic defining relations of the abstract generators of the point groups. The equivalence or inequivalence criteria for the space groups with respect to lattice transformations are established by introducing the definite set $\{U\}$ of the unimodular matrices that leaves a lattice type L of each crystal class G. Based on the set $\{U\}$, it will be shown that mere shifts of the lattice origin are necessary and sufficient to determine the inequivalence of almost all the space groups belonging to a crystal class of high symmetry. It is simpler to construct MGGS of higher symmetry than those of lower symmetry; this contrasts with the existing methods, which are based on the solvability of the space groups.

I. INTRODUCTION

Recently[1–4] there has been renewed interest in the structure of the space groups in order to determine the broken-symmetry (isotropy) groups in connection with the Landau[5] theory of the structural phase transition of the crystalline solids. In generating the isotropy subgroups it is often the most difficult task to identify them from the standard tables[6] of the space groups. In this regard, it is highly desirable to have a compact table of the space groups that clearly exhibits their symmetry properties with respect to the transformation of the lattice basis. Motivated by the renewed interests, we shall reconstruct the space groups such that it will lead to 32 minimal general generator sets (MGGS) for the three-dimensional space groups; each MGGS has a set of parameters that describes the translational parts of the space groups belonging to a crystal class.

An example to MGGS seems in order: It will be shown that eight space groups belonging to the crystal class O are described by two generators with one parameter c,

$$(4_z|0,-c,c),\quad (3_{xyz}|0,0,0), \tag{1.1}$$

which are expressed based on the conventional lattice vectors and using obvious notations (see Sec. III). Here, the parameter c takes $0, \frac{1}{2}, \pm\frac{1}{4}$ for the primitive lattice while taking $0,\frac{1}{4}$ for the face-centered and body-centered lattices.

There are many convenient features for MGGS of the three-dimensional crystal classes.

(1) The number of the generators for every crystal class (or the rank) is less than or equal to 3.

(2) The algebraic defining relations for the minimum number of generators are written down directly from the conditions of regular polygons [see (3.5)]; there is no need to use the solvability of the space groups.

(3) The MGGS for a crystal class provides all the necessary and sufficient information for determining all the (vector or projective) irreducible representations of the space groups belonging to the class.[7]

(4) The number of the translational parameters required for each class is limited (<5); e.g., only one parameter is required for the class O [see (1.1)].

(5) The symmetry properties exhibited by MGGS under lattice transformations help identify the space groups.

Above all, the present construction of 32 MGGS provides so far the simplest algebraic means of arriving at the 230 space groups without the help of a computer. This is partly due to the algebraic equivalence criteria for the space groups completed in this work. Furthermore, MGGS of the crystal classes lead to similar compact expressions for the extended space groups such as magnetic space groups.[8] This gives us control over the large number (a totoal of 1421) and helps us to understand the group structure.

The present method of constructing MGGS is based on the defining relations of the abstract generators of the point groups (or the presentations of the point group[9]). This is possible since the factor group $\hat{G}/T$ of a space group $\hat{G}$ with respect to the translation group T is isomorphic to a point of group G. It is similar to the method due to Zassenhaus[10] and his collaborators,[11,12] which provides a computational algorithm for the determination of all n-dimensional space groups; a program carried out successfully for $n = 4$. Their method is, however, based on the arithmetic classes (73 of them for the three-dimensional space groups). As a result, it is not suitable to construct MGGS for the 32 (geometrical) crystal classes. Their method is good for computer calculation for any dimension, while the present method is good for the back of envelope type hand calculation for two or three dimensions. The reason is as follows: In their method, the required unimodular matrix U, which connects two equivalent space groups, must be found by trial and error,[11] while in the present work we provide a definite set $\{U\}$ for each crystal class to begin with [see (2.7) and (2.8)]. The set is simply given by the symmetry group, which leaves invariant a Bravais lattice type L of the class G. The lattice type L defines the transformation group T with further specification of the lattice basis. With use of the definite set of the unimodular transformations $\{U\}$ one can establish definitely the equivalence or inequivalence of any set of space groups belonging to $\{L:G\}$; without such a set one may need an algorithm to exhaust all the possibilities to establish the inequivalence. It is somewhat surprising to see that such a set $\{U\}$ has

 0022-2488/86/051471-13$02.50

never been reported previously. A probable cause could be the use of the arithmetic classes, which are not very convenient for simple geometrical consideration.

For the three-dimensional space groups, there also exist well-known algebraic methods introduced by Seitz[13] based on the geometrical crystal classes and by Burchkhardt[14] based on the arithmetic classes. Both of them use the fact that the space groups are solvable groups. They closely resemble the geometric method of Schoenflies[15] in that groups of high symmetry are developed in a step by step manner from the cyclic groups. Consequently, their methods also are not suitable for the present purpose of determining MGGS for the crystal classes. Besides, they did not introduce the definite sets of the unimodular matrices $\{U\}$, which determine the equivalence or inequivalence of the space groups. Seitz specifically stated in his work[13] that his treatment does not prove rigorously that there are 230 space groups, but only that there are not more than 230. A similar defect is observed for the work by Bertaut[16] who construct the space groups of the class O_h based on the defining relations of the generators.

The plan of the present work is as follows. The basic development of the present approach already outlined above will be given in Sec. II. The equivalence criteria of the solutions will be discussed there in detail. In Sec. III, some preparatory remarks and the notations are introduced for the point operations. The presentations of the crystal classes are also introduced there. These will make this paper as self-contained as possible. In Sec. IV, all 32 MGGS for the 230 double space groups are constructed starting from the cubic system, which exemplifies the present approach the most effectively. The final results of MGGS are summarized in Table I with the translational parameters that characterize each space group.

II. BASIC DEVELOPMENT AND EQUIVALENCE CRITERIA

For the sake of generality and later convenience for further development we shall base our argument on the double space groups from the outset. This does not require much more effort than for the ordinary space groups, since the extension of a space group to its double groups affects only the rotational part. Hereafter, we mean by a point group or a space group the respective double group unless otherwise specified.

Let $\hat{G}$ be a space group, T be its translation group, and G be a point group isomorphic to the factor group $\hat{G}/T$:

$$G \simeq \hat{G}/T. \tag{2.1}$$

Using the Seitz notation, let $\{R|t_R\}$ be an element of $\hat{G}$, where R is a point operation and t_R is a translation associated with R. Let $(R|v_R)$ be a coset representative of $\hat{G}$ with respect to T, where v_R conveniently may be chosen to be the minimum t_R. Then the set $\{(R|v_R)\}$ is a faithful representation of $\hat{G}/T$ with the multiplication law,

$$\begin{aligned}&(R_1|v_1)(R_2|v_2) = (R_3|v_3),\\ &R_1R_2 = R_3, \quad v_1 + R_1v_2 = v_3 \quad (\text{mod } t\in T).\end{aligned} \tag{2.2}$$

Hereafter, we mean by $\hat{G}/T$ the group $\{(R|v_R)\}$ with this multiplication law.

Now, on account of the isomorphism (2.1), both generator sets of G and $\hat{G}/T$ satisfy the same abstract defining relations, which may be written formally as follows:

$$\prod X_i^{N_i} = \bar{E}, \quad \prod Y_j^{M_j} = E, \quad \bar{E}^2 = E, \tag{2.3}$$

where the $\{N_i\}$'s and the $\{M_j\}$'s are certain integer sets that will be specified later for each G [see (3.5)]. Here, X_i, Y_j are abstract generators. The operators $\bar{E}$ and E for G are given by the 2π rotation $\bar{e}$ and the identity e, respectively, while for $\hat{G}/T$ they are given by

$$\bar{E} = (\bar{e}|0), \quad E = (e|0), \tag{2.4}$$

where 0 is the null vector. Obviously, the presentation of $\hat{G}/T$ is common to all those belonging to a given crystal class, which is also denoted by G.

Now, we are ready to describe the present method of constructing the space groups. Assume a set of realizations for the generators of $\hat{G}/T$ in terms of $(R|v_R)$ with undetermined v_R. Substitution of these into the defining relations (2.3) leads to a set of linear equations of v_R's (mod $t\in T$). Let $\{L;G\}$ be a lattice type L of the crystal class G, which defines the translation group T. Then the solutions of the linear equations yield the space groups belonging to each lattice type $\{L;G\}$. In general there are more than one solution because of mod $t\in T$.

It frequently happens, however, that some of the space groups thus obtained for a given $\{L;G\}$ may be transformed into one another and hence equivalent. Such a transformation may be described by a lattice transformation $\Lambda = [U|s]$, which is a linear transformation U of the lattice basis followed by a shift s of the lattice origin (mod $t\in T$). Let the coordinate system be based on the lattice vectors, then the transformation U is described by a unimodular matrix,[11] which is a matrix with integral elements and det $U = \pm 1$. When the space group has enantiomorphism one should take U as a proper unimodular matrix, i.e., det $U = 1$. Further restriction on U will be discussed later.

Let $\{(R_j|v_j)\}$ and $\{(R_j|u_j)\}$ be a pair of two solutions obtained from the presentation of $\hat{G}/T$ for a given $\{L;G\}$. Then, the pair is equivalent if and only if there exists a lattice transformation of $\Lambda = [U|s]$, which connects the pair through

$$\Lambda^{-1}(R_j|v_j)\Lambda = (R_k|u_k), \tag{2.5}$$

i.e.,

$$\begin{aligned}&(e - R_j)s = v_j - Uu_k (\text{mod } t\in T),\\ &R_k = U^{-1}R_jU\in G,\end{aligned} \tag{2.6}$$

for all $R_j\in G$. Let $R_1,R_2,...,R_r$ be a minimal set of the generators of G. Then from the group property (2.2), only r equations of (2.6) associated with the generators are independent. Since $r\leqslant 3$, the set of equations has more unknowns than the number of equations. Yet, it may not have a solution on account of the condition imposed on U. Thus, we may restate the problem in a more convenient form for calculation: The pair is equivalent if and only if the set of r equations has a solution s for any (proper) unimodular matrix U, which defines an automorphism of G and leaves the lattice type L invariant.

The solvability of the set of equations (2.6) for a given U and $t \in T$ is trivial. If one of the operators R_j is a pure n-fold rotation n or a reflection $\bar{2}$, it has at least one eigenvector Ψ_j belonging to the characteristic value 1. Then the jth equation has a solution s_j (up to Ψ_j) if and only if the right-hand side does not contain a vector component in the space of Ψ_j. On the other hand, if R_j is the inversion i or a rotatory inversion $\bar{n}$ $(n>2)$, it does not have the characteristic value 1 and the corresponding equation has a definite solution s_j for any U. Now the set (2.6) has a solution for the given U and $t \in T$ if all s_j exist and are independent of j.

The crucial part of the problem is to find the allowed set of the unimodular matrices $\{U\}$, unless we resort to trial and error. The collection of all 3×3 unimodular matrices forms a group, which we denote by GL(3,Z). Then the allowed set $\{U\}$ is a normalizer $N(G)$ of the class G in GL(3,Z), which leaves a lattice type L of G invariant. Thus, from the knowledge of the symmetry group of a crystal system that leaves a crystal lattice T invariant, we obtain the following normalizer $N(G)$ $(\rhd G)$ for the required sets $\{U\}$ (ignoring enantiomorphism for the moment):

$$\begin{aligned}
&O_h \rhd O, T_d, T_h, T, D_{2h}, D_2,\\
&D_{6h} \rhd D_6, C_{6v}, D_{3h}, C_{6h}, C_6, C_{3h},\\
&D_{3d} \rhd D_3, C_{3v}, C_{3i} C_3,\\
&D_{4h} \rhd D_4, C_{4v}, D_{2d}, C_{4h}, C_4, S_4, C_{2v},\\
&M \rhd C_{2h}, C_2, C_s,\\
&TR \rhd C_i, C_1.
\end{aligned} \tag{2.7}$$

When there exists enantiomorphism, the normalizers given above should be replaced by their proper subgroups O and D_n $(n>2)$, respectively. The sets M and TR will be defined later.

It should be noted that the normalizer D_{3d} given above for the rhombohedral system is for the rhombohedral lattice. When the lattice is hexagonal, the normalizer for D_{3d}, D_3, C_{3v}, and C_3 is D_{6h}. This does not create any complication in the actual construction of their space groups, since the inequivalences of the allowed solutions for these classes are simply due to the difference in the translational vector component in the invariant eigenvector space of 3_z, the z component (see Sec. IV C). The sets M and TR are infinite groups, which will be given later by their generators. Note that the classes of the orthorhombic system are included in the cubic or tetragonal system. A justification of the above sets given by (2,7) may be seen, for example, from the first set O_h; it describes the group of all possible permutations of the set $(\pm x, \pm y, \pm z)$ (based on the conventional lattice vectors), which accounts for the difference between a lattice type L and an actual lattice T belonging to $(L;G)$. More specifically, the three binary axes of rotation of D_{2h} (or D_2) are regarded as equivalent under their permutations for any of its lattice type in defining the set $\{U\}$ of unimodular matrices. This is due to the fact that equivalence (2.5) under a lattice transformation Λ is defined for the normalized translational parts of the elements $\{(R_j|v_j)\} \in G$ via the (conventional) lattice basis. [For further details, see their symmetry properties described by (4.113a) and (4.113b) in Sec. IV.] This is consistent with the traditional geometric congruence in classification of their space groups. The set $\{U\}$ given by (2.7) may be referred to as the symmetry groups of the crystal type systems.

In the actual calculation, it is sufficient to consider only the coset representatives of G in $N(G)$ for the matrix U in (2.6), since inner automorphisms of $\hat{G}$ do not affect $\hat{G}$. Moreover, when there is no enantiomorphism, a proper operator n can be replaced by an improper operator $\bar{n}$ and vice versa. Consequently, the relevant set of the unimodular transformation $\{U\}$ for every crystal class may be given by one of the following generator sets, up to a multiplicative factor of an element of the respective class:

$$\begin{aligned}
&\{e\};\ O_h, O, T_d, D_{6h}, D_6, C_{6v}, D_{3h}, D_{3d}, D_3, C_{3v},\\
&\qquad D_{4h}, D_4, C_{4v}, D_{2d},\\
&\{2'\};\ C_{6h}, C_6, C_{3h}, C_{3i}, C_3, C_{4h}, C_4, S_4,\\
&\{4_z\};\ T_h, T, C_{2v},\\
&\{(x,y),(x,z)\};\ D_{2h}, D_2,\\
&\{(x,y),(x \pm y, y, z)\};\ C_{2h}, C_2, C_s,\\
&\{(x,y),(x,z),(x \pm y, y, z)\};\ C_i, C_1.
\end{aligned} \tag{2.8}$$

Here $2' \perp n$, (x,y) is the interchange of the x,y axes, and $(x \pm y, y, z)$ is a linear transformation written in terms of the Jones notation. The coordinate systems are based on the conventional lattice vectors. The last set for TR is academic and has no practical use in the present work.

