Projector bases and algebraic spinors

G. Bergdolt
Centre de Recherches Nucleaires, BP20, 67037 Strasbourg Cedex, France

(Received 31 March 1988; accepted for publication 27 July 1988)

In the case of complex Clifford algebras a basis is constructed whose elements satisfy projector relations. The relations are sufficient conditions for the elements to span minimal ideals and hence to define algebraic spinors.

I. INTRODUCTION

Clifford algebras have been used in theoretical physics in a number of applications[1–3] (Ref. 1 contains a list of references on the subject). In the geometric algebra type of application, Clifford algebra is used to express geometric transformations on a linear space. Letting a be an invertible element of the algebra, the group of invertible elements acts on the algebra by means of the transformation $x \to axa^{-1}$, where x is an arbitrary element. Clifford algebra is a graded algebra, and the underlying vector space is the grade 1 part. Invertible elements that leave the grade 1 part invariant form the Clifford group, which contains physically important groups, such as the orthogonal group, as subgroups. Elements belonging to even-grade parts of the algebra form a subalgebra, and in Ref. 4 spinors are defined as members of this subalgebra. The geometric interpretation of those spinors is that they act on the vector space as dilations rotations. Algebraic spinors are defined as elements of minimal left ideals; as shown in Sec. III this is equivalent to the original definition of the concept.[5] Minimal ideals of an algebra may be viewed as representation spaces for irreducible representations, and therefore algebraic spinors belong to an irreducible representation of the Clifford algebra.

Let C_n denote the Clifford algebra based on a complex n-dimensional vector space. Complexifying vector spaces is not in the spirit of geometrical algebra, but the basis used as the starting point in Sec. II is obtained from a basis with arbitrary signature by means of Witt's decomposition. The decomposition of C_n algebras into minimal ideals described below is constructive. Projectors are not postulated but expressed as Clifford products of isotropic basis vectors, a construction due to Schönberg.[6] The results concerning irreducible representations obtained below are not new. Deriving them in the framework of projector bases shows that the latter span minimal ideals. As all minimal left ideals yield equivalent representations, a fact which becomes apparent in the present approach is usually overlooked: any element of the algebra may be decomposed into a sum of algebraic spinors belonging to a set of minimal ideals. Left multiplication leaves left ideals invariant. The situation is reminiscent of irreducible subspaces under the action of a group. In the case of groups this leads to conservation laws; whether the same is true for algebras is a question left for further investigations. As an example testing the validity of the expansions obtained, it is shown that Cartan's matrix representation and defining relation for spinors are recovered in this way. Section II contains the treatment of C_{2m} Clifford algebras based on even-dimensional complex vector spaces, and Sec. III extends the results to C_{2m+1} Clifford algebras based on odd-dimensional complex vector spaces. The relation to Cartan's theory of spinors is described in Sec. IV.

II. C_{2m} ALGEBRAS

A set of basis vectors may be found such that

$$e_i \cdot e_j = \delta_{ij},$$
$$e'_i \cdot e'_j = -\delta_{ij},$$
$$e_i \cdot e'_j = 0, \quad i,j = 1,\dots,m.$$

The set of $2m$ vectors I_i, I^i defined by

$$I_i = \tfrac{1}{2}(e_i + e'_i),$$
$$I^i = \tfrac{1}{2}(e_i - e'_i)$$

are new basis vectors satisfying the scalar product relations

$$I_i \cdot I_j = 0,$$
$$I^i \cdot I^j = 0,$$
$$I_i \cdot I^j = \tfrac{1}{2}\delta^j_i.$$

Expressed as Clifford products the relations above are

$$I_i I_j + I_j I_i = 0,$$
$$I^i I^j + I^j I^i = 0, \tag{1}$$
$$I_i I^j + I^j I_i = \delta^j_i.$$

The vectors I_i and I^i span isotropic subspaces, and the duality of the two sets is indicated by subscripts and superscripts. From the relations above it follows that, restricted to the isotropic subspaces, the Clifford algebra reduces to a Grassman algebra. The element

$$I^{1\cdots m} = I^1 I^2 \cdots I^m$$

has the property

$$I^i I^{1\cdots m} = 0, \tag{2}$$

for all vectors I^i. The same is valid in the dual space. As the basis I_i, I^i is not orthogonal, we have to demonstrate a preliminary proposition.

Proposition A: Letting $e_1,\dots,e_n$ be the basis vectors of an n-dimensional vector space, the 2^n ordered Clifford products $e_{i_1} e_{i_2} \cdots e_{i_m}$ with $i_1 < i_2 \cdots < i_m$ and $m = 0,\dots,n$ are a basis of the algebra.

It is well known that exterior products of basis vectors are a basis of the algebra. In the case of orthogonal vectors Clifford products of vectors coincide with exterior products, and the proposition is trivial. In the general case the proposition results from the following relations, given without proof.

 0022-2488/88/122519-04$02.50

Proposition B: The Clifford product of r vectors can be expressed as the sum of all possible contractions of the exterior product:

$$X_1X_2\cdots X_r$$
$$=\Big(1+\sum_{(ij)} C_{ij}+\sum_{(ij)(kl)} C_{kl}C_{ij}+\cdots\Big)X_1\Lambda X_2\cdots\Lambda X_r.$$

The contraction C_{ij} is defined as

$$C_{ij}(X_1\Lambda X_2\cdots\Lambda X_r)$$
$$=(-)^{\nu}(X_i\cdot X_j)X_1\Lambda X_2\cdots\hat{X}_i\cdots\hat{X}_j\cdots\Lambda X_r,$$

where ν is the number of vectors between X_i and X_j, and $\hat{X}_i, \hat{X}_j$ indicates omission of the vectors. A converse of Proposition B is valid.

Proposition C: The exterior product of r vectors can be expressed as the sum of all possible anticontractions of the Clifford product:

$$X_1\Lambda X_2\cdots\Lambda X_r$$
$$=\Big(1+\sum_{(ij)} \overline{C_{ij}}+\sum_{(ij)(kl)} \overline{C_{kl}C_{ij}}+\cdots\Big)X_1X_2\cdots X_r.$$

The anticontraction is defined as a contraction with opposite sign factor.

According to Proposition A, a basis of the algebra is given by the ordered Clifford products

$$I^{i_r\cdots i_1}I_{j_1\cdots j_s}=I^{i_r}\cdots I^{i_1}I_{j_1}\cdots I_{j_s},\tag{3}$$

where subscripts are increasing, superscripts decreasing. Since the vectors in the subset I_i are orthogonal, $I_{j_1\cdots j_s}$ is an exterior product, and the same holds for $I^{i_r\cdots i_1}$. Following Schönberg[6] introduce elements of the algebra defined by

$$P^{i_1\cdots i_r}_{j_1\cdots j_s}=I^{i_r\cdots i_1}I_{1\cdots m}I^{m\cdots 1}I_{j_1\cdots j_s}.\tag{4}$$

We call these elements projectors to describe their properties outlined below. To simplify the notation let α, β be multi-indices, i.e., an ordered subset of $\{1,2,...,m\}$. By convention the ordering in superscripts is decreasing. The following propositions is crucial or the development below.