A striking aspect of the relevant set $\{U\}$ given by (2.8) is that it is simpler for a crystal class with higher symmetry. In particular, mere shifts $[e|s]$ of the lattice origin will be sufficient to determine the equivalence or inequivalence for almost all classes of high symmetry. Consequently, it is simpler to construct MGGS of the classes with higher symmetry for the present method (this is quite a contrast to the existing methods based on the solvability of the space groups). These conclusions are borne out by the actual calculation given in Sec. IV. For example, one sees immediately the inequivalence of the space groups given in (1.1), since the translational parts of those belonging to each lattice type are all different in the invariant eigenvector space 4_z, the z component. In a special case where one of the pair is the null solution [say $u_k = 0$ for all k in (2.6)], then mere shifts are again sufficient to establish equivalence or inequivalence for all cases. Another interesting use of the equivalence criteria is as follows: For the classes $D_{2h}, D_{3h}, D_3, C_{3v}, D_{3d}$, there exist two types of inequivalent realizations of $\hat{G}/T$ depending on the mutual orientation of G and the lattice type L. Their inequivalences are also treated likewise with the criteria given by (2.6) and (2.8).

The relevant sets of the unimodular matrices $\{U\}$ given by (2.8) greatly simplify the process of removing the redundant solutions as well as ascertaining the inequivalence of the final results to arrive at 230 space groups. In the actual construction given in Sec. IV, we shall first reduce the number of the translational parameters for each class as much as possible by a shift of the lattice origin. Then we obtain possible space group solutions from the presentation of $\hat{G}/T$. By studying the symmetry properties of the solutions with respect to the relevant lattice transformations $[U|s]$, where U is determined from (2.8), we remove the redundant solutions to arrive at the final independent solutions. The equivalence criteria (2.6) supplemented by (2.8) must correspond to the geometric congruence criteria used by the early workers. A similar relevant set of unimodular matrices for identifying the isotropy groups will be discussed in a forthcoming paper.

When two space groups $\{(R_j|v_j)\}$ and $\{(R_j|u_j)\}$ are equivalent under Λ we frequently say simply v_j and u_j are equivalent under Λ and write

$$v_j \sim u_j \quad \text{under } \Lambda. \tag{2.9}$$

III. THE NOTATIONS AND THE DEFINING RELATIONS

A. The notations

As in the International Tables, we shall use the following notations for the lattice type L: primitive (P), base centered (A,B,C) face centered (F), body centered (I), and rhombohedral (R). When the lattice R is regarded as the double-centered hexagonal lattice, it is denoted by R^*. The coordinate systems will be based on the conventional lattice vectors (the rhombohedral coordinates for the R lattice and the hexagonal coordinates for the R^* lattice).

In determining the translational parameters of any space group operator $(R|a,b,c)$, we frequently use the following simplification for the nonprimitive lattices. Suppose that one of the parameters, say a, is known to be binary, i.e.,

$$a = 0,\tfrac{1}{2}, \tag{3.1}$$

and the remaining parameters b,c are yet to be determined. Then we can set $a=0$ before further calculation in the case of a B, C, F, or I lattice.

Next we shall introduce the notations for the pure point operations in terms of the Jones faithful representations, which conveniently replace their three-dimensional matrix representations. These are given to the extent that is needed for the multiplications of the group generators and for the lattice transformations Λ of (2.5).

(i) For Cartesian coordinates,

$$\begin{aligned} &2_x = (x,\bar{y},\bar{z}), \quad 2_y = (\bar{x},y,\bar{z}),\\ &2_z = (\bar{x},\bar{y},z), \quad 2_{xy} = (y,x,\bar{z}),\\ &2_{x\bar{y}} = (\bar{y},\bar{x},\bar{z}), \quad 3_{xyz} = (z,x,y), \quad 3_{\bar{x}\bar{y}\bar{z}} = (y,z,x),\\ &4_z = (\bar{y},x,z), \quad i = (\bar{x},\bar{y},\bar{z}), \quad m_x = i2_x = \bar{2}_x,\\ &m_{xy} = i2_{xy} = \bar{2}_{xy}, \quad \bar{n} = \text{in}. \end{aligned} \tag{3.2}$$

(ii) For hexagonal coordinates,

$$\begin{aligned} &2_z = (\bar{x},\bar{y},z), \quad 3_z = (\bar{y},x-y,z), \quad 6_z = (x-y,x,z),\\ &u_0 = (x-y,\bar{y},\bar{z}), \quad u_1 = (x,x-y,\bar{z}),\\ &i = (\bar{x},\bar{y},\bar{z}), \quad m_0 = iu_0, \quad m = iu_1. \end{aligned} \tag{3.3}$$

Here u_ν is a binary rotation in the x,y plane about an axis which makes an angle $\nu\pi/6$ with the x axis.

(iii) For rhombohedral coordinates,

$$3_{xyz} = (z,x,y), \quad u_{x\bar{z}} = (\bar{z},\bar{y},\bar{x}), \quad i = (\bar{x},\bar{y},\bar{z}). \tag{3.4}$$

Here the coordinate system is taken such that the first two operators 3_{xyz} and $u_{x\bar{z}}$ correspond to the operators 3_z and u_0 in the hexagonal coordinates for the R^* lattice, respectively.

B. The defining relations of the crystal classes

A point group G is either isomorphic to a proper point group P or a direct product group $P_i = P \times C_i$, where C_i is the group of inversion. The presentation of P may be given by

$$A^n = B^m = (AB)^l = \bar{E}, \quad \bar{E}^2 = E, \tag{3.5a}$$

where A,B are abstract generators, $\bar{E}$ corresponds to the 2π rotation that commutes with all the elements, and E is the identity. The set of integers $\{n,m,l\}$ is limited to $\{n,0,0\}$ for a cyclic group C_n and $\{n,m,2\}$ for a polyhedral group of a regular Euclidean solid, where n is the number of regular m-gons meeting at any vertex so that $1/n + 1/m > \frac{1}{2}$. Thus we have $\{n,2,2\}$ for D_n, $\{3,3,2\}$ for T, and $\{4,3,2\}$ for O. The presentation of P_i is given by

$$X \in P, \quad I^2 = [X,I] = E, \tag{3.5b}$$

where $[X,I] = XIX^{-1}I^{-1}$ is the commutator. The realizations of $\bar{E}$ and E were discussed before in (2.3) and (2.4). The realization of I for G is the pure inversion i while for $\hat{G}/T$ it is expressed by $(i|v_i)$.

Let us discuss the commutator in some detail. Let $X_1 = (R_1|v_1) \in \hat{P}/T$, then

$$[X_1,I] = (e|2v_1 + R_1 v_i - v_i). \tag{3.6}$$

In a special case when $2v_1$ is a primitive translation of T we have

$$R_1 v_i = v_i \quad (\text{mod } t \in T), \tag{3.7}$$

which characterizes v_i (mod $t \in T$) as an invariant eigenvector of R_1. Such a characterization of v_i for a generator X_1 of $\hat{P}/T$ frequently reduces the labor involved in determining v_i from the commutators with the remaining generators of P.

Based on (2.4) and (3.5), the defining relations of every crystal class will be given in the beginning of the actual calculations of the space groups belonging to the class. To facilitate the realizations of the improper operators for the improper point groups we have denoted the crystal class in terms of the Schoenflies symbols together with the new symbol invented by the author.[17] According to this system of the symbols, the improper point groups corresponding to a proper point group P (with the n-fold principal axis) augmented by i, $2n$, and $2'$ ($2' \perp n$) are expressed by P_i, P_p, and P_v, respectively. This symbol is the most effective in describing their isomorphisms with the proper point groups.

IV. CONSTRUCTION OF 230 SPACE GROUPS

A. The cubic system

The Cartesian coordinates based on the conventional lattice vectors will be used for this system. It is characterized by a threefold axis in the (1,1,1) direction, which may be written as $(3_{xyz}|0)$ for every lattice type adjusting the lattice origin. From (2.8), the relevant generator sets of the unimodular transformations $\{U\}$ are given by $\{4_z\}$ for T, T_h, and by $\{e\}$ for the O, T_d, O_h classes. It turns out, however, that for the class T one needs only shift $[e|s]$ to ascertain the inequivalence of the solutions since there exist only one or two of them for each lattice type of the class.

1. The class T; $A^2 = B^3 = \bar{E}$, $(AB)^3 = E$

Adjusting the lattice origin one may set

$$A = (2_z|a,0,c), \quad B = (3_{xyz}|0), \tag{4.1}$$

where a and c are parameters to be determined for each lattice type from the defining relations and the products

$$A^2 = (\bar{e}|0,0,2c), \quad (AB)^3 = (\bar{e}|a-c,-a+c,-a+c). \tag{4.2}$$

(i) P lattice: The allowed values of the parameters are

$$c = a = 0,\tfrac{1}{2}, \tag{4.3}$$

which define two independent space groups with the generators,

$$(2_z|a,0,a), \quad (3_{xyz}|0); \quad a = 0,\tfrac{1}{2}. \tag{4.4}$$

(ii) F lattice: The above solutions (4.4) are reduced to one solution with the parameter

$$a = 0. \tag{4.5}$$

(iii) I lattice: Directly from (4.2) we obtain $a,c = 0,\frac{1}{2}$. Since however $c \sim c + \frac{1}{2}$ under $[e|\frac{1}{4},\frac{1}{4},\frac{1}{4}]$ we may set $a = c$ to obtain (4.4) again for the I lattice.

The five space groups obtained above for the class T are summarized by Nos. 195–199 in Table I.

2. The class $T_h(=T_i)$; $A,B \in T$, $I^2 = [A,I] = [B,I] = E$

The space groups of this class may be constructed by augmenting those of the class T with the inversion translation I. We may set, using (3.7) for 3_{xyz},

$$A = (2_z|a,0,a), \quad B = (3_{xyz}|0), \quad I = (i|\alpha,\alpha,\alpha). \tag{4.6}$$

Then, from the commutators

$$[A,I] = (e|-2\alpha,-2\alpha,0), \quad [B,I] = (e|0), \tag{4.7}$$

we determine the parameter α for each lattice type. Here, the augmentation does not affect the original parameter a.

(i) P lattice: From (4.4) and (4.7) we have a, $\alpha = 0,\frac{1}{2}$. We first rewrite the generator set (4.6) in the form

$$(2_z|a+\alpha,\alpha,a), \quad (3_{xyz}|0), \quad (i|0), \tag{4.8a}$$

then observe that there exist only three inequivalent sets for the ordered pair (a,α) given by

$$(0,0),\ (0,\tfrac{1}{2}),\ (\tfrac{1}{2},0) \sim (\tfrac{1}{2},\tfrac{1}{2}). \tag{4.8b}$$

The equivalence of the last two pairs follows from

$$(2_z|\tfrac{1}{2},0,\tfrac{1}{2}) \to (2_z|0,\tfrac{1}{2},\tfrac{1}{2}), \quad (3_{xyz}|0) \to (3_{\bar{x}\bar{y}\bar{z}}|0), \tag{4.9}$$

under $[4_z|0]$, which leaves $(i|0)$ invariant.

(ii) F lattice: From (4.5) and (4.7) we obtain

$$a = 0, \quad \alpha = 0,\tfrac{1}{4}, \tag{4.10}$$

which yield two independent solutions.

(iii) I lattice: From the result of the I lattice of the class T and (4.7) we have a, $\alpha = 0,\frac{1}{2}$. Since however $\alpha \sim \alpha + \frac{1}{2}$ for the I lattice, we obtain two independent solutions characterized by

$$a = 0,\tfrac{1}{2}, \quad \alpha = 0. \tag{4.11}$$

The seven space groups obtained above for the class T_h are summarized by Nos. 200–206 in Table I.

3. The class O; $A^4 = B^3 = (AB)^2 = \bar{E}$

Let

$$A = (4_z|0,b,c), \quad B = (3_{xyz}|0). \tag{4.12}$$

Then

$$A^4 = (\bar{e}|0,0,4c), \quad (AB)^2 = (\bar{e}|0,b+c,b+c). \tag{4.13}$$

(i) P lattice: The allowed values of the parameters are

$$-b = c = 0,\tfrac{1}{4},\tfrac{2}{4},\tfrac{3}{4}, \tag{4.14}$$

which yield four independent solutions.

(ii) F and I lattices: Under the shift of the lattice origin by $[e|\frac{1}{4},\frac{1}{4},\frac{1}{4}]$ we have $(4_z|0,b,c) \to (4_z|\frac{1}{2},b,c)$, which yields the equivalences $b \sim b + \frac{1}{2}$, $c \sim c + \frac{1}{2}$ for the F lattice and $(b,c) \sim (b+\frac{1}{2},c+\frac{1}{2})$ for the I lattice. Thus, from (4.13) we arrive at

$$-b = c = 0,\tfrac{1}{4}, \tag{4.15}$$

which provide two independent solutions each for both lattices.

4. The class $T_d(=T_p)$; $A^4 = B^3 = (AB)^2 = \bar{E}$

Let

$$A = (\bar{4}_z|a,0,c), \quad B = (3_{xyz}|0). \tag{4.16}$$

Then

$$A^4 = (\bar{e}|0), \quad (AB)^2 = (\bar{e}|2a,-c,c). \tag{4.17}$$

(i) P lattice: The allowed values of the parameters are

$$a = 0,\tfrac{1}{2}, \quad c = 0, \tag{4.18}$$

which yield two independent solutions of (4.16).

(ii) F lattice: We may set one of the parameters a or c of A equal to zero since both of them are binary from (4.17). Thus we obtain again two independent solutions given by (4.18) for the F lattice.

(iii) I lattice: From (4.17) we obtain

$$a = c = 0 \quad \text{or} \quad 2a = c = \tfrac{1}{2}, \tag{4.19}$$

which give two independent solutions of (4.16).

It is interesting to note that all these six space groups obtained above for T_d can be expressed in the following form with one parameter c:

$$(\bar{4}_z|c,-c,c), \quad (3_{xyz}|0). \tag{4.20}$$

This follows since $(\bar{4}_z|a,0,c) \to (\bar{4}_z|a,-\alpha,c-\alpha)$ under $[e|\alpha/2,\alpha/2,\alpha/2]$. The allowed values of c are $0,\frac{1}{2}$ for P and F lattices and $0,\frac{1}{4}$ for the I lattice (see Nos. 215–220, Table I).

5. The class $O_h(=O_i)$; $A,B \in O$, $I^2 = [A,I] = [B,I] = E$

By augmenting the class O with I we set

$$A = (4_z|0,-c,c), \quad B = (3_{xyz}|0), \quad I = (i|\alpha,\alpha,\alpha). \tag{4.21}$$

Then, the required commutators are

TABLE I. The 32 minimal general generator sets of the 230 space groups.

A. The cubic system

§T; $(2_z|c,0,c)$, $(3_{xyz}|0)$ with $L(c)$:
195. $P(0)$, 196. $F(0)$, 197. $I(0)$, 198. $P(\frac{1}{2})$, 199. $I(\frac{1}{2})$.

§$T_h(=T_i)$; $(2_z|c+a,a,c)$, $(3_{xyz}|0)$, $(i|0)$ with $L(c,a)$:
200. $P(0,0)$, 201. $P(0,\frac{1}{2})$, 202. $F(0,0)$, 203. $F(0,\frac{1}{4})$, 204. $I(0,0)$, 205. $P(\frac{1}{2},0)$, 206. $I(\frac{1}{2},0)$.

§O; $(4_z|0,-c,c)$, $(3_{xyz}|0)$ with $L(c)$:
207. $P(0)$, 208. $P(\frac{1}{2})$, 209. $F(0)$, 210. $F(\frac{1}{4})$, 211. $I(0)$, 212. $P(\frac{3}{4})$, 213. $P(\frac{1}{4})$, 214. $I(\frac{1}{4})$.

§$T_d(=T_p)$; $(\bar{4}_z|c,-c,c)$, $(3_{xyz}|0)$ with $L(c)$:
215. $P(0)$, 216. $F(0)$, 217, $I(0)$, 218. $P(\frac{1}{2})$, 219. $F(\frac{1}{2})$, 220. $I(\frac{1}{4})$.

§$O_h(=O_i)$; $(4_z|-a,-c,c)$, $(3_{xyz}|0)$, $(i|0)$ with $L(a,c)$:
221. $P(0,0)$, 222. $P(\frac{1}{2},0)$, 223. $P(\frac{1}{2},\frac{1}{2})$, 224. $P(0,\frac{1}{2})$, 225. $F(0,0)$, 226. $F(\frac{1}{2},0)$, 227. $F(\frac{1}{2},\frac{1}{4})$, 228. $F(0,\frac{1}{4})$, 229. $I(0,0)$, 230. $I(\frac{1}{4},\frac{1}{4})$.

B. The hexagonal system

§C_6; $P(6_z|0,0,c)$ with c:
168. 0, 169. $\frac{1}{6}$, 170. $\frac{5}{6}$, 171. $\frac{1}{3}$, 172. $\frac{2}{3}$, 173. $\frac{1}{2}$.

§$C_{3h}(=C_{3p})$:
174. $P(\bar{6}_z|0)$.

§$C_{6h}(=C_{6i})$; $P(6_z|0,0,c)$, $(i|0)$ with c:
175. 0, 176. $\frac{1}{2}$.

§D_6; $P(6_z|0,0,c)$, $(u_0|0)$ with c:
177. 0, 178. $\frac{1}{6}$, 179. $\frac{5}{6}$, 180. $\frac{1}{3}$, 181. $\frac{2}{3}$, 182. $\frac{1}{2}$.

§C_{6v}; $P(6_z|0,0,c)$, $(m_0|0,0,c')$ with (c,c'):
183. $(0,0)$, 184. $(0,\frac{1}{2})$, 185. $(\frac{1}{2},\frac{1}{2})$, 186. $(\frac{1}{2},0)$.

§$D_{3h}(=D_{3p})$; $P(\bar{6}_z|0,0,0)$, $(m_0|0,0,c)$ or $(u_0|0,0,c)$ with $(m;c)$ or $(u;c)$:
187. $(m;0)$, 188. $(m;\frac{1}{2})$, 189. $(u;0)$, 190. $(u;\frac{1}{2})$.

§$D_{6h}(=D_{6i})$; $P(6_z|0,0,c)$, $(u_0|0,0,c')$, $(i|0)$ with (c,c'):
191. $(0,0)$, 192. $(0,\frac{1}{2})$, 193. $(\frac{1}{2},\frac{1}{2})$, 194. $(\frac{1}{2},0)$.

C. The rhombohedral system

§C_3; $(3_z|0,0,c)$ with $L(c)$:
143. $P(0)$, 144. $P((\frac{1}{3})$, 145. $P(\frac{2}{3})$, 146. $R^*(0)$.

§C_{3i}; $(3_z|0)$, $(i|0)$ with L:
147. P, 148. R^*.

§D_3; $(3_z|0,0,c)$, $(u_v|0)$ with $L(u_v;c)$:
149. $P(u_1;0)$, 150. $P(u_0;0)$, 151. $P(u_1;\frac{1}{3})$, 152. $P(u_0;\frac{1}{3})$, 153. $P(u_1;\frac{2}{3})$, 154. $P(u_0;\frac{2}{3})$, 155. $R^*(u_0;0)$.

§C_{3v}; $(3_z|0)(m_v|0,0,c)$ with $L(m_v;c)$:
156. $P(m_0;0)$, 157. $P(m_1;0)$, 158. $P(m_0;\frac{1}{2})$, 159. $P(m_1;\frac{1}{2})$, 160. $R^*(m_0;0)$, 161. $R^*(m_0;\frac{1}{2})$.

§$D_{3d}(=D_{3i})$; $(3_z|0)$,$(u_v|0)$,$(i|0,0,c)$ with $L(u_v;c)$:
162. $P(u_1;0)$, 163. $P(u_1;\frac{1}{2})$, 164. $P(u_0;0)$, 165. $P(u_0;\frac{1}{2})$, 166. $R^*_0(u_0;0)$, 167. $R^*_0(u_0;\frac{1}{2})$.

D. The tetragonal system

§C_4; $(4_z|0,0,c)$ with $L(c)$:
75. $P(0)$, 76. $P(\frac{1}{4})$, 77. $P(\frac{1}{2})$, 78. $P(\frac{3}{4})$, 79. $I(0)$, 80. $I(\frac{1}{4})$.

§$S_4(=C_{2p})$; $(\bar{4}_z|0)$ with L:
81. P, 82. I.

§$C_{4h}(=C_{4i})$; $(4_z|a,b,c)(i|0)$ with $L(a,b,c)$:
83. $P(0,0,0)$, 84. $P(0,0,\frac{1}{2})$, 85. $P(\frac{1}{2},0,0)$, 86. $P(\frac{1}{2},0,\frac{1}{2})$, 87. $I(0,0,0)$, 88. $I(\frac{1}{4},\frac{1}{4},\frac{1}{4})$.

§D_4; $(4_z|0,0,c)$, $(2_x|a,a,0)$ with $L(c,a)$:
89. $P(0,0)$, 90. $P(0,\frac{1}{2})$, 91. $(\frac{1}{4},0)$, 92. $P(\frac{1}{4},\frac{1}{2})$, 93. $P(\frac{1}{2},0)$, 94. $P(\frac{1}{2},\frac{1}{2})$, 95. $P(\frac{3}{4},0)$, 96. $P(\frac{3}{4},\frac{1}{2})$, 97. $I(0,0)$, 98. $I(\frac{1}{4},0)$.

§C_{4v}; $(4_z|0,0,c)(m_x|a+2c,a,c')$ with $L(c,a,c')$:

99. $P(0,0,0)$, 100. $P(0,\frac{1}{2},0)$, 101. $P(\frac{1}{2},0,\frac{1}{2})$, 102. $P(\frac{1}{2},\frac{1}{2},\frac{1}{2})$, 103. $P(0,0,\frac{1}{2})$, 104. $P(0,\frac{1}{2},\frac{1}{2})$, 105. $P(\frac{1}{2},0,0)$, 106. $P(\frac{1}{2},\frac{1}{2},0)$, 107. $I(0,0,0)$, 108. $I(0,0,\frac{1}{2})$, 109. $I(\frac{1}{4},0,0)$, 110. $I(\frac{1}{4},\frac{1}{2},0)$.

§$D_{2d}(=D_{2p})$; $(\bar{4}_z|0)(2_x$ or $m_x|a+2c,a,c)$ with $L(2$ or $m;a,c)$:
111. $P(2;0,0)$, 112. $P(2;0,\frac{1}{2})$, 113. $P(2;\frac{1}{2},0)$, 114. $P(2;\frac{1}{2},\frac{1}{2})$, 115. $P(m;0,0)$, 116. $P(m;0,\frac{1}{2})$, 117. $P(m;\frac{1}{2},0)$, 118. $P(m;\frac{1}{2},\frac{1}{2})$, 119. $I(m;0,0)$, 120. $I(m;0,\frac{1}{2})$, 121. $I(2;0,0)$, 122. $I(2;\frac{1}{2},\frac{1}{4})$.

§$D_{4h}(=D_{4i})$; $(4_z|0,0,c)(2_x|a,a,0)(i|\alpha,\alpha+2c,\gamma)$ or $(4_z|\alpha+c,c,c)(2_x|a,a+\alpha+2c,\gamma)(i|0)$ with (c,a) and (α,γ).

L	(c,a) \ (α,γ)	$(0,0)$	$(0,\frac{1}{2})$	$(\frac{1}{2},0)$	$(\frac{1}{2},\frac{1}{2})$
P	$(0,0)$	123	124	125	126
	$(0,\frac{1}{2})$	127	128	129	130
	$(\frac{1}{2},0)$	131	132	133	134
	$(\frac{1}{2},\frac{1}{2})$	135	136	137	138
I	$(0,0)$	139	140		
	$(\frac{1}{4},0)$	142	141		

E. The orthorhombic system

§D_2; $(2_z|c,0,c)(2_x|a,a,0)$ with $L(c,a)$: 16. $P(0,0)$, 17. $P(\frac{1}{2},0)$, 18. $P(0,\frac{1}{2})$, 19. $P(\frac{1}{2},\frac{1}{2})$, 20. $C(\frac{1}{2},0)$, 21. $C(0,0)$, 22. $F(0,0)$, 23. $I(0,0)$, 24. $I(\frac{1}{2},\frac{1}{2})$.

§C_{2v}; $(2_z|a',-b,c+c')(m_x|0,b,c)(m_y|a',0,c')$

L	(b,c) \ (a',c')	$(0,0)$	$(0,\frac{1}{2})$	$(\frac{1}{2},0)$	$(\frac{1}{2},\frac{1}{2})$
P	$(0,0)$	25	26	28	31
	$(0,\frac{1}{2})$		27	29	30
	$(\frac{1}{2},0)$			32	33
	$(\frac{1}{2},\frac{1}{2})$				34
C	$(0,0)$	35	36		
	$(0,\frac{1}{2})$		37		

TABLE I. (*Continued.*)

A	$(0,0)$	38		40
	$(0,\frac{1}{2})$	39		41
I	$(0,0)$	44	46	
	$(0,\frac{1}{2})$		45	
F	$b=c=a'=c'=d$; 42. $d=0$, 43. $d=\frac{1}{4}$			

$\S D_{2h}(=D_{2i})$

$(2_z|c,0,c)(2_x|a,a,0)(i|\alpha,\beta,\gamma)$ or $(2_z|c+\alpha,\beta,c)(2_x|a,a+\beta,\gamma)(i|0)$

L	(c,a)		(α,β,γ)		
P	$(0,0)$	47. $(0,0,0)$	48. $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$	49. $(0,0,\frac{1}{2})$	50. $(\frac{1}{2},\frac{1}{2},0)$
	$(\frac{1}{2},0)$	51. $(\frac{1}{2},0,0)$	52. $(0,\frac{1}{2},0)$	53. $(0,0,0)$	54. $(\frac{1}{2},\frac{1}{2},0)$
	$(0,\frac{1}{2})$	55. $(0,0,0)$	56. $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$	57. $(0,\frac{1}{2},0)$	58. $(0,0,\frac{1}{2})$
		59. $(\frac{1}{2},\frac{1}{2},0)$	60. $(0,\frac{1}{2},\frac{1}{2})$		
	$(\frac{1}{2},\frac{1}{2})$	61. $(0,0,0)$	62. $(0,0,\frac{1}{2})$		
C	$(0,0)$	65. $(0,0,0)$	66. $(0,0,\frac{1}{2})$	67. $(0,\frac{1}{2},0)$	68. $(0,\frac{1}{2},\frac{1}{2})$
	$(\frac{1}{2},0)$	63. $(0,\frac{1}{2},0)$	64. $(0,0,0)$		
F	$(0,0)$	69. $(0,0,0)$	70. $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$		
I	$(0,0)$	71. $(0,0,0)$	72. $(0,0,\frac{1}{2})$		
	$(\frac{1}{2},\frac{1}{2})$	73. $(0,0,0)$	74. $(0,0,\frac{1}{2})$		

F. The monoclinic system

$\S C_2$; $(2_z|0,0,c)$ with $L(c)$: 3. $P(0)$, 4. $P(\frac{1}{2})$, 5. $B(0)$.

$\S C_s(=C_{1p})$; $(m_z|0,b,0)$ with $L(b)$: 6. $P(0)$, 7. $P(\frac{1}{2})$, 8. $B(0)$, 9. $B(\frac{1}{2})$.

$\S C_{2h}(=C_{2i})$; $(2_z|0,b,c)(i|0)$ with $L(b,c)$: 10. $P(0,0)$, 11. $P(0,\frac{1}{2})$, 12. $B(0,0)$, 13. $P(\frac{1}{2},0)$, 14. $P(\frac{1}{2},\frac{1}{2})$, 15. B. $(\frac{1}{2},0)$.

G. The triclinic system

$\S C_1$: 1. $P(\bar{e}|0)$. $\S C_i$: 2. $P(\bar{e}|0)(i|0)$

Notes: (i) $L(a,\ldots)$ means the lattice type L with the parameters $a,\ldots$.
(ii) The number assigned to each space group is in accordance with Ref. 6.

$$[A,I]=(e|-2\alpha,-2c,2c),$$
$$[B,I]=(e|0,0,0), \tag{4.22}$$

from which the parameters $c\in\{0,\ldots,\frac{3}{4}\}$ and α are to be specified further for each lattice type.

(i) P lattice: The allowed values of the parameters are

$$c,\alpha=0,\tfrac{1}{2}, \tag{4.23}$$

which yield four independent solutions of (4.21).

(ii) F lattice: From (4.22) we obtain

$$c=0,\tfrac{1}{4},\quad \alpha=0,\tfrac{1}{2}, \tag{4.24}$$

which provide four independent solutions of (4.21) for the F lattice.

(iii) I lattice: From (4.22) we obtain two sets of the parameters

$$c=\alpha=0,\tfrac{1}{4}, \tag{4.25}$$

which yield two independent solutions of (4.21).

If the lattice origin is chosen at the center of symmetry, the generator set (4.21) is rewritten as follows:

$$(4_z|\alpha,-c,c),\quad (3_{xyz}|0),(i|0). \tag{4.26}$$

B. The hexagonal system

The hexagonal coordinates based on the primitive lattice vectors are used throughout to describe the symmetry elements. As mentioned in Sec. II, there exist two types of realizations of the generators for the class D_{3h}. The construction of the space groups for this system is straightforward since there exists not much redundance except for D_{3h}.

1. The class C_6; $A^6=\bar{E}$

$$A=(6_z|00c),\quad c=0,\tfrac{1}{6},\ldots,\tfrac{5}{6}. \tag{4.27}$$

2. The class $C_{3h}(=C_{3p})$; $A^6=\bar{E}$

One simply sets

$$A=(\bar{6}_z|0), \tag{4.28}$$

since there exists no invariant eigenvector space for $\bar{6}_z$.

3. The class $C_{6h}(=C_{6i})$; $A^6=\bar{E}$, $I^2=[A,I]=E$

Let

$$A=(6_z|0,0,c),\quad I=(i|a,b,0). \tag{4.29}$$

Then

$$[A,I] = (e|-b,a-b,2c). \qquad (4.30)$$

The allowed values of the parameters are

$$a=b=0, \quad c=0,\tfrac{1}{2}, \qquad (4.31)$$

which yield two inequivalent solutions of (4.29).

4. The class D_6; $A^6 = B^2 = (AB)^2 = \bar{E}$

Let

$$A = (6_z|0,0,c), \quad B = (u_0|a,b,0), \qquad (4.32)$$

where u_0 is the diad parallel to the x axis defined in (3.3). From

$$A^6 = (\bar{e}|0,0,6c), \quad B^2 = (\bar{e}|2a-b,0,0),$$
$$(AB)^2 = (\bar{e}|2a-2b,a-b,0), \qquad (4.33)$$

we obtain

$$a=b=0, \quad c=0,\tfrac{1}{6},\ldots,\tfrac{5}{6}, \qquad (4.34)$$

which yield six independent solutions for (4.32).