Proposition D: The projectors P^{α}_{β} satisfy the relations

$$P^{\alpha}_{\beta}P^{\mu}_{\nu}=\delta^{\mu}_{\beta}P^{\alpha}_{\nu},\tag{5}$$

$$\sum_{\alpha} P^{\alpha}_{\alpha}=1.\tag{6}$$

Relation (5), given by Schönberg,[6] may be obtained as follows. Let P be the projector

$$P=I_{1\cdots m}I^{m\cdots 1}=I_1I^1I_2I^2\cdots I_mI^m.$$

Since P is the product of commuting idempotent elements I_iI^i, P is idempotent: $P^2=P$.

Letting $\mu=1,...,m$ denote the full set of indices, from relations (1), (2), and the dual of (2) the intermediate result

$$I^{\mu}I_{\beta}I^{\gamma}I_{\mu}=\delta^{\gamma}_{\beta}I^{\mu}I_{\mu}$$

is obtained. Relation (5) follows then from the fact that P is idempotent.

Relation (6) may be derived as follows: from (1) it follows that

$$P=\sum_{\alpha}(-)^{\nu(\alpha)}I^{\alpha}I_{\alpha},$$

where $\nu(\alpha)$ is the number of indices in the subset α. The sum includes the null set for which $\nu(\alpha)=0$, $I^{\alpha}=I_{\alpha}=1$. We have

$$\sum_{\beta}P^{\beta}_{\beta}=\sum_{\beta}I^{\beta}PI_{\beta}=\sum_{\alpha,\beta}(-)^{\nu(\alpha)}I^{\beta}I^{\alpha}I_{\alpha}I_{\beta};$$

only terms with $\alpha\cap\beta=0$ are nonzero. Consider the sum of terms such that $\alpha\cup\beta=\gamma$:

$$\sum_{\beta}P^{\beta}_{\beta}=\sum_{\gamma=0}^{\mu}I^{\gamma}I_{\gamma}\sum_{\alpha=0}^{\gamma}(-)^{\nu(\alpha)}.$$

The rearrangement of indices to obtain an ordered set does not produce a sign change as it is done symmetrically in subscripts and superscripts. The proof is completed by realizing that for a finite set, the number of subsets with an even number of elements is equal to the number of subsets with an odd number, so that $\Sigma^{\gamma}_{\alpha=0}(-)^{\nu(\alpha)}=0$ except for $\gamma=0$, where the result is 1. The proposition about finite sets used above may be derived by induction. The proposition is valid for one-element sets that have two improper subsets: the null set and the full set.

To demonstrate the basis property of the projectors, we express them as linear combinations of ordered products and conversely. We have

$$P^{\alpha}_{\beta}=I^{\alpha}PI_{\beta}=\sum_{\gamma}(-)^{\nu(\gamma)}I^{\alpha}I^{\gamma}I_{\gamma}I_{\beta}$$
$$=\sum_{\gamma}(-)^{\nu(\gamma)}I^{\gamma\cup\alpha}I_{\gamma\cup\beta}.$$

Nonzero terms are those with $\gamma\cap\alpha=0$ and $\gamma\cap\beta=0$. Reordering of indices in the antisymmetric exterior products $I^{\gamma\cup\alpha}$ and $I_{\gamma\cup\beta}$ may be required to obtain ordered products. The converse relations are obtained using relation (6):

$$I^{\beta}I_{\gamma}=\sum_{\alpha}I^{\beta}P^{\alpha}_{\alpha}I_{\gamma}=\sum_{\alpha}I^{\beta}I^{\alpha}PI_{\alpha}I_{\gamma}=\sum_{\alpha}P^{\alpha\cup\beta}_{\alpha\cup\gamma}.\tag{7}$$

Conditions for nonzero terms and rearrangements are as before. Let X and Y be elements of the algebra expanded in a projector basis:

$$X=X^{\alpha}_{\beta}P^{\beta}_{\alpha},\quad Y=Y^{\gamma}_{\delta}P^{\delta}_{\gamma},$$

where $X^{\alpha}_{\beta}, Y^{\gamma}_{\delta}$ are complex numbers and the summation convention is used. From relation (5)

$$XY=(X^{\alpha}_{\beta}Y^{\gamma}_{\alpha})P^{\beta}_{\gamma}.$$

Let $C(2^m)$ designate a $2^m\times 2^m$ complex matrix. We have the following proposition.

Proposition E: C_{2m} algebras admit a $C(2^m)$ matrix representation.

The second consequence of relation (5) is stated as follows.

Proposition F: The idempotent projector P^{α}_{α} generates a minimal left ideal spanned by the projectors P^{β}_{α} with α fixed.

Let α_0 denote a fixed value of the multi-index α (the summation convention does not apply to α_0). We have

$$XP^{\alpha_0}_{\alpha_0}=X^{\gamma}_{\beta}P^{\beta}_{\gamma}P^{\alpha_0}_{\alpha_0}=X^{\alpha_0}_{\beta}P^{\beta}_{\alpha_0}.$$

Letting A denote the algebra, algebraic spinor members of the minimal left ideal $AP^{\alpha_0}_{\alpha_0}$ may then be expanded as

$$\eta = \eta_\beta P^\beta_{\alpha_0}.$$

To $X \in A$ the mapping $\rho(X): AP^{\alpha_0}_{\alpha_0} \to AP^{\alpha_0}_{\alpha_0}$ defined by $\eta \to X\eta$ can be associated; $\rho(X)$ is a representation of the algebra $\rho(XY) = \rho(X)\rho(Y)$. The representation is irreducible since $AP^{\alpha_0}_{\alpha_0}$ is minimal. A minimal ideal is generated by each idempotent projector P^β_β. The basis property of the projectors shows that the ideals do not overlap, and relation (6) that the algebra is completely reduced. All representations defined by the minimal ideals are equivalent, since a one-to-one mapping $AP^\alpha_\alpha \to AP^\beta_\beta$ is defined by $\eta \to \eta P^\alpha_\beta$. We have thus recovered a theorem already given by Weyl,[7] which states that central simple algebras have, up to equivalence, one irreducible representation contained in the regular representation with a multiplicity equal to the dimensions of the former. [The regular representation is the mapping $\rho(a): A \to A$ defined by $X \to aX$, and the representation space is obviously identical to the algebra.]