5. The class C_{6v}; $A^6 = B^2 = (AB)^2 = \bar{E}$

If we replace the diad u_0 of D_6 by the reflection $m_0 = iu_0$, we arrive at C_{6v}. We set

$$A = (6_z|0,0,c), \quad B = (m_0|a',b',c'). \qquad (4.35)$$

Then, from

$$A^6 = (\bar{e}|0,0,6c), \quad B^2 = (\bar{e}|b',2b',2c'),$$
$$(AB)^2 = (\bar{e}|0,a'+b',2c+2c'), \qquad (4.36)$$

we obtain

$$a'=b'=0, \quad c,c'=0,\tfrac{1}{2}, \qquad (4.37)$$

which yield four independent solutions of (4.35).

6. The class $D_{3h}(=D_{3p})$; $A^6 = B^2 = (AB)^2 = \bar{E}$

There exist two possibilities; either the second generator B is the mirror plane m_0 or the diad u_0. Thus we set

$$A = (\bar{6}_z|0),$$
$$B_m = (m_0|a,b,c), \quad B_u = (u_0|a,b,c). \qquad (4.38)$$

For the AB_m type, we have

$$B_m^2 = (\bar{e}|b,2b,2c),$$
$$(AB_m)^2 = (\bar{e}|-2a+2b,-a+b,0), \qquad (4.39)$$

which yield

$$a=b=0, \quad c=0,\tfrac{1}{2}. \qquad (4.40)$$

Thus, we obtain two independent solutions for the AB_m type,

$$(\bar{6}_z|0), \quad (m_0|0,0,c); \quad c=0,\tfrac{1}{2}. \qquad (4.41)$$

For the AB_u type, we have

$$B_u^2 = (\bar{e}|2a-b,0,0), \quad (AB_u)^2 = (\bar{e}|0,-a-b,-2c), \qquad (4.42)$$

which yield

$$a=-b, \quad b=0,\tfrac{1}{3},\tfrac{2}{3}, \quad c=0,\tfrac{1}{2}. \qquad (4.43)$$

However, $(u_0|a,b,c)\rightarrow(u_0|0,0,c)$ under $[e|-a,-b,0]$, which leaves $(\bar{6}_z|0)$ invariant. Thus, we again arrive at two independent solutions given by

$$(\bar{6}_z|0), \quad (u_0|0,0,c); \quad c=0,\tfrac{1}{2}. \qquad (4.44)$$

7. $D_{6h}(=D_{6i})$; $A,B\in D_6$, $I^2 = [A,I] = [B,I] = E$

Augmenting D_6 we may set

$$A = (6_z|0,0,c), \quad B = (u_0|0), \quad I = (i|0,0,\gamma). \qquad (4.45)$$

Then, from the commutators

$$[A,I] = (e|0,0,2c), \quad [B,I] = (e|0,0,-2\gamma), \qquad (4.46)$$

we obtain

$$c,\gamma = 0,\tfrac{1}{2}, \qquad (4.47)$$

which provide four independent solutions.

C. The rhombohedral system

The hexagonal coordinates are used for the primitive hexagonal lattice P and for the double-centered hexagonal lattice R^*, while the rhombohedral coordinates are used when the latter is regarded as the rhombohedral lattice R. On account of the difference in the coordinate systems of the P and R lattices we shall discuss their space groups separately. Then we simply write down the results for the R^* lattice from those of the R lattice. It is noted that there exist two types of realizations of the generators for the classes D_3, C_{3v}, and D_{3i} with the P lattice but not with the R (or R^*) lattice.

1. The class C_3; $A^3 = \bar{E}$

(i) P lattice; $A=(3_z/0,0,c)$; $c=0,\tfrac{1}{3},\tfrac{2}{3}$ (4.48)

(ii) R lattice; $A=(3_{xyz}/0)$. R^* *lattice;* $A=(3_z/0)$ (4.49)

2. The class C_{3i}; $A^3 = \bar{E}$, $I^2 = [A,I] = E$

(i) P lattice: We may set[18]

$$B = (\bar{3}_z|0), \quad \bar{E} = (\bar{e}|0), \qquad (4.50)$$

and obtain

$$A = \bar{E}B^4 = (3_z|0), \quad I = \bar{E}B^3 = (i|0). \qquad (4.51)$$

(ii) R lattice: In terms of the rhombohedral coordinates and using (3.7) we obtain only one space group characterized by

$$A = (3_{xyz}|0), \quad I = (i|0), \qquad (4.52a)$$

which may be rewritten for the R^* lattice as follows:

$$(3_z|0), \quad (i|0). \qquad (4.52b)$$

3. The class D_3; $A^3 = B^2 = (AB)^2 = \bar{E}$

(i) P lattice: There exist two types of realizations for the second generator B, which is a diad; either it is parallel or it is perpendicular to one of the primitive lattice vectors in the x,y, plane. We set

$$A = (3_z|0,0,c) \quad B_\nu = (u_\nu|a,b,0), \quad \nu=0,1, \qquad (4.53)$$

where $u_0 \| x$ and $u_1 \perp y$.

For the AB_0 type we have

$$A^3=(\bar{e}|0,0,3c),\quad B_0^2=(\bar{e}|2a-b,0,0),$$
$$(AB_0)^2=(\bar{e}|a-2b,a-2b,0), \tag{4.54}$$

which yield

$$b=2a,\quad c,a=0,\tfrac{1}{3},\tfrac{2}{3}. \tag{4.55}$$

For each given value of c, however, there exists only one independent solution; this can be shown as in the case of AB_u type of D_{3h}. Thus, we are left with only three inequivalent sets of the parameters for the AB_0 type,

$$a=b=0,\quad c=0,\tfrac{1}{3},\tfrac{2}{3}. \tag{4.56}$$

For the AB_1 type, we have

$$B_1^2=(\bar{e}|2a,a,0),\quad (AB_1)^2=(\bar{e}|a-b,2a-2b,0), \tag{4.57}$$

which immediately yield the same parameter sets as those given by (4.56).

Thus, there exist altogether six independent space groups belonging to the P lattice given by

$$(3_z|0,0,c),\quad (u_\nu|0);\quad \nu=0,1,\quad c=0,\tfrac{1}{3},\tfrac{2}{3}. \tag{4.58}$$

(iii) R lattice: In terms of the rhombohedral coordinates we set

$$A=(3_{xyz}|0),\quad B=(u_{x\bar{z}}|a,b,c). \tag{4.59}$$

Then

$$B^2=(\bar{e}|a-c,0,-a+c),$$
$$(AB)^2=(\bar{e}|0,a-b,-a+b), \tag{4.60}$$

which yield $a=b=c$. Consequently, we can reduce (4.59) into the following form, via $[e|a/2,a/2,a/2]$,

$$(3_{xyz}|0),\quad (u_{x\bar{z}}|0), \tag{4.61a}$$

which may be rewritten for the R * lattice as follows:

$$(3_z|0),\quad (u_0|0). \tag{4.61b}$$

4. The class C_{3v}; $A^3=B^2=(AB)^2=\bar{E}$

(i) P lattice: Replacing the binary rotations u_ν of D_3 by the reflections $m_\nu=iu_\nu$ we arrive at the class C_{3v}. We set

$$A=(3_z|0,0,c),\quad B_\nu=(m_\nu|a',b',c'),\quad \nu=0,1. \tag{4.62}$$

Then, for the AB_0 type we have

$$A^3=(\bar{e}|0,0,3c),$$
$$B_0^2=(\bar{e}|b',2b',2c'),\quad (AB_0)^2=(\bar{e}|-a',a',2c+2c'), \tag{4.63}$$

which yield two inequivalent sets of the parameters,

$$c=a'=b'=0,\quad c'=0,\tfrac{1}{2}. \tag{4.64}$$

For the AB_1 type

$$B_1^2=(\bar{e}|0,2b'-a',2c'),$$
$$(AB_1)^2=(\bar{e}|-a'-b',0,2c+2c'), \tag{4.65}$$

which give again the same parameter sets as given by (4.64) through a shift $[e|b',-b',0]$. Thus, we obtain altogether four independent solutions for the P lattice,

$$(3_z|0),\quad (m_\nu|0,0,c');\quad \nu=0,1,\quad c'=0,\tfrac{1}{2}. \tag{4.66}$$

(ii) R lattice: Let

$$A=(3_{xyz}|0),\quad B=(m_{x\bar{z}}|a,b,c). \tag{4.67}$$

Then, from

$$B^2=(\bar{e}|a+c,2b,a+c),$$
$$(AB)^2=(\bar{e}|2c,a+b,a+b), \tag{4.68}$$

we obtain two inequivalent sets of the parameters

$$a=b=c=0,\tfrac{1}{2}, \tag{4.69}$$

which define two independent solutions of (4.67). These may be rewritten for the R * lattice as follows:

$$(3_z|0),\quad (m_0|0,0,c);\quad c=0,\tfrac{1}{2}. \tag{4.70}$$

5. The class $D_{3d}(=D_{3i})$; $A,B\in D_3$, $I^2=[A,I]=[B,I]=E$

(i) P lattice: Augmenting the class D_3 we may set, using (3.7),

$$A=(3_z|0,0,c),\quad B_\nu=(u_\nu|0),\quad \nu=0,1,$$
$$I=(i|0,0,\gamma), \tag{4.71}$$

where $c\in\{0,\tfrac{1}{3},\tfrac{2}{3}\}$ and γ are to be specified further by the commutators,

$$[A,I]=(e|0,0,2c),$$
$$[B_0,I]=(e|0,0,-2\gamma),$$
$$[B_1,I]=(e|0,0,-2\gamma). \tag{4.72}$$

The allowed values are

$$c=0,\quad \gamma=0,\tfrac{1}{2}. \tag{4.73}$$

Thus we obtain two independent solutions each for both types which may be rewritten as follows:

$$(3_z|0),\quad (u_\nu|0,0,\gamma),\quad (i|0);\quad \nu=0,1,\quad \gamma=0,\tfrac{1}{2}. \tag{4.74}$$

(ii) R lattice: From (4.16a) we may set

$$A=(3_{xyz}|0),\quad B=(u_{x\bar{z}}|0),\quad I=(i|\alpha,\alpha,\alpha). \tag{4.75}$$

Then

$$[A,I]=(e|0),\quad [B,I]=(e|-2\alpha,-2\alpha,-2\alpha), \tag{4.76}$$

which yield two independent solutions for the R lattice characterized by

$$\alpha=0,\tfrac{1}{2}. \tag{4.77}$$

The solutions may be rewritten for the R * lattice as follows:

$$(3_z|0),\quad (u_0|0),\quad (i|0,0,\alpha)\quad \text{or}$$
$$(3_z|0),\quad (u_0|0,0,\alpha),\quad (i|0);\quad \alpha=0,\tfrac{1}{2}. \tag{4.78}$$

D. Tetragonal system

The coordinate system based on the conventional lattice vectors is used with the z axis being parallel to the fourfold axis of rotation. The redundancies of the solutions are low for this system and easily removed by simple shifts of the lattice origin. In particular the following transformations will be very useful:

$$(4_z|a,b,c)\sim(4_z|a-\tfrac{1}{2},b+\tfrac{1}{2},c)$$
$$\text{via } [e|\tfrac{1}{2},0,0], \tag{4.79a}$$
$$(4_z|a,b,c)\sim(4_z|a-\tfrac{1}{2},b,c)\quad \text{via } [e|\tfrac{1}{4},\tfrac{1}{4},\tfrac{1}{4}], \tag{4.79b}$$

where (4.79a) leaves $(i|a,b,c)$ invariant always while (4.79b) leaves $(i|a,b,c)$ invariant only for the I lattice.

1. The class C_4; $A^4 = \bar{E}$

We set

$$A = (4_z|0,0,c). \tag{4.80}$$

Then, for the P lattice,

$$c = 0,\tfrac{1}{4},\tfrac{2}{4},\tfrac{3}{4}, \tag{4.81}$$

while for the I lattice, using (4.79a),

$$c = 0,\tfrac{1}{4}. \tag{4.82}$$

2. The class $S_4(=C_{2p})$; $A^4 = \bar{E}$

One may simply set

$$A = (\bar{4}_z|0), \tag{4.83}$$

for both p and I lattices, since there exists no invariant eigenvector space for $\bar{4}_z$.

3. The class $C_{4h}(=C_{4I})$; $A^4 = \bar{E}$, $I^2 = [A,I] = E$

It is simpler to set

$$A = (4_z|a,b,c), \quad I = (i|0), \tag{4.84}$$

instead of augmenting C_4 by I. Then

$$A^4 = (\bar{e}|0,0,4c), \quad [A,I] = (e|2a,2b,2c). \tag{4.85}$$

(i) P lattice: All three parameters are binary,

$$a,b,c = 0,\tfrac{1}{2}. \tag{4.86}$$

One can set, however, one of the parameters, a or b, equal to zero using the transformation (4.79a). Thus, we obtain only four independent solutions

$$(4_z|a,0,c), \quad (i|0); \quad a,c = 0,\tfrac{1}{2}. \tag{4.87}$$

(ii) I lattice: First, the parameters of (4.87) are further reduced to $a = c = 0$ for the I lattice using (4.79a) and (4.79b). Second, directly from (4.85) one obtains another set of the parameters given by

$$a = b = c = \tfrac{1}{4}. \tag{4.88}$$

Thus, together we have two independent solutions given by

$$(4_z|d,d,d), \quad (i|0); \quad d = 0,\tfrac{1}{4}. \tag{4.89}$$

4. The class D_4; $A^4 = B^2 = (AB)^2 = \bar{E}$

Let

$$A = (4_z|0,0,c), \quad B = (2_x|a,b,0). \tag{4.90}$$

Then

$$A^4 = (\bar{e}|0,0,4c), \quad B^2 = (\bar{e}|2a,0,0),$$
$$(AB)^2 = (\bar{e}|a-b,a-b,0). \tag{4.91}$$

(i) P lattice: The allowed values of the parameters are

$$a = b = 0,\tfrac{1}{2}, \quad c = 0,\tfrac{1}{4},\tfrac{2}{4},\tfrac{3}{4}, \tag{4.92}$$

which yield eight inequivalent solutions of (4.90) without any redundancy.

(ii) I lattice: We again arrive at (4.92), which are, however, further reduced to two inequivalent sets

$$a = b = 0, \quad c = 0,\tfrac{1}{4}, \tag{4.93}$$

since $(a,b) \sim (0,0)$ under $[e|0,0,\tfrac{1}{4}]$ while $c \sim c + \tfrac{1}{2}$ by (4.79a).

5. The class C_{4v}; $A^4 = B^2 = (AB)^2 = \bar{E}$

Let

$$A = (4_z|0,0,c), \quad B = (m_x|a',b',c'). \tag{4.94}$$

Then

$$A^4 = (\bar{e}|0,0,4c), \quad B^2 = (\bar{e}|0,2b',2c'),$$
$$(AB)^2 = (\bar{e}|-a'-b',a'+b',2c+2c'). \tag{4.95}$$

(i) P lattice: The allowed values of the parameters are

$$a' = b', \quad c,a',c' = 0,\tfrac{1}{2}, \tag{4.96}$$

which yield eight solutions for (4.94) without any redundancy.

(ii) I lattice: Since b' is binary from (4.95), we may set $b' = 0$ and then obtain

$$b' = 0, \quad c' = 0,\tfrac{1}{2}, \quad 2c = a' = 0,\tfrac{1}{2}, \tag{4.97}$$

which yield four independent solutions given by

$$(4_z|0,0,a'/2), \quad (m_x|a',0,c'); \quad a',c' = 0,\tfrac{1}{2}. \tag{4.98}$$

6. The class $D_{2d}(=D_{2p})$; $A^4 = B^2 = (AB)^2 = \bar{E}$

Two types of the space groups exist for this class depending on the second generator B; either it is 2_x or m_x. Thus, we set

$$A = (\bar{4}_z|0), \quad B_2 = (2_x|a,b,c), \quad B_m = (m_x|a,b,c). \tag{4.99}$$

For the AB_2 type,

$$B_2^2 = (\bar{e}|2a,0,0), \quad (AB_2)^2 = (\bar{e}|a+b,-a-b,2c), \tag{4.100}$$

and for the AB_m type,

$$B_m^2 = (\bar{e}|0,2b,2c), \quad (AB_m)^2 = (\bar{e}|-a+b,-a+b,0). \tag{4.101}$$

(i) P lattice: The allowed values of the parameters are the same for both types

$$a = b, \quad a,c = 0,\tfrac{1}{2}, \tag{4.102}$$

which yield four independent solutions for each type.

(ii) I lattice: For the AB_2 type, the parameter a is binary from (4.100). Thus from (4.99) we may set $a = 0$ first then obtain $b = 2c = 0,\tfrac{1}{2}$. These yield two independent solutions

$$(\bar{4}_z|0), \quad (2_x|0,b,b/2); \quad b = 0,\tfrac{1}{2}. \tag{4.103}$$

For the AB_m type, from (4.99) and (4.101) we first set $b = 0$ then obtain $a = 0$ and $c = 0,\tfrac{1}{2}$. These provide two independent solutions given by

$$(\bar{4}_z|0), \quad (m_x|0,0,c); \quad c = 0,\tfrac{1}{2}. \tag{4.104}$$

7. The class $D_{4h}(=D_{4I})$; $A,B \in D_4$, $I^2 = [A,I] = [B,I] = E$

In view of (4.92) and (4.93) we set

$$A = (4_z|0,0,c), \quad B = (2_x|a,a,0), \quad I = (i|\alpha,\beta,\gamma), \tag{4.105}$$

where $c \in \{0,...,\tfrac{3}{4}\}$, $a \in \{0,\tfrac{1}{2}\}$, and α,β,γ are specified further from the commutators

$$[A,I] = (e| -\alpha-\beta,\alpha-\beta,2c), \quad (4.106a)$$

$$[B,I] = (e|0,-2\beta,-2\gamma). \quad (4.106b)$$

(i) P lattice: The allowed values of the parameters are

$$c,a = 0,\tfrac{1}{2}, \quad \alpha=\beta=0,\tfrac{1}{2}, \quad \gamma=0,\tfrac{1}{2}, \quad (4.107)$$

which yield 16 space groups without any redundancy,

$$(4_z|0,0,c),\quad (2_x|a,a,0),\quad (i|\alpha,\alpha,\gamma) \quad \text{or}$$
$$(4_z|\alpha,0,c),\quad (2_x|a,a-\alpha,\gamma),\quad (i|0);\quad c,a,\alpha,\gamma=0,\tfrac{1}{2}. \quad (4.108)$$

(ii) I lattice: First, from (4.93) we have $a=0$, $c=0,\frac{1}{4}$. Second, we can set $\beta=0$, $\gamma=0,\frac{1}{2}$ for the I lattice from (4.105) and (4.106b). Then, from (4.106a) we have $\alpha=2c=0,\frac{1}{2}$. Thus, we obtain four independent solutions given by

$$(4_z|0,0,\alpha/2),\quad (2_x|0),\quad (i'|\alpha,0,\gamma) \quad \text{or}$$
$$(4_z|-\alpha/2,\alpha/2,\alpha/2),\quad (2_x|0,0,\gamma),\quad (i|0);\quad \alpha,\gamma=0,\tfrac{1}{2}. \quad (4.109)$$

E. The orthorhombic system

The rectangular coordinates based on the conventional lattice vectors are used. On account of high redundancy of the solutions, their removal will be the major problem. This is easily achieved, however, with the use of the relevant generator sets of the unimodular transformations $\{U\}$ given in (2.8); i.e., $\{(x,y),(x,z)\}$ for the classes D_2, D_{2h} and $\{4_z\} \sim \{(x,y)\}$ for the class C_{2v}.

1. The class D_2; $A^2=B^2=(AB)^2=\bar{E}$

Let

$$A=(2_z|c,0,c),\quad B=(2_x|a,b,0),\quad C=(2_y|c-a,-b,c), \quad (4.110)$$

where we have introduced an auxilary generator C for later discussion of the symmetry of the set $\{A,B,C;a,b,c\}$. By definition, the parameters c,a,b describe the screw properties of the diads A,B,C, respectively. Now, from

$$A^2=(\bar{e}|0,0,2c),\quad B^2=(\bar{e}|2a,0,0),$$
$$(AB)^2=(\bar{e}|0,-2b,0), \quad (4.111)$$

we conclude that three parameters are all binary, i.e.,

$$a,b,c=0,\tfrac{1}{2}, \quad (4.112)$$

for any lattice type belonging to D_2. In removing the possible redundant solutions we may use the following symmetry properties: Under the transformation $[x\rightleftarrows y|c/2,c/2,c/2]$,

$$\{A,B,C;a,b,c\}\rightarrow\{A,B,C;b,a,c\} \quad (4.113a)$$

and under $[x\rightleftarrows z|c/2,b/2,a/2]$,

$$\{A,B,C;a,b,c\}\rightarrow\{A,B,C;c,b,a\}, \quad (4.113b)$$

where $x\rightleftarrows y$ means the interchange of the x,y axes.

(i) P lattice: From (4.113), the set of three diads is equivalent with respect to any permutation of the parameters a,b,c. Accordingly, the inequivalent sets of the parameters are obtained by assigning 0 and $\frac{1}{2}$ to the set $\{a,b,c\}$, with repetitions. Since two of the parameters are always equal we may set $a=b$, $a,c=0,\frac{1}{2}$ to obtain four independent solutions given by

$$(2_z|c,0,c),\quad (2_x|a,a,0);\quad c,a=0,\tfrac{1}{2}. \quad (4.114)$$

Obviously these hold for the remaining lattices with further simplifications specified below.

(ii) F lattice: The solutions (4.114) are further reduced to a single solution with

$$c=a=0. \quad (4.115)$$

(iii) I lattice: The solutions (4.114) are reduced to two independent solutions with

$$c=a=0,\tfrac{1}{2}, \quad (4.116)$$

with the use of the shift $[e|0,0,\frac{1}{4}]$.

(iv) C lattice: From (4.114) we may set the binary parameter a equal to zero and then obtain two independent solutions given by

$$(2_z|c,0,c),\quad (2_x|0,0,0);\quad c=0,\tfrac{1}{2}. \quad (4.117)$$

2. The class C_{2v}; $A^2=B^2=(AB)^2=\bar{E}$

Let

$$A=(m_x|0,b,c),\quad B=(m_y|a',0,c'). \quad (4.118)$$

Then, from

$$A^2=(\bar{e}|0,2b,2c),\quad B^2=(\bar{e}|2a',0,2c'),$$
$$(AB)^2=(\bar{e}|0,0,2c-2c'), \quad (4.119)$$

we obtain

$$b,c,a',c'=0,\tfrac{1}{2}, \quad (4.120)$$

which hold for all the lattices belonging to C_{2v} except for the F lattice. Moreover, we have the following equivalence:

$$\{A(b,c),B(a',c')\}\sim\{A(a',c'),B(b,c)\}, \quad (4.121)$$

under the interchange of the x,y axes.

(i) P lattice: On account of the symmetry (4.121), the inequivalent solutions are given by choosing two sets with repetitions out of the four sets

$$(0,0),(0,\tfrac{1}{2}),(\tfrac{1}{2},0),(\tfrac{1}{2},\tfrac{1}{2}), \quad (4.122)$$

and assigning them to $\{A,B\}$. Thus one arrives at ten inequivalent space groups given by Nos. 25–34 in Table I. In the conventional notations, the sets in (4.122) define the glidings m,c,g,n of the mirror plane A or B.

(ii) C and I lattices: We may set $b=a'=0$ in (4.118) via $[e|b/2,a'/2,0]$ and then using the symmetry (4.121) obtain three inequivalent solutions

$$(m_x|0,0,c),\quad (m_y|0,0,c');\quad c<c'=0,\tfrac{1}{2}. \quad (4.123)$$

(iii) A lattice: From (4.118) we may set the binary parameter b equal to zero then $c'\rightarrow 0$ via $[e|0,c'/2,0]$. Thus, we obtain four independent solutions

$$(m_x|0,0,c),\quad (m_y|a',0,0);\quad c,a'=0,\tfrac{1}{2}. \quad (4.124)$$

(iv) F lattice: Directly from (4.119) we obtain

$$b=c=a'=c'=0,\tfrac{1}{4}, \quad (4.125)$$

which yield two independent solutions of (4.118).

When we summarize the space groups of C_{2v} in Table I, we shall present them in terms of the generators $(2_z|v_z)$ and $(2_x|v_x)$ for convenience, e.g., in constructing their irreducible representations or the magnetic space groups.

3. The class $D_{2h}(=D_{2l})$; $A,B \in D_2$, $I^2=[A,I]=[B,I]=E$

This is the largest crystal class containing 28 space groups. Augmenting the class D_2, we set

$$A=(2_z|c,0,c),\quad B=(2_x|a,a,0),\qquad C=(2_y|c+a,a,c),\quad I=(i|\alpha,\beta,\gamma), \tag{4.126}$$

where $c,a=0,\frac{1}{2}$, and an auxiliary generating element C has been introduced as before. Then from

$$[A,I]=(e|-2\alpha,-2\beta,0),\qquad [B,I]=(e|0,-2\beta,-2\gamma), \tag{4.127}$$

we obtain

$$\alpha,\beta,\gamma=0,\tfrac{1}{2}, \tag{4.128}$$

which hold for the P, I, and C lattices.

Since the augmentation does not affect the parameters c and a, the redundant parameters of α,β,γ will be removed for each given pair (c,a) by the transformations which leave the three diad set $\{A,B,C;a,a,c\}$ invariant. For this purpose we need only the symmetry properties introduced by (4.113).

(i) P lattice: (ia) When $(c,a)=(0,0)$, through permutations of the x,y,z axes we may set

$$\alpha=\beta,\gamma=0,\tfrac{1}{2}, \tag{4.129}$$

which provide four inequivalent sets.

(ib) When $(c,a)=(\frac{1}{2},0)$ by $(x\rightleftarrows y|\frac{1}{4},\frac{1}{4},\frac{1}{4})$ we may set $\gamma=0$ and obtain four inequivalent sets

$$\alpha,\beta=0,\tfrac{1}{2},\quad \gamma=0. \tag{4.130}$$

(ic) When $(c,a)=(0,\frac{1}{2})$, by interchanging the x,y axes we may set $\alpha\leqslant\beta$ and obtain six inequivalent sets

$$\alpha\leqslant\beta=0,\tfrac{1}{2},\quad \gamma=0,\tfrac{1}{2}. \tag{4.131}$$

(id) When $(c,a)=(\frac{1}{2},\frac{1}{2})$, proceeding as in (ia) we first obtain (4.129) then reduce them into

$$\alpha=\beta=0,\quad \gamma=0,\tfrac{1}{2}, \tag{4.132}$$

via $[x\rightleftarrows y\frac{1}{4},\frac{1}{4},\frac{1}{4}]$.

These 16 space groups obtained above are summarized by Nos. 47–62 in Table I.

(ii) F lattice: From (4.115) we have $c=a=0$ and then from (4.127)

$$\alpha=\beta=\gamma=0,\tfrac{1}{4}. \tag{4.133}$$

(iii) I lattice: From (4.116) we have $c=a=0,\frac{1}{2}$. Now for the set $(c,a)=(0,0)$, we obtain, through (4.129)

$$\alpha=\beta=0,\quad \gamma=0,\tfrac{1}{2}. \tag{4.134}$$

It can be shown that these also hold for $(c,a)=(\frac{1}{2},\frac{1}{2})$ through (4.132). Thus, altogether we have four independent solutions.

(iv) C lattice: From (4.117) we have $c=0,\frac{1}{2}$, $a=0$, and also we may set $\alpha=0$ for the C lattice. Thus for $(c,a)=(0,0)$, we have four inequivalent sets

$$\alpha=0,\quad \beta,\gamma=0,\tfrac{1}{2}. \tag{4.135}$$

For $(c,a)=(\frac{1}{2},0)$, proceeding as in (ib) of the P lattice, we may again set $\gamma=0$ to obtain two inequivalent sets

$$\alpha=\gamma=0,\quad \beta=0,\tfrac{1}{2}. \tag{4.136}$$

F. Monoclinic system

The coordinate system based on the conventional lattice vectors is used with the z axis being parallel to the binary axis of rotation. The redundant solutions may be removed by using the relevant set of the unimodular transformations $\{U\}$ given by $\{(x,y),(x\pm y,y,z)\}$ in (2.8).

1. The class C_2; $A^2=\bar{E}$

Let

$$A=(2_z|0,0,c), \tag{4.137}$$

then for the P lattice we have two inequivalent solutions with

$$c=0,\tfrac{1}{2}, \tag{4.138}$$

while for the B lattice we have only one solution with

$$c=0, \tag{4.139}$$

since $c\sim c+\frac{1}{2}$ under $[e|c/2,0,0]$.

2. The class $C_s(=C_{1p})$; $A^2=\bar{E}$

We set

$$A=(m_z|a,b,0), \tag{4.140}$$

and obtained $a,b=0,\frac{1}{2}$. Then using the relevant set $\{U\}$, we arrive at

$$a=0,\quad b=0,\tfrac{1}{2}, \tag{4.141}$$

which provide two inequivalent solutions each for both the P and B lattices.

3. The class $C_{2h}(=C_{2l})$; $A^2=\bar{E}$, $I^2=[A,I]=E$

Let

$$A=(2_z|0,0,c),\quad I=(i|a,b,0). \tag{4.142}$$

Then we see that the parameters a,b,c are all binary. Proceeding as in the class C_s we arrive at

$$a=0,\quad b,c=0,\tfrac{1}{2}, \tag{4.143}$$

which provide four inequivalent solutions for the P lattice. For the B lattice, from (4.139) and (4.143) we obtain

$$a=c=0,\quad b=0,\tfrac{1}{2}, \tag{4.144}$$

which yield two independent solutions.

G. The triclinic system

There exists only one lattice type P and two classes, each of which consists of one space group: the class C_1; $\bar{E}=(\bar{e}|0)$ and the class $S_2(=C_i)$; $\bar{E}=(\bar{e}|0)$, $I=(i|0)$.