III. C_{2m+1} ALGEBRAS

There is now an unpaired basis vector I_0 with the following scalar products:

$$I_0^2 = 1,$$
$$I_0 \cdot I_i = 0,$$
$$I_0 \cdot I^i = 0.$$

Introduce the elements

$$I_\pm = \tfrac{1}{2}(1 \pm I_0).$$

It is easily seen that the following relations are satisfied:

$$I_\pm^2 = I_\pm,$$
$$I_+ I_- = 0 = I_- I_+.$$

The products above are Clifford products. Let ω be an index with values $+$, $-$. As linear combinations of the elements of an ordered product basis, the 2^n elements $I^\alpha I_\omega I_\beta$ are a basis of the algebra. We define a set of projectors

$$P^\alpha_{\omega\beta} = I^\alpha I_\mu I_\omega I^\mu I_\beta. \tag{8}$$

By straightforward computation it can be checked that the projectors satisfy the relations

$$P^{\alpha'}_{\omega'\beta'} P^{\alpha''}_{\omega''\beta''} = \delta_{\omega',\omega''} \delta^{\alpha''}_{\beta'} P^{\alpha'}_{\omega'\beta'} \tag{9}$$

and

$$\sum_{\omega,\alpha} P^\alpha_{\omega\alpha} = 1.$$

The basis property of the projectors derives from the relations to the elements of an ordered product basis,

$$P^\beta_{\omega\gamma} = \sum_\alpha (-)^{\nu(\alpha)} I^{\alpha\cup\beta} I_{\omega\epsilon(\alpha)} I_{\alpha\cup\gamma},$$
$$I^\beta I_\omega I_\gamma = \sum_\alpha P^{\alpha\cup\beta}_{\omega\epsilon(\alpha)\alpha\cup\gamma}, \tag{10}$$

where $\epsilon(\alpha) = (-)^{m+\nu(\alpha)}$ is a sign factor. From relation (9) it follows that the projectors $P^\alpha_{\omega\beta}$ with ω fixed span a subalgebra, and that C_{2m+1} is the direct sum of the two subalgebras. The relation among components is now

$$(X_\beta^{+\alpha}, X_\beta^{-\alpha})(Y_\delta^{+\gamma}, Y_\delta^{-\gamma}) = (X_\beta^{+\alpha} Y_\alpha^{+\gamma}, X_\beta^{-\alpha} Y_\alpha^{-\gamma}),$$

and we have the following proposition.

Proposition G: C_{2m+1} algebras admit $C(2^m) \otimes C(2^m)$ matrix representations.

In this case, the proposition about minimal ideals becomes the following.

Proposition H: The idempotent projector $P^\alpha_{\omega\alpha}$ generates a minimal left ideal spanned by the projectors $P^\beta_{\omega\alpha}$ with ω and α fixed.

The theorem about irreducible representations used for C_{2m} algebras cannot be applied here, since C_{2m+1} algebras are not central and simple. According to the proposition above we have that C_{2m+1} algebras have two irreducible representations of dimension 2^m contained in the regular representation with multiplicity 2^m.

IV. CONNECTION WITH CARTAN'S THEORY OF SPINORS

We show that Cartan's matrix representation is contained in relation (10). Going back to indices, the expansions of the basis vectors are

$$I^i = P^i + \sum_j P^{ji}_j + \sum_{jk} P^{jki}_{jk} + \cdots,$$
$$I_0 = (-)^m \Big(P' - \sum_i P'^i_i + \sum_{ij} P'^{ij}_{ij} - \cdots\Big), \tag{11}$$
$$I_i = P_i + \sum_j P^j_{ji} + \sum_{jk} P^{jk}_{jki} + \cdots.$$

Here $P = P_+ + P_-$ and $P' = P_+ - P_-$. Numerical results for $m = 2$, $n = 5$ are

$$I^1 = P^1 - P^{12}_2,$$
$$I^2 = P^2 + P^{12}_1,$$
$$I_0 = P' - P'^1_1 - P'^2_2 + P'^{12}_{12},$$
$$I_1 = P_1 - P^2_{12},$$
$$I_2 = P_2 + P^1_{12}.$$

If multi-indices are ordered none, 1, 2, 12, the matrices of components relative to the projectors $P^\alpha_{+\beta}$ are

$$H^1 = \begin{Bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{Bmatrix},$$

$$H^2 = \begin{Bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Bmatrix},$$

$$H_0 = \begin{Bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{Bmatrix},$$

$$H_1 = \begin{Bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{Bmatrix},$$

$$H_2 = \begin{Bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Bmatrix}.$$

The matrices are obtained by inspection of the preceding relations and are identical with Cartan's expression[8] (subscripts are identified with Cartan's primed indices). Cartan's system of equations defining spinors[9] is obtained by working out the components of the relation

$$\eta = X\xi,$$

where X is a vector,

$$X = X_i I^i + X^0 I_0 + X^i I_i,$$

and ξ, η are algebraic spinors,

$$\eta = \sum_{p=0}^{m} \eta_{i_1 \cdots i_p} P_+^{i_1 \cdots i_p},$$

$$\xi = \sum_{p=0}^{m} \xi_{i_1 \cdots i_p} P_+^{i_1 \cdots i_p}.$$

The result is obtained using (11) and (8):

$$\eta_{i_1 \cdots i_p} = \sum_{k=1}^{p} (-)^{p-k} X_{i_k} \xi_{i_1 \cdots \hat{i}_k \cdots i_p} + (-)^p X^0 \xi_{i_1 \cdots i_p} + \sum_j X^j \xi_{i_1 \cdots i_p j}.$$

This is Cartan's relation. However, Cartan's theory goes beyond the relation given above. The concept of pure spinor is defined by nonlinear constraints on the components of algebraic spinors, and lies outside of the theory of irreducible representations.

V. CONCLUSION

The construction of projector bases leads to the results of irreducible representation theory by a simple algebraic approach. All quantities involved in the approach, projectors and spinors, are members of the algebra. A projector basis is obviously the basis appropriate to calculations involving algebraic spinors.

[1]N. Salingaros and Y. Ilamed, Found. Phys. **14**, 777 (1984).
[2]K. R. Greider, Found. Phys. **14**, 467 (1984).
[3]Z. Hasiewicz, A. K. Kwansniewski, and P. Moraviec, J. Math. Phys. **25**, 2031 (1984).
[4]D. Hestenes, Am. J. Phys. **39**, 1013 (1971).
[5]E. Cartan, *Theory of Spinors* (Herrman, Paris, 1966).
[6]M. Schönberg, Nuovo Cimento Suppl. **6**, 356 (1957).
[7]H. Weyl, the *Theory of Groups and Quantum Mechanics* (Dover, New York, 1950), p. 316.
[8]See Ref. 5, p. 82.
[9]See Ref. 5, p. 81.

SU(1,1), its connections with SU(2), and the vector model

E. de Prunelé
Service de Physique des Atomes et des Surfaces, Centre d'Etudes Nucléaires de Saclay, 91191 Gif-Sur-Yvette Cedex, France

(Received 3 May 1988; accepted for publication 3 August 1988)

A relation between representation functions (RF's) of positive discrete unitary irreducible representations (UIR's) of SU(1,1) and the RF's of the UIR's of SU(2) is given. The classical vector model is worked out for the positive discrete UIR's of SU(1,1). Classical domains, probability densities, and algorithms for numerical computations of SU(1,1) RF's and Clebsch–Gordan coefficients are derived, in full analogy with the SU(2) case.

I. INTRODUCTION

The SU(1,1) group occurs in many areas of physics. For example, it is locally isomorphic to the three-dimensional Lorentz group SO(2,1); the so(2,1) Lie algebra is also relevant for the hydrogen atom and the isotropic harmonic oscillator.[1–2] The SU(1,1) symmetry is involved in recent publications concerning various subjects (see, e.g., Refs. 3–8): path integration methods, coherent states, squeezed states, laser–plasma scattering, spin wave in solid-state physics, field theory, etc.

In the present paper we consider only positive discrete unitary irreducible representations (UIR's). The restriction to unitary representations is necessary if a classical vector model is used (the Hermitic generators then have real eigenvalues). Once a relation is obtained for positive discrete UIR's, the analogous relation for negative discrete UIR's follows easily. The continuous supplementary and principal series of UIR's are not considered here.

The SU(1,1) Clebsch–Gordan (CG) coefficients (also called Wigner coefficients), associated to the coupling of two discrete UIR's of the same sign, are known to be related to the SU(2) CG coefficients.[9–11] More precisely, these SU(1,1) CG coefficients can be expressed as SU(2) CG coefficients, their arguments being given functions of those of the SU(1,1) CG coefficients. These relations, which take their simplest form when expressed in terms of $3j$ coefficients, correspond[9] to symmetry properties of generalized CG coefficients. Generalized CG coefficients were introduced[12,13] to take into account analyticity properties of the CG coefficients when their arguments move in the complex plane.

To our knowledge, however, no such relations between representation functions (RF's) of SU(1,1) and RF's of SU(2) (also called rotation matrix elements) have been given up to now. The RF's of SU(1,1) are known to be analytically related to RF's of SU(2). Thus, otherwise stated, an appropriate symmetry relation for generalized RF's appears to be still lacking in the literature.

In the first part of this paper, Sec. II, it is pointed out that such a relation [see Eq. (23)] can easily be obtained from the explicit expressions of the RF's.

The vector model for SU(2) is extremely useful, for both physical intuition and practical calculations.[14–19] Despite the close interconnections between SU(1,1) and SU(2) discussed above, it appears desirable to develop the vector model for SU(1,1) independently of its relations to SU(2). This provides a direct and deeper insight into the relation between the SU(1,1) discrete UIR's and the classical vectors in three-dimensional space. The vector model for SU(1,1) will be shown (Sec. IV) to provide suitable algorithms for numerical computation of RF and CG coefficients based on three term recursion relations (Sec. III), in full analogy with previous works[18,19] for SU(2). When only a few SU(1,1) RF's are needed, one can express them as SU(2) RF's and use the SU(2) algorithms. In physical applications involving the SU(1,1) symmetry, the simultaneous computation of a whole set of RF's differing only by one SU(1,1) argument is generally required. When a large number of such SU(1,1) RF's are needed, the SU(1,1) algorithms presented below become very appropriate. Indeed the SU(2) algorithms correspond to the simultaneous variation of several SU(1,1) RF arguments. The SU(1,1) algorithms are, in fact, quite similar to the SU(2) ones. The main differences occur in the values of the arguments of the coefficients involved in the three term recursion relation. For SU(1,1) CG coefficients the SU(2) algorithms[18] are directly appropriate since the variation of one SU(2) independent argument also corresponds to the variation of only one SU(1,1) independent argument in the three term recursion relations considered.

Section III is devoted to the derivation of the recursion relations needed in Sec. IV. There are several methods for deriving these recursion relations. The methods presented here avoid the use of explicit analytical expressions for RF and CG coefficients. They start only from the defining commutation relations for the Hermitic generators. They are appropriate to both SU(2) and SU(1,1) cases. These methods are not new, except the one given in Sec. III B 1. We believe that it is useful to present a unified and direct way for deriving all these recursion relations, with a complete exposition of the phase conventions.

In Sec. IV we present the vector model for SU(1,1). The classical domains and probability are determined, as well as the algorithm for numerical computation of SU(1,1) RF's and CG coefficients.

II. SU(1,1) REPRESENTATION FUNCTIONS IN TERMS OF SU(2) ROTATION MATRICES

A. Notation and definitions

The h and g denote diagonal matrices associated with SU(2) and SU(1,1), respectively:

$$h^{11} = h^{22} = h^{33} = -g^{11} = -g^{22} = g^{33} = 1. \tag{1}$$

The covariant components h_{jk}, g_{jk} corresponding to the inverse matrices are equal to the contravariant components defined in Eq. (1). The J denote the generators of SU(2), the K those of SU(1,1). Their commutation relations can be cast into the form[20]

$$[J^j, J^k] = i\eta^{jkl} J_l \,, \tag{2}$$

$$J^l = h^{lj} J_j = J_l \,, \tag{3}$$

$$[K^j, K^k] = -i\epsilon^{jkl} K_l \,, \tag{4}$$

$$K^l = g^{lj} K_j \,, \tag{5}$$

where summation occurs for repeated covariant and contravariant indices, and both η and ϵ denote the completely antisymmetric tensor ($\eta^{123} = \epsilon^{123} = 1$). Two different symbols have been used since the covariant components of η are obtained with the metric h whereas those of ϵ are obtained with the metric g. Here, $J \cdot J$ and $K \cdot K$ are the Casimir operators of SU(2) and SU(1,1), respectively:

$$J \cdot J = J^k J_k = (J^1)^2 + (J^2)^2 + (J^3)^2 \,, \tag{6}$$

$$K \cdot K = K^k K_k = -(K^1)^2 - (K^2)^2 + (K^3)^2 \,. \tag{7}$$

The orthonormal basis vectors of the UIR's of SU(2) are denoted by $|j,m\rangle$, j being a positive integer or half-integer; $m = -j, -j+1, \ldots, j$. We have

$$J \cdot J \, |j,m\rangle = j(j+1) |j,m\rangle \,, \tag{8}$$

$$J^3 |j,m\rangle = m |j,m\rangle \,. \tag{9}$$

The orthonormal basis vectors of the positive discrete UIR's of SU(1,1) are denoted by $|\gamma,\mu)$, γ being a negative integer or half-integer; $\mu = -\gamma, -\gamma+1, \ldots$ up to infinity. [The notation $|\gamma,\mu)$ is used in place of $|\gamma,\mu\rangle$ in order to avoid confusion in Sec. III B 1.] We have