V. CONCLUDING REMARKS

Based on the defining relations (3.5) of the point groups we have constructed 32 minimal general generator sets (MGGS) for 230 space groups. The algebraic equivalence criteria of the space groups with respect to lattice transformation $\Lambda=[U|s]$ are completed by introducing a relevant set (2.8) of the unimodular matrices $\{U\}$ for each crystal class. The definite set $\{U\}$ greatly reduces the labor involved in removing the redundant solutions to arrive at the independent space groups. According to the relevant set $\{U\}$, mere shifts $[e|s]$ of the lattice origin are necessary and sufficient to determine the equivalence or inequivalence for almost all classes of high symmetry. As a result, it is simpler to construct MGGS of higher symmetry (with minor exceptions) in contrast to the existing methods,[13,14] which are based on the solvability of the space groups. The symmetry properties of MGGS with respect to $[U|s]$ play the essential role in identifying a space group and also in constructing the extended space groups.

As one can see from the total results given in Table I, the number of the translational parameters of MGGS is very much limited. Their maximum number for each crystal system is two for cubic, one for the hexagonal, rhombohedral, and monoclinic, four for tetragonal, five for orthorhombic, and zero for the triclinic system. Quite obviously, these numbers roughly measure the number of independent space groups belonging to the crystal system.

The present method is easily extended to construct the magnetic space groups since the defining relations of the magnetic point groups have already been worked out by the author.[19] In terms of MGGS of the space groups belonging to each crystal class one can construct the similar general expressions for the magnetic space groups. This will be treated, however, in a forthcoming paper.[8]

ACKNOWLEDGMENTS

Part of this paper has been presented at the 13th Internation Colloquim on the Group Theoretical Methods in Physics held at the University of Maryland, 20–25, May 1984. The author is very grateful to professor Joseph L. Birman, Department of Physics, The City College, The City University of New York, New York for his encouragement and helpful suggestion to this work.

[1]J. L. Birman, Phys. Rev. Lett. **17**, 1216 (1966); Physica A **114**, 564 (1982).
[2]M. Senechal, Acta Crystallogr. Sect. A **36**, 845 (1980).
[3]M. V. Jaric, J. Math. Phys. **24**, 2865 (1983).
[4]D. M. Hatch and H. T. Stokes, Phys. Rev. B **31**, 2908 (1985).
[5]L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Addison–Wesley, New York, 1958).
[6]N. F. M. Henry and K. Lonsdale, *International Tables for x-ray Crystallography* (Kynoch, Birmingham, U.K., 1965), Vol. 1.
[7]S. K. Kim, J. Math. Phys. **24**, 411 (1983).
[8]S. K. Kim, J. Math. Phys. **27**, 1484 (1986).
[9]H. S. M. Coxeter and W. O. J. Moser, *Generators and Relations for Discrete Groups* (Springer, New York, 1980), 4th ed.
[10]H. Zassenhaus, Comment. Math. Helv. **21**, 117 (1948).
[11]H. Brown, Math. Comput. **23**, 499 (1969).
[12]H. Brown, R. Bülow, J. Neubüser, H. Wondratschek, and H. Zassenhaus, *Crystallographic Groups of Four-Dimensional Space* (Wiley, New York, 1978).
[13]P. Seitz, Z. Kristallogr. Kristallgeom. Kristallphys. Kristalchem. **88**, 433 (1934); **90**, 289 (1935); **91**, 336 (1935); **94**, 100 (1936).
[14]J. J. Burckhardt, Comm. Math. Helvet. **2**, 91 (1930); **6**, 159 (1934); **9**, 284 (1937). See also *Die Bewegungsgruppen der Kristallographie* (Birkhäuser, Basel, 1966), 2nd ed.
[15]A. Schoenflies, *Kristallsystem und Kristallstruktur* (Teubner, Leipzig, 1891).
[16]E. F. Bertaut, Acta Crystallogr. Sect. A **26**, 409 (1970).
[17]S. K. Kim, J. Math. Phys. **24**, 414 (1983).
[18]Note that there exists no invariant eigenvector space for $\bar{3}_z$.
[19]S. K. Kim, J. Math. Phys. **25**, 189 (1984).

The 38 assemblies of the general generator sets for 1421 magnetic double space groups

Shoon K. Kim
Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122

(Received 24 July 1985; accepted for publication 11 December 1985)

Thirty eight assemblies of general expressions of the generator sets for 1421 magnetic double space groups in terms of the 32 minimal general generator sets of the unitary space groups are provided.

I. INTRODUCTION

Let us understand by the magnetic space groups $\hat{M}$ the trivial magnetic (or gray) space groups as well as the nontrivial magnetic (or black and white) space groups. Belov, Neronova, and Smirnova[1] provided a listing of all 1421 of the magnetic space groups and the manner of deriving them through geometrical consideration. However, no rigorous analytical method of constructing them has been reported in the previously available literature. There exist extensive reviews by Opechowski and Guccione[2] and by Bradley and Cracknell[3] on the subject.

In a previous work using the defining relations of the point groups the author[4] has constructed the 32 minimal general generator sets (MGGS) of the unitary space groups $\hat{G}$ (this work will be referred to as I hereafter). In an analogous manner we can construct the generator sets of $\hat{M}$ such that one can describe all 1421 of them by 38 assemblies of general expressions in terms of MGGS of the unitary space groups $\hat{G}$. One can achieve this with a minor amount of additional effort on account of MGGS of $\hat{G}$.

For convenience of their construction we shall first subdivide the magnetic space groups $\hat{M}$ into two kinds[5] denoted as $\hat{G}^e$ and $\hat{H}^z$; the factor group of the first kind $\hat{G}^e/T$ with respect to the translational group T is isomorphic to a gray point group, while the factor group of the second kind $\hat{H}^z/T$ is isomorphic to a magnetic point group. The first kind will be constructed in Sec. II by augmenting a given space group $\hat{G}$ with the time inversion translation operators $\hat{\theta}$. These are determined from the defining relations of the time inversion operator θ. The magnetic Bravais lattices follow automatically from the lattice type L of the crystal class G and the allowed augmentors $\hat{\theta}$'s. The second kind will be constructed in Sec. III through the left coset decomposition of $\hat{G}$ with respect to its halving subgroup $\hat{H}$ as usual. This will be achieved by introducing simple algebraic lemmas on the generators of the point group $G(\simeq \hat{G}/T)$. The lemmas directly construct the coset decompositions of $\hat{G}$ from its generators with the help of their defining relations. The two kinds of the magnetic space groups thus constructed based on the MGGS of a crystal class G will be combined together and presented as one assembly. In some cases, however, more than one assembly of the magnetic space groups $\hat{M}$ may correspond to a crystal class G on account of the difference in the symmetry properties of the space groups $\hat{G}$. As a result, we arrive at 38 (instead of 32) assemblies of the general expressions of the generator sets, which describe 1421 magnetic space groups. These will be presented in Table I. Unless otherwise specified we shall use the same notations as in I. In particular, we mean by a point group or a space group the respective double group.

II. THE MAGNETIC SPACE GROUPS OF THE FIRST KIND

Let G be a point group and G^e be the gray group defined by

$$G^e = G + \theta G, \quad \theta^2 = \bar{e}, \tag{2.1}$$

where θ is the time inversion operator, which is antiunitary, and $\bar{e}$ is a 2π rotation. Let $\hat{G}$ be a space group corresponding to G, then the magnetic space group of the first kind is defined by

$$\hat{G}^e = \hat{G} + \hat{\theta}\hat{G}, \quad \hat{\theta} = (\theta|v_\theta), \tag{2.2}$$

where $\hat{\theta}$ commutes with all elements of $\hat{G}$ and v_θ is a minimum translation associated with θ. Let T be the translation group of $\hat{G}$ and let $\hat{R} = (R|v_R)$ be an element of the factor group $\hat{G}/T$. Then from the isomorphism $\hat{G}^e/T \simeq G^e$ and using the defining relations of G^e, we obtain

$$\hat{\theta}^2 = (\theta^2|2v_\theta) = (\bar{e}|0), \tag{2.3a}$$

$$[R, \hat{\theta}] = (e|Rv_\theta - v_\theta) = (e|0), \quad \forall R \in G. \tag{2.3b}$$

These can be rewritten as

$$2v_\theta \in T, \quad Rv_\theta = v_\theta \quad (\text{mod}\, t \in T), \quad \forall R \in G, \tag{2.4}$$

which characterize v_θ (mod $t \in T$) as a binary vector belonging to the invariant eigenvector space of all R's of G.

All the possible $\hat{G}^e$ are constructed from (2.3) or (2.4) by determining the allowed values of the parameters of $v_\theta = (\xi, \eta, \zeta)$ for each lattice type L. An immediate conclusion is that these parameters are all binary, i.e., they are limited to

$$\xi, \eta, \zeta = 0, \tfrac{1}{2} \tag{2.5}$$

for any lattice type L. Further restrictions on these follow from the commutation relation with a characteristic generator R_c of a given G whose order is higher than 2. Here we may restrict R_c to be proper since the corresponding improper operator $\bar{R}_c$ leads to the same conclusion. According to the defining relations of the point groups [see Eq. (3.5)] such an operator R_c is characteristic to each crystal system. Thus, the additional restrictions imposed on the parameter sets are also characteristic to each crystal system.

Obviously, it is possible that some of $\hat{G}^e$ thus obtained may be equivalent. Let $\Lambda = [U|s]$ be a lattice transformation that leaves invariant the halving subgroup $\hat{G}$ of $\hat{G}^e$ [see Eq. (2.5) of I]. Then, from

 0022-2488/86/051484-06$02.50

$$\Lambda(\theta|v_\theta)\Lambda^{-1} = (\theta|Uv_\theta), \quad (2.6a)$$

we have the equivalence

$$v_\theta \sim Uv_\theta \quad \text{under } \Lambda. \quad (2.6b)$$

Thus, the inequivalent sets of $\hat{\theta}$'s may depend not only on the crystal system but also on the symmetry of the space group $\hat{G}$. According to (2.4)–(2.6) the equivalence of different $\hat{G}^e$ may occur only under nontrivial outer automorphisms that leave $\hat{G}$ invariant. This severely limits the equivalence of those belonging to the crystal classes of high symmetry [see Eq. (2.8) of I]. In fact, it is observed only for those belonging to the orthorhombic or monoclinic systems.

Before actual construction of $\hat{G}^e$, it is convenient for its presentation to regard $\hat{G}^e$ as a direct product of two (double) space groups.

$$\hat{G}^e = M_L \times \hat{G}, \quad M_L/T = \hat{e} + \hat{\bar{e}} + \hat{\theta}, \quad (2.7)$$

and call the group M_L as a magnetic Bravais lattice (MBL). Since L of G defines the transformation group T, a MBL may be denoted by L and $\hat{\theta}$. Thus, all the possible MBL's corresponding to (2.5) are expressed by

$$\begin{aligned} L_0 &= L(\theta|0,0,0), & L_a &= L(\theta|\tfrac{1}{2},0,0), \\ L_b &= L(\theta|0,\tfrac{1}{2},0), & L_c &= L(\theta|0,0,\tfrac{1}{2}), \\ L_A &= L(\theta|0,\tfrac{1}{2},\tfrac{1}{2}), & L_B &= L(\theta|\tfrac{1}{2},0,\tfrac{1}{2}), \\ L_C &= L(\theta|\tfrac{1}{2},\tfrac{1}{2},0), & L_I &= L(\theta|\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}). \end{aligned} \quad (2.8)$$

There exist a total of 48 MBL's since L may take one of the six conventional lattice types (P,A,B,C,F,I). Obviously, some of them may be equal or equivalent or may not be allowed for a certain crystal system. Each allowed L_x defines an additional set of equivalent points with the time inversion θ in the conventional unit cell of the Bravais lattice L of G. The above notations are in accordance with those introduced by Belov *et al.*[1] for the black and white lattices in terms of figures.[6] They introduced 36 topologically independent MBL's through geometrical consideration. We shall construct them simply through further restriction of (2.5) with the characteristic generator R_c for each crystal system or removing possible redundance through equivalence (2.6). We shall begin with the simplest case of the cubic system.

A. The cubic system

The characteristic proper point operation R_c for this system is 3_{xyz}. Its invariant eigenvectors with the binary parameters are given by

$$\xi = \eta = \zeta = 0, \tfrac{1}{2}. \quad (2.9)$$

Thus, for $L = P, F$, or I, we obtain five independent MBL's

$$\{P_0, P_i\}, \ \{F_0, F_s\}, \ I_0, \quad (2.10)$$

where $F_s = F(\theta|\frac{1}{2},\frac{1}{2},\frac{1}{2})$, which is equal to F_a, F_b, or F_c.

B. The hexagonal system

With $R_c = 6_z$ in (2.3) we have

$$[6_z, \theta] = (e| -\eta, \xi - \eta, 0), \quad (2.11)$$

which yields $\xi = \eta = 0$, $\zeta = 0, \frac{1}{2}$, for $L = P$. Thus we obtain

$$P_0, P_c. \quad (2.12)$$

C. The rhombohedral system

With $R_c = 3_{xyz}$ in the rhombohedral coordinates for the R lattice we obtain

$$R_0, R_I, \quad (2.13a)$$

as in the case of the cubic system. These may be reexpressed by

$$R_0^*, R_c^*, \quad (2.13b)$$

in the hexagonal coordinates for the double-centered hexagonal lattice R^*. For the primitive hexagonal lattice P of the system we obtain again two MBL's given by (2.12).

D. The tetragonal system

With $R_c = 4_z$, we obtain $\xi = \eta = 0, \frac{1}{2}$, $\zeta = 0, \frac{1}{2}$, which yield

$$\{P_0, P_c, P_C, P_I\}, \quad \{I_0, I_c\}, \quad (2.14)$$

corresponding to $L = P, I$.

E. The orthorhombic system

Since there is no characteristic R_c for this system, all the parameter sets given by (2.5) are allowed with possible redundances depending on the symmetry of the space group $\hat{G}$. According to the relevant set of the outer automorphisms for G given in (2.8) of I, we need to consider only the permutations of the x,y,z axes followed by appropriate shifts for this system.

(i) If the three axes of the coordinate system are inequivalent for a given $\hat{G}$, we have the following independent sets of MBL's corresponding to

$L = P, C, A, F, I$:

$$\begin{aligned} M(8P) &= \{P_0, P_a, P_b, P_c, P_A, P_B, P_C, P_I\}, \\ M(C) &= \{C_0, C_b, C_I, C_B\}, \\ M(A) &= \{A_0, A_b, A_I, A_B\}, \\ M(F) &= \{F_0, F_s\}, \\ M(4I) &= \{I_0, I_a, I_b, I_c\}, \end{aligned} \quad (2.15)$$

where $F_s = F(\theta|\frac{1}{2},\frac{1}{2},\frac{1}{2})$ and the lattice types C and A are in accordance with those of the space groups $\hat{G}$ given in I.

(ii) If $\hat{G}$ is invariant under $\Lambda = [(x,y)|s]$, where (x,y) denotes the interchange of the x, y axes, we may characterize the independent parameters by

$$\xi \leq \eta = 0, \tfrac{1}{2}, \quad \gamma = 0, \tfrac{1}{2}. \quad (2.16)$$

Then we obtain

$$\begin{aligned} M(6P) &= \{P_0, P_b, P_c, P_A, P_C, P_I\}, \\ M(3I) &= \{I_0, I_b, I_c\}. \end{aligned} \quad (2.17)$$

The above symmetry does not affect $M(C)$ or $M(F)$ of (2.15) while it cannot apply for the A lattice.