$$K \cdot K \, |\gamma,\mu) = \gamma(\gamma+1) |\gamma,\mu) \,, \tag{10}$$

$$K^3 |\gamma,\mu) = \mu |\gamma,\mu) \,. \tag{11}$$

Defining

$$J^{\pm} = J^1 \pm i J^2 \,, \tag{12}$$

$$K^{\pm} = K^1 \pm i K^2 \,, \tag{13}$$

one has, with the usual phase conventions,[1]

$$J^{\pm} |j,m\rangle = ((j \pm m + 1)(j \mp m))^{1/2} |j, m \pm 1\rangle = c(\pm m, j) |j, m \pm 1\rangle \,, \tag{14}$$

$$K^{\pm} |\gamma,\mu) = (-(\gamma \pm \mu + 1)(\gamma \mp \mu))^{1/2} |\gamma, \mu \pm 1) = c(\gamma, \pm\mu) |\gamma, \mu \pm 1) \,, \tag{15}$$

with

$$c(a,b) \equiv ((b + a + 1)(b - a))^{1/2} = c(a, -b-1) \,. \tag{16}$$

The advantage of the above phase conventions for raising and lowering operators is that both the d SU(2) RF's and δ SU(1,1) RF's defined below are real. The disadvantage is that the action of K cannot be viewed as the analytic continuation of the action of J, due to the minus sign inside the square root of Eq. (15), or, equivalently, due to the interchange of the arguments of the function c in Eqs. (14) and (15). When using the Euler angle parametrization, the RF's of SU(2) and SU(1,1) are given by

$$\langle j,m' | \exp(-i\alpha' J^3) \exp(-i\beta J^2) \exp(-i\alpha J^3) | j,m\rangle = \exp(-i(\alpha' m' + \alpha m)) d^j_{m',m}(\beta) \,, \tag{17}$$

$$(\gamma,\mu' | \exp(-i\alpha' K^3) \exp(-i\beta K^2) \exp(-i\alpha K^3) | \gamma,\mu) = \exp(-i(\alpha'\mu' + \alpha\mu)) \delta^{\gamma}_{\mu',\mu}(\beta) \,. \tag{18}$$

A relation between the d and δ RF's will now be established.

B. Relation between d and δ RF's

The phase conventions fixed by Eqs. (14) and (15) completely determine the matrices d and δ. Their explicit expressions can be derived by several methods.[16] For $m' + m \geqslant 0$, d can be expressed in terms of the hypergeometric function:

$$\begin{aligned} d^j_{m',m}(\beta) &= [(-1)^{j-m}/(m'+m)!]((j+m')!(j+m)! \\ &\quad \times ((j-m')!(j-m)!)^{-1})^{1/2} \sin^{2j}(\beta/2) \\ &\quad \times \cot^{m'+m}(\beta/2)\, {}_2F_1(m'-j, m-j; m'+m \\ &\quad + 1; -\cot^2(\beta/2)) \,. \end{aligned} \tag{19}$$

Similarly δ can be expressed[2,21,22] as

$$\begin{aligned} \delta^{\gamma}_{\mu',\mu}(\beta) &= (-1)^{\gamma+\mu'}(\mu'+\mu-1)!((\gamma+\mu')!(\gamma+\mu)! \\ &\quad \times (-\gamma+\mu'-1)!(-\gamma+\mu-1)!)^{-1/2} \\ &\quad \times \sinh^{2\gamma}(\beta/2) \tanh^{\mu'+\mu}(\beta/2)\, {}_2F_1(-\mu'-\gamma, \\ &\quad -\mu-\gamma; -\mu'-\mu+1; \coth^2(\beta/2)) \,. \end{aligned} \tag{20}$$

One then uses the transformation

$$\begin{aligned} &{}_2F_1(-\mu'-\gamma, -\mu-\gamma; -\mu'-\mu+1; \coth^2(\beta/2)) \\ &\quad = (\mu'-\gamma-1)!(\mu-\gamma-1)! \\ &\quad \times ((\mu'+\mu-1)!(-2\gamma-1)!)^{-1} \\ &\quad \times {}_2F_1(-\gamma-\mu', -\gamma-\mu; \\ &\quad -2\gamma; 1-\coth^2(\beta/2)) \,. \end{aligned} \tag{21}$$

This transformation, which can be verified by expressing explicitly the hypergeometric functions in terms of polynomials, is a particular case of a more general transformation valid for other values of the arguments.[23] The last argument of ${}_2F_1$ in the right-hand side of Eq. (21) is now always negative, as in Eq. (19). A direct comparison between Eq. (19) and Eq. (20) can now be performed. Using the relation[16]

$$d^j_{m',m}(\beta) = (-1)^{m'-m} d^j_{m,m'}(\beta) \,, \tag{22}$$

one finally obtains

$$\delta^{\gamma}_{\mu',\mu}(\beta) = (\cosh(\beta/2))^{-1} d^{(\mu'+\mu-1)/2}_{-\gamma+(\mu'-\mu-1)/2,\, -\gamma+(\mu-\mu'-1)/2}(2\arctan(\sinh(\beta/2))) \,. \tag{23}$$

III. THREE TERM RECURSION RELATIONS FOR SU(1,1)

A. Recursion relations within a given UIR

Three term recursion relations within a given positive discrete UIR are now obtained exactly as for SU(2).

1. Representation functions

For RF's one starts from the following equality:

$$(\gamma,\mu'|\exp(-i\beta K^2)K^3|\gamma,\mu)=\mu\delta^{\gamma}_{\mu',\mu}(\beta)\,. \quad (24)$$

Then, using the Baker–Campbell–Hausdorff relation and Eq. (4), one obtains

$$\exp(-i\beta K^2)K^3 = (K^3\cosh(\beta)+((K^+ + K^-)/2)\sinh(\beta)) \times\exp(-i\beta K^2)\,, \quad (25)$$

and one easily obtains the recursion relation

$$\delta^{\gamma}_{\mu',\mu}(\beta)(\mu-\mu'\cosh(\beta)) = \sinh(\beta)(c(\gamma,-\mu')\delta^{\gamma}_{\mu'-1,\mu}(\beta) + c(\gamma,\mu')\delta^{\gamma}_{\mu'+1,\mu}(\beta))/2\,. \quad (26)$$

The recursion relation for the SU(2) rotation matrix d is obtained in a similar way and is

$$d^{j}_{m',m}(\beta)(m-m'\cos(\beta)) = \sin(\beta)(c(-m',j)d^{j}_{m'-1,m}(\beta) + c(m',j)d^{j}_{m'+1,m}(\beta))/2\,. \quad (27)$$

2. Clebsch–Gordan coefficients

The SU(1,1) CG coefficients coupling two positive discrete UIR's are matrix elements of a unitary transformation and therefore satisfy

$$|(\gamma_a\gamma_b)\gamma,\mu)=\sum_{\mu_a,\mu_b}|\gamma_a,\mu_a;\gamma_b,\mu_b)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu)\,, \quad (28)$$

$$|\gamma_a,\mu_a;\gamma_b,\mu_b)=\sum_{\gamma,\mu}|(\gamma_a\gamma_b)\gamma,\mu)(\gamma,\mu|\gamma_a,\mu_a;\gamma_b,\mu_b)\,, \quad (29)$$

$$\sum_{\mu_a,\mu_b}(\gamma',\mu'|\gamma_a,\mu_a;\gamma_b,\mu_b)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu)=\delta_{\gamma'\gamma}\delta_{\mu'\mu}\,, \quad (30)$$

$$\sum_{\gamma,\mu}(\gamma_a,\mu_a';\gamma_b,\mu_b'|\gamma,\mu)(\gamma,\mu|\gamma_a,\mu_a;\gamma_b,\mu_b)=\delta_{\mu_a'\mu_a}\delta_{\mu_b'\mu_b}\,, \quad (31)$$

$$(\gamma,\mu|\gamma_a,\mu_a;\gamma_b,\mu_b)=(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu)^*\,. \quad (32)$$

The following relations are always implicit for the CG coefficients:

$$\mu_a+\mu_b=\mu,\quad \mu_a\geqslant-\gamma_a,\quad \mu_b\geqslant-\gamma_b,\quad \gamma_a+\gamma_b\geqslant\gamma\,.$$

A three term recursion relation is most easily obtained from the action of the operator K^-K^+, with

$$K^{\pm}=K_a^{\pm}+K_b^{\pm}\,, \quad (33)$$

on Eq. (28).

A simple calculation leads to the recursion relation

$$(c^2(\gamma,\mu)-c^2(\gamma_a,\mu_a)-c^2(\gamma_b,\mu_b))(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu) = c(\gamma_a,\mu_a)c(\gamma_b,\mu_b-1)(\gamma_a,\mu_a+1;\gamma_b,\mu_b-1|\gamma,\mu) + c(\gamma_b,\mu_b)c(\gamma_a,\mu_a-1) \times(\gamma_a,\mu_a-1;\gamma_b,\mu_b+1|\gamma,\mu)\,. \quad (34)$$

In a similar way, one obtains, for SU(2),

$$(c^2(m,j)-c^2(m_a,j_a)-c^2(m_b,j_b))\langle j_a,m_a;j_b,m_b|j,m\rangle = c(m_a,j_a)c(m_b-1,j_b) \times\langle j_a,m_a+1;j_b,m_b-1|j,m\rangle + c(m_b,j_b)c(m_a-1,j_a) \times\langle j_a,m_a-1;j_b,m_b+1|j,m\rangle\,. \quad (35)$$

B. Recursion relations between different UIR's

1. Representation functions

Recursion relations for SU(2) RF's involving different values of j are most easily derived from the general relation[16]

$$d^{j_a}_{m_a',m_a}(\beta)d^{j_b}_{m_b',m_b}(\beta) = \sum_j\langle j_a,m_a';j_b,m_b'|j,m'\rangle \times\langle j,m|j_a,m_a;j_b,m_b\rangle d^{j}_{m',m}(\beta)\,. \quad (36)$$

For $j_b=1$, $m_b'=m_b=0$, the above equation directly gives a three term recursion relation for the SU(2) RF's d where only j varies, m' and m being fixed. The analog of Eq. (36) for SU(1,1),

$$\delta^{\gamma_a}_{\mu_a',\mu_a}(\beta)\delta^{\gamma_b}_{\mu_b',\mu_b}(\beta) = \sum_{\gamma}(\gamma_a,\mu_a';\gamma_b,\mu_b'|\gamma,\mu') \times(\gamma,\mu|\gamma_a,\mu_a;\gamma_b,\mu_b)\delta^{\gamma}_{\mu',\mu}(\beta)\,, \quad (37)$$

however, does not lead to a three term recursion relation. The reason is that γ on the right-hand side of Eq. (37) varies betwen the maximum value of $-\mu'$, $-\mu$ up to $\gamma_a+\gamma_b$.

A method valid for both SU(2) and SU(1,1), which avoids the use of the CG coefficients, is now described. The starting point is the bosonic realization of both SU(2) and SU(1,1),[24]

$$[\bar a_1,a_1]=[\bar a_2,a_2]=1\,, \quad (38)$$

all other commutators being equal to zero. Here $\bar a$ is the annihilation operator. Its adjoint is the creation operator a. One has

$$J^3=(a_1\bar a_1-a_2\bar a_2)/2\,, \quad (39)$$

$$J^+=a_1\bar a_2\,, \quad (40)$$

$$J^-=a_2\bar a_1\,, \quad (41)$$

$$K^3=(a_1\bar a_1+\bar a_2a_2)/2\,, \quad (42)$$

$$K^+=a_1a_2\,, \quad (43)$$

$$K^-=\bar a_2\bar a_1 \quad (44)$$

$$|j,m\rangle=((j+m)!(j-m)!)^{-1/2}a_1^{j+m}a_2^{j-m}|0\rangle\,. \quad (45)$$

It follows that

$$K^3|j,m\rangle=(j+\tfrac12)|j,m\rangle\,, \quad (46)$$

$$K^+|j,m\rangle = ((j+1+m)(j+1-m))^{1/2} \times |j+1,m\rangle, \quad (47)$$

$$K^-|j,m\rangle = ((j+m)(j-m))^{1/2}|j-1,m\rangle. \quad (48)$$

From now on, it is supposed that m is negative or zero, and that γ,μ are associated to j and m according to the involutive transformation

$$\gamma = m - \tfrac{1}{2}, \quad \mu = j + \tfrac{1}{2}. \quad (49)$$

Thus Eq. (45) can equivalently be rewritten as

$$|j,m\rangle = ((\gamma+\mu)!(\mu-\gamma-1)!)^{-1/2} \times a_1^{\gamma+\mu} a_2^{\mu-\gamma-1}|0\rangle \equiv |\gamma,\mu\rangle. \quad (50)$$

The SU(1,1) recursion relation is finally obtained by evaluating the following matrix elements, denoted by A and B:

$$A \equiv (\gamma,\mu'|J^+ \exp(-i\beta K^2)J^-|\gamma,\mu), \quad (51)$$

$$B \equiv (\gamma,\mu'|J^- \exp(-i\beta K^2)J^+|\gamma,\mu). \quad (52)$$

A direct evaluation yields

$$A = ((\gamma-\mu')(\gamma+\mu')(\gamma-\mu)(\gamma+\mu))^{1/2}\delta^{\gamma-1}_{\mu',\mu}(\beta). \quad (53)$$