(iii) If $\hat{G}$ is invariant under $\Lambda = [(x,y,z)|s]$, where (x,y,z) denotes the permutations of the x, y, and z axes, we may set

$$\xi \leq \eta \leq \zeta = 0, \tfrac{1}{2}, \quad (2.18)$$

which yield, for $L = P, I$,

$$M(4P) = \{P_0, P_c, P_A, P_I\},$$
$$M(2I) = \{I_0, I_c\}. \quad (2.19)$$

This symmetry does not affect $M(F)$ of (2.15), and it cannot apply for $L = C$ or A.

F. The monoclinic system

Let the binary axis of rotation be parallel to the z axis. Then the angle between the x and y axes is arbitrary. As in the previous case of E, the possible redundance of (2.5) may be removed by the outer automorphisms with a set of linear transformations $\{(y, x, z), (x \pm y, y, z)\}$ and appropriate shifts for this system. If the x and y axes are inequivalent for $\widehat{G}$, we have, for $L = P, B$,

$$M(6P) = \{P_0, P_a, P_b, P_c, P_A, P_B\},$$
$$M(B) = \{B_0, B_b, B_c\}. \quad (2.20)$$

If $\widehat{G}$ is invariant under $[(x, y)|s]$ we have

$$M(4P) = \{P_0, P_a, P_c, P_A\}. \quad (2.21)$$

This symmetry does not apply for $L = B$.

G. The triclinic system

Since all three angles of the coordinate system are arbitrary, we have only two MBL's given by

$$M(P) = \{P_0, P_s\}, \quad (2.22)$$

where $P_s = P(\theta|0, 0, \frac{1}{2})$, taking the z axis in the direction of (ξ, η, ζ).

Thus we have constructed altogether 48 MBL's, of which only 36 are for the holohedral space groups in agreement with the result of Belov *et al.*[1] In Table I we have presented the magnetic space groups of the first kind $\widehat{G}^e$ in terms of MBL's and MGGS of each crystal class. It contains also the magnetic space groups of the second kind $\widehat{H}^z$, which will be discussed in Sec. III.

III. THE MAGNETIC SPACE GROUPS OF THE SECOND KIND

Let $\widehat{H}$ be a halving subgroup of a space group $\widehat{G}$, then the left coset decomposition of $\widehat{G}$ with respect to $\widehat{H}$ yields

$$\widehat{G} = \widehat{H} + \hat{z}\widehat{H}, \quad (3.1)$$

where $\hat{z}$ is the coset representative that is a generator of $\widehat{G}$. Since $\widehat{H}$ is an invariant subgroup, $\hat{z}$ must satisfy the following compatibility conditions expressed by their angular parts:

$$z^2 \in H, \quad zHz^{-1} = H. \quad (3.2)$$

Accordingly, not every generator of $\widehat{G}$ may be acceptable for $\hat{z}$, as will be discussed in detail. For each left coset decomposition $\{\hat{z}, \widehat{H}\}$ of $\widehat{G}$ one can define a magnetic space group by

$$\widehat{H}^z = \widehat{H} + \hat{z}'\widehat{H}, \quad \hat{z}' = \theta\hat{z}. \quad (3.3)$$

It is obvious from the definition that the factor group $\widehat{H}^z/T$ is isomorphic to the magnetic point group defined by

$$H^z = H + z'H, \quad z' = \theta z. \quad (3.4)$$

We shall now show how to construct $\widehat{H}^z$ from a given space group $\widehat{G}$ when its factor group $\widehat{G}/T(\simeq G)$ is expressed by its generators. As discussed before in I the abstract generators of the point group G are characterized by the following defining relations:

$$A^n = B^l = (AB)^m = \overline{E}, \quad I^2 = E, \quad (3.5)$$

where I is the inversion operator that is in the center of G if it is contained in G. The set of orders $\{n, l, m\}$ is characteristic to G: $\{n, 0, 0\}$ for C_n, $\{n, 2, 2\}$ for D_n, $\{3, 3, 2\}$ for T, and $\{4, 3, 2\}$ for O. Let an element with an odd (even) order with respect to $\overline{E}$ be simply called an odd (even) element. Then, from (3.2) and (3.5) we arrive at the following lemmas, which play the essential role for constructing $\widehat{H}^z$ from a given $\widehat{G}$.

Lemma 1: An odd generator of G is not acceptable for z. The highest even generator of a cyclic group and the inversion operator I of any G are always acceptable for z.

Lemma 2: If there exist two generators $A, B(\neq I)$ for G, choose them such that $(AB)^2 = \overline{E}$. Then, even one, say A, is acceptable for z. The corresponding halving subgroup H is formed by A^2, B, and I (if I is contained in G).

The proof is elementary. For Lemma 2, it follows from $zBz^{-1} = EB^{-1}z^{-2} \in H$. Hereafter it is assumed that two generators $A, B(\neq I)$ are chosen in accordance with Lemma 2 as in (3.5). The effectiveness of the lemmas is obvious since all the required orders are given by the defining relations (3.5). In the traditional approach, the antiunitary operators of H^z are determined from the character table of the point group.[3,7]

According to (3.5) the generator set of a space group $\widehat{G}$ with a lattice type L can be expressed by one of the following five sets:

$$L\{A\}, L\{I\}, L\{A, I\}, L\{A, B\}, L\{A, B, I\}. \quad (3.6)$$

The corresponding $\widehat{H}^z$'s are immediately constructed by using the lemmas and (3.5). In a special case of $L\{A, B\}$ belonging to the class T, there exists no $\widehat{H}^z$ since both generators are odd. Now for the even dihedral crystal systems (orthorhombic, tetragonal, and hexagonal), the generators A, B are all even. In such a case, we obtain the following seven $\widehat{H}^z$ for the most general space group $L\{A, B, I\}$:

$$L.\ \{A', B, I\},\ \{A', (AB), I\},\ \{A, B', I\},$$
$$\{A, B, I'\},\ \{A, \overline{B}, I'\},\ \{\overline{A}, B, I'\},\ \{\overline{A}, (AB), I'\}, \quad (3.7)$$

provided that A, B, and AB are all inequivalent. Here $X' = \theta X$ and $\overline{X} = XI$, and for simplicity the lattice type L is not repeated. If only one of the generators A or B is even, say A is even, then one obtains the following $\widehat{H}^z$ corresponding to $L\{A, B, I\}$,

$$L.\ \{A', B, I\},\ \{A, B, I'\},\ \{\overline{A}, B, I'\}. \quad (3.8)$$

This case is applicable to the class D_{3h} or O_h. Thus, we have exhausted all the possibilities of $\widehat{H}^z$ arising from a given $\widehat{G}$ since the remaining cases of (3.6) may be regarded as special cases of $L\{A, B, I\}$.

We shall now discuss the completeness of the sets given by (3.7). It can be seen from the following equalities:

$$\{A', (AB), I\} = \{A', B', I\} = \{(AB), B', I\},$$
$$\{A, \overline{B}, I'\} = \{A, B', I'\},\ \{\overline{A}, (AB), I'\} = \{A', B', I'\}. \quad (3.9)$$

These equalities also give an alternative basis for construct-

TABLE I. The 38 assemblies of the general generator sets for the 1421 magnetic space groups.

	A. The cubic system
	$L = P, F, I;\quad M_L = \{P_0, P_I\},\quad F_S,\quad I_0$
$T(195\text{–}199)$:	$M_L\,(2_z\,\|c, 0\,c)(3_{xyz}\,\|0)$
$T_h\,(200\text{–}206)$:	$M_L\,(2_z\,\|c+a, a, c)(3_{xyz}\,\|0)(i\|0), L\{23i'\}$
$O(207\text{–}214)$:	$M_L\,(4_z\,\|0, -c, c)(3_{xyz}\,\|0), L\{4'3\}$
$T_d\,(215\text{–}220)$:	$M_L\,(\bar{4}_z\,\|c, -c, c)(3_{xyz}\,\|0), L\{\bar{4}'3\}$
$O_h\,(221\text{–}230)$:	$M_L\,(4_z\,\|a, -c, c)(3_{xyz}\,\|0)(i\|0), L.\,\{43i'\}, \{\bar{4}3i'\}, \{4'3i\}$
	B. The hexagonal system
	$L = P, M_L = \{P_0, P_c\}$
$C_6(168\text{–}173)$:	$M_L\,(6_z\,\|0, 0, c), P\{6'\}$
$C_{3h}\,(174)$:	$M_L\,(\bar{6}_z\,\|0), P\{\bar{6}'\}$
$C_{6h}\,(175\text{–}176)$:	$M_L\,(6_z\,\|0, 0, c)(i\|0), P.\{6'i\}, \{6i'\}, \{\bar{6}i'\}$
$D_6(177\text{–}182)$:	$M_L\,(6_z\,\|0, 0, c)(u_0\|0)(u_1\|0, 0, c)^\dagger, P.\{6'u_0\}, \{6'u_1\}, \{6u_0'\}$
$C_{6v}\,(183\text{–}186)$:	$M_L\,(6_z\,\|0, 0, c)(m_0\|0, 0, c')(m_1\|0, 0, c+c')^\dagger, P.\{6'm_0\}, \{6'm_1\}, \{6m_0'\}$
$D_{3h}\,(187\text{–}188)$:	$M_L\,(\bar{6}_z\,\|0)(m_0\|0, 0, c)(u_1\|0, 0, c)^\dagger, P.\{\bar{6}'m_0\}, \{\bar{6}'u_1\}, \{\bar{6}m_0'\}$
$(189\text{–}190)$:	$M_L\,(\bar{6}_z\,\|0)(u_0\|0, 0, c)(m_1\|0, 0, c)^\dagger, P.\{\bar{6}'u_0\}, \{\bar{6}'m_1\}, \{\bar{6}u_0'\}$
$D_{6h}\,(191\text{–}194)$:	$M_L\,(6_z\,\|0, 0, c)(u_0\|0, 0, c')(u_1\|0, 0, c+c')^\dagger(i\|0), P.\{6'u_0i\}, \{6'u_1i\}, \{6u_0'i\}, \{6u_0i'\}, \{6\bar{u}_0i'\}, \{\bar{6}u_0i'\}, \{\bar{6}u_1i'\}$
	C. The rhombohedral system
	$L = P, R^*, M_L = \{P_0, P_c\}, \{R_0^*, R_c^*\}$
$C_3(143\text{–}146)$:	$M_L\,(3_z\,\|0, 0, c)$
$C_{3i}\,(147\text{–}148)$:	$M_L\,(3_z\,\|0)(i\|0), L\{3i'\}$
$D_3(149\text{–}155)$:	$M_L\,(3_z\,\|0, 0, c)(u_v\,\|0), L\{3u_v'\}$
$C_{3v}\,(156\text{–}161)$:	$M_L\,(3_z\,\|0)(m_v\,\|0, 0, c), L\{3m_v'\}$
$D_{3d}\,(162\text{–}167)$:	$M_L\,(3_z\,\|0)(u_v\,\|0)(m_v\,\|0, 0, c)^\dagger(i\|0, 0, c), L.\{3u_v'i\}, \{3u_vi'\}, \{3m_vi'\}$
	D. The tetragonal system
	$L = P, I, M_L = \{P_0, P_c,$ $P_C, P_I), \{I_0, I_c)$
$C_4(75\text{–}80)$:	$M_L\,(4_z\,\|0, 0, c), L\{4'\}$
$S_4(81\text{–}82)$:	$M_L\,(\bar{4}_z\,\|0), L\{\bar{4}'\}$
$C_{4h}\,(83\text{–}88)$:	$M_L\,(4_z\,\|a, b, c)(i\|0), L.\{4'i\}, \{4i'\}, \{\bar{4}i'\}$
$D_4(89\text{–}98)$:	$M_L\,(4_z\,\|0, 0, c)(2_x\,\|a, a, 0)(2_{xy}\,\|a, a, c)^\dagger, L.\{4'2_x\}, \{4'2_{xy}\}, \{42_x'\}$
$C_{4v}\,(99\text{–}110)$:	$M_L\,(4_z\,\|0, 0, c)(m_x\,\|a+2c, a, c')(m_{xy}\,\|a, a+2c, c+c')^\dagger,\ L.\{4'm_x\}, \{4'm_{xy}\}, \{4m_x'\}$
$D_{2d}\,(111\text{–}114, 121, 122)$:	$M_L\,(\bar{4}_z\,\|0)(2_x\,\|a+2c, a, c)(m_{xy}\,\|a, a+2c, -c)^\dagger, L.\{\bar{4}'2\}, \{\bar{4}'m\}, \{\bar{4}2'\}$
$(115\text{–}118, 119, 120)$:	$M_L\,(\bar{4}_z\,\|0)(m_x\,\|a+2c, a, c)(2_{xy}\,\|a, a+2c, -c)^\dagger, L.\{\bar{4}'m\}, \{\bar{4}'2\}, \{\bar{4}m'\}$
$D_{4h}\,(123\text{–}142)$:	$M_L\,(4_z\,\|\alpha+c, c, c)(2_x\,\|a, a+2c+\alpha, \gamma)(2_{xy}\,\|a-c, a+c, c+\gamma)^\dagger(i\|0),$ $L.\{4'2_xi\}, \{4'2_{xy}i\}, \{42'i\}, \{42i'\}, \{\bar{4}2i'\}, \{\bar{4}2_xi'\}, \{\bar{4}2_{xy}i'\}$
	E. The orthorhombic system
	$L = P, C, A, F, I$
	$M(4P) = \{P_0, P_c, P_A, P_I\}$
	$M(6P) = \{P_0, P_b, P_c, P_A, P_C, P_I\}$
	$M(8P) = \{P_0, P_a, P_b, P_c, P_A, P_B, P_C, P_I\}$
	$M(C) = \{C_0, C_b, C_I, C_B\}$
	$M(A) = \{A_0, A_b, A_I, A_B\}$
	$M(F) = \{F_0, F_s\}$
	$M(2I) = (I_0, I_c\}, M(3I) = \{I_0, I_b, I_c\},$
	$M(4I) = \{I_0, I_a, I_b, I_c\}$
	D_2
$(16, 19; 22; 23, 24)$:	$M(4P; F; 2I)(2_z\,\|c, 0, c)(2_x\,\|c, c, 0), L(P; F; I)\{22'\}$
$(17, 18; 20, 21)$:	$M(6P; C)(2_z\,\|c, 0, c)(2_x\,\|a, a, 0), L(P, C).\{2'2\}, \{22'\}$
	C_{2v}
$(25, 27, 32, 34; 35, 37; 42, 43; 44, 45)$:	$M(6P; C; F; 3I)(2_z\,\|b, -b, 0)(m_x\,\|0, b, c), L(P; C; F; I).\,\{2'm\}, \{2m'\}$
$(26, 28\text{–}31, 33; 36; 38\text{–}41; 46)$:	$M(8P; C; A; 4I)(2_z\,\|a', -b, c+c)(m_x\,\|0, b, c)(m_y\,\|a', 0, c')^\dagger,$ $L(P; C; A; I).\{2'm_x\}, \{2'm_y\}, \{2m_x'\}$
	D_{2h}
$(47, 48, 61; 69, 70; 71, 73)$:	$M(4P; F; 2I)(2_z\,\|c+\alpha, \alpha, c)(2_x\,\|c, c+\alpha, \alpha)(i\|0), L(P; F; I).\{2'2i\}, \{22i'\}, \{2\bar{2}i'\}$
$(49, 50, 55, 56, 58, 59; 65\text{–}68; 72, 74)$:	$M(6P; C; 3I)(2_z\,\|\alpha, \alpha, 0)(2_x\,\|a, a+\alpha, \gamma)(i\|0), (P; C; I).\{2'2i\}, \{22'i\}, \{22i'\}, \{2\bar{2}i'\}, \{\bar{2}2i'\}$
$(51\text{–}54, 57, 60, 62; 63, 64)$:	$M(8P; C)(2_z\,\|c+\alpha, \beta, c)(2_x\,\|a, a+\beta, \gamma)(2_y\,\|a+c+\alpha, a, c+\gamma)^\dagger(i\|0),$ $L(P; C)\{2_z'2_xi\}, \{2_z2_x'i\}, \{2_z'2_yi\}, \{2_z2_xi'\}, \{\bar{2}_z2_xi'\}, \{2_z\bar{2}_xi'\}, \{\bar{2}_z2_yi'\}$
	F. The monoclinic system
	$L = P, B$
	$M(4P) = (P_0, P_a, P_c, P_A\}$
	$M(6P) = \{P_0, P_a, P_b, P_c, P_A, P_B\}$
	$M(B) = \{B_0, B_b, B_c\}$