The evaluation of the two commutators

$$[K^2,J^+] = i(a_1a_1 + \bar{a}_2\bar{a}_2)/2, \quad (54)$$

$$[K^2,[K^2,J^+]] = -J^+, \quad (55)$$

allows one to use the Baker–Campbell–Hausdorff relation and to obtain

$$\exp(i\beta K^2)J^+\exp(-i\beta K^2) = \cosh(\beta)J^+ - \sinh(\beta)(a_1a_1+\bar{a}_2\bar{a}_2)/2. \quad (56)$$

Using the additional relations

$$a_1a_1|\gamma-1,\mu) = ((\gamma+\mu+1)(\gamma+\mu))^{1/2}|\gamma,\mu+1), \quad (57)$$

$$\bar{a}_2\bar{a}_2|\gamma-1,\mu) = ((\mu-\gamma)(\mu-\gamma-1))^{1/2}|\gamma,\mu-1), \quad (58)$$

one obtains

$$A = -(\gamma-\mu)(\gamma+\mu)\cosh(\beta)\delta^{\gamma}_{\mu',\mu}(\beta) - \sinh(\beta)(-(\gamma-\mu)(\gamma+\mu))^{1/2}(((\gamma+\mu+1) \times (\gamma+\mu))^{1/2}\delta^{\gamma}_{\mu',\mu+1}(\beta) + ((\mu-\gamma) \times (\mu-\gamma-1))^{1/2}\delta^{\gamma}_{\mu,\mu-1}(\beta))/2. \quad (59)$$

Following a procedure analogous to the one above, one obtains

$$B = ((\gamma+1-\mu')(\gamma+1+\mu')(\gamma+1-\mu) \times (\gamma+1+\mu))^{1/2}\delta^{\gamma+1}_{\mu',\mu}(\beta), \quad (60)$$

$$B = -(\gamma+1-\mu)(\gamma+1+\mu)\cosh(\beta)\delta^{\gamma}_{\mu',\mu}(\beta) - \sinh(\beta)(-(\gamma+1-\mu)(\gamma+1+\mu))^{1/2} \times (((\mu-\gamma)(\mu-\gamma-1))^{1/2}\delta^{\gamma}_{\mu',\mu+1}(\beta) + ((\gamma+1+\mu)(\gamma+\mu))^{1/2}\delta^{\gamma}_{\mu',\mu-1}(\beta)/2. \quad (61)$$

Finally the desired recursion relation is obtained by evaluating $A/\gamma + B/(\gamma+1)$ and using Eq. (26) together with the symmetry property

$$\delta^{\gamma}_{\mu',\mu}(\beta) = \delta^{\gamma}_{\mu,\mu'}(-\beta). \quad (62)$$

The result is

$$(((\gamma-\mu')(\gamma+\mu')(\gamma-\mu)(\gamma+\mu))^{1/2}/\gamma)\delta^{\gamma-1}_{\mu',\mu}(\beta) + (((\gamma+1-\mu')(\gamma+1+\mu')(\gamma+1-\mu) \times (\gamma+1+\mu))^{1/2}/(\gamma+1))\delta^{\gamma+1}_{\mu',\mu}(\beta) = (2\gamma+1)(-\cosh(\beta)+\mu\mu'/(\gamma(\gamma+1)))\delta^{\gamma}_{\mu',\mu}(\beta). \quad (63)$$

A similar method yields, for SU(2),

$$(((j-m')(j+m')(j-m)(j+m))^{1/2}/j)d^{j-1}_{m',m}(\beta) + (((j+1-m')(j+1+m')(j+1-m) \times (j+1+m))^{1/2}/(j+1))d^{j+1}_{m',m}(\beta) = (2j+1)(\cos(\beta) - mm'/(j(j+1)))d^{j}_{m',m}(\beta). \quad (64)$$

2. Clebsch–Gordan coefficients

The method presently used is the analog of the well-known method for SU(2) based on the concept of vector operators (see, e.g., Refs. 1 and 25). Therefore it is only briefly outlined here. We define an SU(1,1) vector operator V as a set of three operators V^j satisfying the following commutation rules:

$$[K^j,V^k] = -i\epsilon^{jkl}V_l. \quad (65)$$

Defining

$$V^{\pm} = V^1 \pm iV^2, \quad (66)$$

one obtains, from Eqs. (4) and (13),

$$[K^{\pm},V^{\mp}] = \mp 2V^3, \quad (67)$$

$$[K^3,V^{\pm}] = \pm V^{\pm}, \quad (68)$$

$$[K^{\pm},V^3] = \mp V^{\pm}, \quad (69)$$

$$[K^+,V^+] = [K^-,V^-] = [K^j,V^j] = 0, \quad (70)$$

$$K\cdot V \equiv K^jV_j = V\cdot K, \quad (71)$$

$$[K\cdot V,K^j] = 0. \quad (72)$$

From Eqs. (70) and (68), one obtains selection rules relative to μ:

$$(\gamma',\mu'|V^3|\gamma,\mu) = 0, \quad \text{if } \mu' \neq \mu, \quad (73)$$

$$(\gamma',\mu'|V^{\pm}|\gamma,\mu) = 0, \quad \text{if } \mu' \neq \mu \pm 1. \quad (74)$$

We now define a vector product rule for two SU(1,1) vector operators A and B:

$$(A\wedge B)^j = -\epsilon^j{}_{kl}A^kB^l. \quad (75)$$

It is easy to verify that if A and B are SU(1,1) vector operators, then $A\wedge B$ is also a SU(1,1) vector operator. One also verifies that

$$[K^j,(A\wedge B)^k] = i(A^jB^k - A^kB^j), \quad (76)$$

$$[K\cdot K,V] = i(V\wedge K - K\wedge V), \quad (77)$$

$$[K\cdot K,K\wedge V] = 2i((K\cdot K)V - (K\cdot V)K), \quad (78)$$

$$[K\cdot K,V\wedge K] = 2i(-V(K\cdot K) + (K\cdot V)K), \quad (79)$$

$$[K\cdot K,[K\cdot K,V]] = 2(V(K\cdot K) - 2(K\cdot V)K + (K\cdot K)V). \quad (80)$$

Taking the matrix elements of the last equation, one obtains

$$((\gamma'-\gamma)^2-1)((\gamma'+\gamma+1)^2-1)(\gamma',\mu'|V|\gamma,\mu)$$
$$=-4(\gamma',\mu'|(K\cdot V)K|\gamma,\mu)$$
$$=-4(\gamma'\|(K\cdot V)\|\gamma)(\gamma',\mu'|K|\gamma,\mu)\,, \quad (81)$$

where the last equation results from the fact that $K\cdot V$ commutes with the K^j, i.e., $K\cdot V$ is an SU(1,1) scalar operator. From the above equation, one easily deduces the following selection rules:

$$(\gamma\|(K\cdot V)\|\gamma)=0,\quad \text{if } \gamma=-1\,, \quad (82)$$

$$(\gamma',\mu'|V|\gamma,\mu)=0,\quad \text{if } \gamma'-\gamma \text{ is different from } 1,0,-1\,. \quad (83)$$