$C_2(3, 4; 5)$:	$M(4P; B)(2_z\|0, 0, c), L(P; B)\{2'\}$
$C_s(6; 7; 8, 9)$:	$M(4P; 6P; B)(m_z\|0, b, 0), L(P; P; B)\{m'\}$
$C_{2h}(10, 11; 12, 15; 13, 14)$:	$M(4P; B; 6P)(2_z\|0, b, c)(i\|0), L(P; B; P).\{2'i\}, \{2i'\}, \{\bar{2}i'\}$
	G. The triclinic system $L = P, M)P) = \{P_0, P_s\}$
$C_1(1)$:	$M(P)(\bar{e}\|0)$
$C_i(2)$:	$M(P)(\bar{e}\|0)(i\|0), P\{\bar{e}i'\}$

Notes: (i) The translational parameters $a, b, ..., \gamma$ are given in Table I of I.
(ii) The numbers in the parentheses after the class symbols are the space group numbers.
(iii) The superscripts †'s denote auxiliary generators needed to describe $\hat{H}^z$.
(iv) Abbreviations of the symbols are used whenever no confusion exists.

ing $\hat{H}^z$ in terms of the international notations of the space groups as has been carried out by Belov *et al.*[1] We prefer the expressions given by (3.7) since these are directly related to their irreducible corepresentations (coirreps) determined through the projective coirreps of their point groups H^z introduced by the author.[5]

It should be noted that one assembly of $\hat{H}^z$ as given by (3.7) corresponds to one minimal general generator set (MGGS) of a crystal class except for the classes D_{3h} and D_{2d} and all three classes of the orthorhombic system. For the former two classes, it is due to two inequivalent realizations of each generator set. For the latter three classes, it is due to possible equivalence of some or all of the three diads of $\hat{G}$ as discussed in Sec. II E. Thus, two or three assemblies of $\hat{H}^z$ correspond to a crystal class for these cases. It should also be noted that the equivalence of some or all of A, B, and AB of $\hat{G}$ under lattice transformations Λ can occur only for the orthorhombic system.

All the magnetic space groups $\hat{H}^z$ thus determined are given in Table I together with $\hat{G}^e$ constructed in Sec. II. Here, for convenience, we have expressed a space group $L\{A, B\}$ or $L\{A, B, I\}$ by

$$L\{A, B, (AB)^\dagger\} \text{ or } L\{A, B, (AB)^\dagger, I\}, \quad (3.10)$$

in the case when the product (AB) is needed to describe $\hat{H}^z$ [the superscript on (AB) simply means that (AB) is an auxiliary generator].

IV. ILLUSTRATIVE EXAMPLES

On account of the compact nature of Table I, it seems worthwhile to give some illustrative examples for its use. For each example, we first write down a space group $\hat{G}$ by its generators determined from Table I of I. Then we give the corresponding magnetic space groups $\hat{G}^e$ and $\hat{H}^z$ from the present table. These are reexpressed in terms of the notations introduced by Belov, Neronova, and Smirnova[1] (BNS) for comparison.

(i) No. 230. $I(4_z|\frac{1}{4}, -\frac{1}{4}, \frac{1}{4})(3_{xyz}|0)(i|0) \in O_h$. (4.1)

Since $L = I$, we have $M_L = I_0$ from Table I-A. Then from the line O_h (221–230) we obtain

$$I_0 43i, I43i', I\bar{4}3i', I4'3i, \quad (4.2a)$$

where 4, 3 are obvious abbreviations of the generators given in the beginning. These correspond to

$$Ia3'd, Ia'3d', Ia'3d, Ia3d', \quad (4.2b)$$

in the BNS notations, respectively.

(ii) No. 70. $F(2_z|\frac{1}{4}, \frac{1}{4}, 0)(2_x|0, \frac{1}{4}, \frac{1}{4})(i|0) \in D_{2h}$. Since $L = F$, from Table I E we have $M(F), = \{F_0, F_s\}$ and then from No. 70 of D_{2h}.

$$F_0 22i, F_s 22i, F2'2i, F22i', F2\bar{2}i', \quad (4.3a)$$

which correspond to

$$Fddd1', F_s ddd, Fd'd'd, Fd'd'd', Fd'dd, \quad (4.3b)$$

in the notation of BNS.

The illustrative examples given above show that the present table is much more explicit than those given by BNS.

V. THE CONCLUDING REMARKS

Using the general generator sets of the space groups $\hat{G}$ given in I and the defining relations of the point groups, we have rigorously constructed all the generator sets of the magnetic space groups $\hat{M}$ with ease. These are presented in Table I. It describes all the generators of 1421 magnetic space groups by mere 38 assemblies of general expressions with the translational parameters that are predetermined for the space groups $\hat{G}$. In almost all cases, one assembly corresponds to each crystal class. The exceptions occur only for the five classes D_{3h}, D_{2d}, D_2, C_{2v}, and D_{2h}, where two or three assemblies of the generator sets are required to describe $\hat{H}^z$'s for each class. This has been discussed in Sec. III.

As it has been shown by the author[5] the generator sets given in Table I are essential and sufficient to construct all the irreducible corepresentations (coirreps) through the isomorphisms $\hat{G}^e/T \simeq G^e$ and $\hat{H}^z/T \simeq H^z$ from the projective coirreps of the gray point groups G^e and the magnetic point groups H^z, respectively. These general generator sets of $\hat{M}$ are also very convenient in discussing their symmetry properties as we have discussed those of the space groups $\hat{G}$.

Undoubtedly, the present compact presentation of the magnetic space groups gives us control over their large number and thus helps us to study the symmetry properties of the magnetically ordered crystalline solids even more systematically.

[1]N. V. Belov, N. N. Neronova, and T. S. Smirnova, Trudy Inst. Kristallogr. Akad. Nauk SSR **11**, 33 (1955); translated in English by A. V. Shubnikov and N. V. Belov, *Colored Symmetry* (Pergamon, Oxford, 1964).
[2]W. Opechowski and R. Guccione, "Magnetic symmetry," in *Magnetism*, edited by G. T. Rado and H. Suhl (Academic, New York, 1965), Vol. 2A, p. 105.
[3]C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford 1972).
[4]S. K. Kim J. Math. Phys. **27**, 1471 (1986).
[5]S. K. Kim, J. Math. Phys. **25**, 189 (1984). See also **24**, 419 (1983).
[6]The factor group of a black and white lattice is defined by $\hat{e} + \hat{\theta}$ with $\hat{\theta}^2 = e$.
[7]V. L. Indenbom, Sov. Phys. Crystallogr. **4**, 578 (1960); E. F. Bertaut, Acta Crystallogr. A **24**, 217 (1968).

Transition of a Kantowski–Sachs cosmological model into an inflationary era

Ø. Grøn
Institute of Physics, University of Oslo, Oslo, Norway

(Received 3 September 1985; accepted for publication 31 October 1985)

An empty Kantowski–Sachs universe model with a cosmological constant is found. It emerges from a pancake singularity at $t = 0$. If the cosmological constant is of the same magnitude as that induced by a GUT vacuum phase transition, this universe enters an exponentially expanding era at the GUT time $t_G = 1{,}0\cdot 10^{-35}$ sec. The universe then isotropizes rapidly, and the shear diminishes by a factor of 10^{-56} if this inflationary era lasts for $t_1 = 1{,}3\cdot 10^{-33}$ sec, as is the case in Guth's inflationary universe model.

I. INTRODUCTION

A qualitative study of Kantowski–Sachs cosmological models[1] has recently been performed by Weber.[2,3] These models are spatially homogeneous, have shear, and have no rotation. They do not belong to the Bianchi classes. Weber investigated such models with a cosmological constant and found that there exist models evolving towards the de Sitter universe.

Such cosmological models are of particular interest due to the possible existence of a GUT phase transition producing a vacuum-dominated inflationary era in the very early history of the universe.[4]

I will here find an exact solution of Einstein's vacuum field equations with a cosmological constant for a space-time of the Kantowski–Sachs type. This will be used to investigate if such a universe model will show a transition into an inflationary era at the GUT time.

II. EMPTY KANTOWSKI-SACHS UNIVERSE WITH COSMOLOGICAL CONSTANT

The Kantowski–Sachs metric takes the form

$$ds^2 = dt^2 - A^2\,dr^2 - B^2\,d\Omega^2, \quad d\Omega^2 = d\theta^2 + \sin^2 d\phi^2\,, \qquad (1)$$

where A and B are functions of the cosmic time t. Einstein's vacuum field equations with a cosmological constant for this metric take the form

$$2\frac{\dot A\dot B}{AB} + \frac{1+\dot B^2}{B^2} = \Lambda\,, \qquad (2)$$

$$2\frac{\ddot B}{B} + \frac{1+\dot B^2}{B^2} = \Lambda\,, \qquad (3)$$

$$\frac{\ddot A}{A} + \frac{\ddot B}{B} + \frac{\dot A\dot B}{AB} = \Lambda\,. \qquad (4)$$

Integration of Eq. (3) gives

$$\dot B^2 = H_0^2B^2 - 1 + K_1/B, \quad H_0 = (\Lambda/3)^{1/2}\,. \qquad (5)$$

Putting the integration constant K_1 equal to zero and integrating once more leads to

$$B = B_0\cosh(H_0t), \quad B_0 = B(0)\,. \qquad (6)$$

Equations (2) and (3) give

$$A = K_2\dot B\,. \qquad (7)$$

Inserting the solution (6), (7) into the field equations, one finds that they are satisfied for $B_0 = K_2 = H_0^{-1}$ after a constant adjustment of the r coordinate. The solution may now be written

$$ds^2 = dt^2 - H_0^{-2}\sinh^2(H_0t)dr^2 - H_0^{-2}\cosh^2(H_0t)d\Omega^2\,. \qquad (8)$$

This line element describes a Kantowski–Sachs vacuum space-time with cosmological constant.

III. TRANSITION INTO AN INFLATIONARY ERA

The characteristic feature of an inflationary era is that space-time expands exponentially owing to the repulsive gravitation of a dominating vacuum energy. This vacuum energy is due to Higgs fields that produce a large cosmological constant. At the GUT time this is of a magnitude so that $t_G = (H_0)^{-1} = 1{,}0\cdot 10^{-35}$ sec.

From the line element (8), it is seen that at this point of time there is a transition from an anisotropic Kantowski–Sachs universe to an isotropic de Sitter universe, with an exponentially expanding scale factor.

The expansion is

$$\Theta = \dot A/A + 2\dot B/B = H_0[\coth(H_0t) + 2\tanh(H_0t)]\,, \qquad (9)$$

which tends towards $3H_0$. The shear is

$$\sigma = 3^{-1/2}(\dot A/A - \dot B/B) = 3^{-1/3}2H_0\,\mathrm{csch}(2H_0t)\,, \qquad (10)$$

which decays exponentially towards zero when $H_0t \gg 1$.

The Guth inflationary era[5] lasts from t_G to $t_1 = 1{,}3\cdot 10^{-33}$ sec. During this era the shear diminishes by a factor 10^{-56}.

IV. CONCLUSIONS

A Kantowski–Sachs universe model describing vacuum with a cosmological constant has been found. This universe model emerges from a pancake singularity and develops towards an isotropic de Sitter universe.

If the cosmological constant is of the same magnitude as that induced by a GUT phase transition, the universe will evolve extremely rapidly from a shear-dominated state before $t_G = 1{,}0\cdot 10^{-35}$ sec to a vacuum-dominated isotropic state after this point of time. The shear diverges initially, in an infinitely thin universe with finite extension in its own plane. From $t = 0$ until $t = t_G$ the thickness of the universe

 0022-2488/86/051490-02$02.50

increases approximately linearly with time, while its extension in the other two directions remains approximately constant. For $t > t_G$ the universe expands exponentially in all directions, and the shear decays exponentially towards zero. If this inflationary era lasts until $t_1 = 1{,}3 \cdot 10^{-33}$ sec, as in Guth's inflationary universe model,[5] then the shear diminishes by a factor of the order 10^{-56} during this era.

[1]R. Kantowski and R. K. Sachs, J. Math. Phys. 7, 443 (1966).
[2]E. Weber, J. Math. Phys. **25**, 3279 (1984).
[3]E. Weber, J. Math. Phys. **26**, 1308 (1985).
[4]A. D. Linde, Rep. Progr. Phys. **47**, 925 (1984).
[5]A. H. Guth, Phys. Rev. D **23**, 347 (1981).

ERRATA

Erratum: Operators for the two-dimensional harmonic oscillator in an angular momentum basis [J. Math. Phys. 24, 2340 (1983)]

Olúwolé Òdúndún
University of Ife, Ile-Ife, Nigeria

(Received 16 December 1985; accepted for publication 8 January 1986)

The paragraph immediately above the Acknowledgments should read as follows:
Whereas the commutation relation

$$(2\boldsymbol{\lambda}^{\dagger}\cdot\boldsymbol{\lambda}+1)[\lambda_i,\lambda_j^{\dagger}]=(2\boldsymbol{\lambda}^{\dagger}\cdot\boldsymbol{\lambda}+1)\delta_{ij}-2\lambda_i^{\dagger}\lambda_j$$

holds in the three-dimensional case, the commutation relation

$$M[\lambda_i,\lambda_j^{\dagger}]=M\delta_{ij}-\lambda_i^{\dagger}\lambda_j$$

is found to hold in the two-dimensional situation.

The author would like to thank Dr. A. J. Bracken and Mr. A. S. McAnally, both of the University of Queensland, Australia, for pointing out this error.

Erratum: Functional integrals for spin-Bose systems [J. Math. Phys. 27, 221 (1986)]

G. J. Papadopoulos
Department of Physics, Laboratory of Mechanics, University of Athens, Panepistimiopolis, GR.157 71, Zografos, Athens, Greece

(Received 19 December 1985; accepted for publication 14 January 1986)

The final sentence of the second paragraph on p. 223, left column, should be deleted and the first sentence of the third paragraph should be replaced by the following.

It should be noted that in pursuing evaluations using angular momentum operators there will correspond for each value of $J(=\frac{1}{2},1,\frac{3}{2},...)$ the same pair of complex variables β,γ. This mode of description, given a fixed number of atoms, essentially gives each atomic state by its upper level occupancy.