The three terms recursion relation is now derived from the action of K_a^3 on a coupled state, expressed in two different ways:

$$K_a^3|(\gamma_a\gamma_b)\gamma,\mu)=\sum_{\mu_a\mu_b}\mu_a|\gamma_a,\mu_a;\gamma_b,\mu_b)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu)$$
$$=\sum_{\mu_a\mu_b}\sum_{\gamma'}|\gamma_a,\mu_a;\gamma_b,\mu_b)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma',\mu)$$
$$\times((\gamma_a\gamma_b)\gamma',\mu|K_a^3|(\gamma_a\gamma_b)\gamma,\mu)\,. \quad (84)$$

From now on, the symbols $(\gamma_a\gamma_b)$ will be dropped for the sake of clarity. In the above relation, γ' takes the values $\gamma-1$, γ, $\gamma+1$, since K_a is an SU(1,1) vector operator. A three term recursion follows:

$$\mu_a(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu)$$
$$=\sum_{\gamma'}(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma',\mu)(\gamma',\mu|K_a^3|\gamma,\mu)\,. \quad (85)$$

It remains to evaluate the matrix elements of K_a^3 in the above equation. For the case $\gamma'=\gamma$, Eq. (80) gives

$$(\gamma,\mu|K_a^3|\gamma,\mu)$$
$$=\mu(\gamma(\gamma+1)+\gamma_a(\gamma_a+1)-\gamma_b(\gamma_b+1))$$
$$\times(2\gamma(\gamma+1))^{-1}\,. \quad (86)$$

For the general case, one deduces from Eqs. (70) that the μ dependence of the matrix elements of V^- can be factorized as follows:

$$(\gamma,\mu|V^-|\gamma,\mu+1)=c(\gamma,\mu)(\gamma\|V^-\|\gamma)\,, \quad (87)$$

$$(\gamma,\mu|V^-|\gamma-1,\mu+1)$$
$$=((\gamma-1-\mu)(\gamma-\mu))^{1/2}(\gamma\|V^-\|\gamma-1)\,, \quad (88)$$

$$(\gamma,\mu|V^-|\gamma+1,\mu+1)$$
$$=((\gamma+1+\mu)(\gamma+\mu+2))^{1/2}(\gamma\|V^-\|\gamma+1)\,. \quad (89)$$

The above relations define the reduced matrix elements and up to now one has no relation between $(\gamma\|V^-\|\gamma-1)$ and $(\gamma-1\|V^-\|\gamma)$. Using Eq. (67), one obtains

$$(\gamma,\mu|V^3|\gamma,\mu)=\mu(\gamma\|V^-\|\gamma)\,, \quad (90)$$

$$(\gamma,\mu|V^3|\gamma-1,\mu)$$
$$=(-(\gamma-\mu)(\gamma+\mu))^{1/2}(\gamma\|V^-\|\gamma-1)\,, \quad (91)$$

$$(\gamma,\mu|V^3|\gamma+1,\mu)$$
$$=(-(\gamma+1-\mu)(\gamma+1+\mu))^{1/2}(\gamma\|V^-\|\gamma+1)\,. \quad (92)$$

Now V^3 is supposed Hermitic. From Eqs. (91) and (92), it follows that

$$(\gamma\|V^-\|\gamma-1)=(\gamma-1\|V^-\|\gamma)^*\,. \quad (93)$$

By evaluating the matrix elements of the commutator between K_a^- and K_a^3, one obtains

$$(\gamma\|K_a^-\|\gamma)$$
$$=|(\gamma\|K_a^-\|\gamma)|^2+(-2\gamma+1)|(\gamma\|K_a^-\|\gamma-1)|^2$$
$$+(2\gamma+3)|(\gamma\|K_a^-\|\gamma+1)|^2 \quad (94)$$

or, using Eqs. (86) and (90),

$$(-2\gamma+1)|(\gamma\|K_a^-\|\gamma-1)|^2$$
$$+(2\gamma+3)|(\gamma\|K_a^-\|\gamma+1)|^2$$
$$=[(\gamma(\gamma+1))^2-p]/(2\gamma(\gamma+1))^2\,, \quad (95)$$

where

$$p\equiv(\gamma_a(\gamma_a+1)-\gamma_b(\gamma_b+1))^2\,. \quad (96)$$

Taking the matrix element of the relation

$$K_a\cdot K_a=K_a^3K_a^3-(K_a^+K_a^-+K_a^-K_a^+)/2\,, \quad (97)$$

one obtains

$$\gamma(1-2\gamma)|(\gamma\|K_a^-\|\gamma-1)|^2$$
$$-(\gamma+1)(2\gamma+3)|(\gamma\|K_a^-\|\gamma+1)|^2$$
$$=[2q\gamma(\gamma+1)-p-(\gamma(\gamma+1))^2]/4\gamma(\gamma+1)\,, \quad (98)$$

where

$$q\equiv\gamma_a(\gamma_a+1)+\gamma_b(\gamma_b+1)\,. \quad (99)$$

Finally, from Eqs. (95) and (98), one obtains

$$(\gamma\|K_a^-\|\gamma-1)$$
$$=\pm((\gamma^2-(\gamma_a-\gamma_b)^2)$$
$$\times(\gamma^2-(\gamma_a+\gamma_b+1)^2))^{1/2}/(2\gamma(4\gamma^2-1)^{1/2})\,. \quad (100)$$

The phase convention chosen in the Appendix requires that the sign + should be retained on the right-hand side of the above equation, as γ is negative. To summarize, the recursion relation is

$$\gamma A(\gamma+1)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma+1,\mu)$$
$$+B(\gamma)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu)$$
$$+(\gamma+1)A(\gamma)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma-1,\mu)=0\,, \quad (101)$$

where

$$A(\gamma)\equiv[((\gamma)^2-(\gamma_a-\gamma_b)^2)((\gamma_a+\gamma_b)+1)^2-(\gamma)^2)$$
$$\times((\gamma)^2-(\mu)^2)/(4(\gamma)^2-1)]^{1/2} \quad (102)$$

and

$$B(\gamma)\equiv[(\gamma_a(\gamma_a+1)-\gamma_b(\gamma_b+1))\mu$$
$$+\gamma(\gamma+1)(\mu_b-\mu_a)]\,. \quad (103)$$

Within the usual phase conventions for SU(2) (see, e.g., Ref. 26), the SU(2) recursion relation is obtained from the one of SU(1,1) by the replacement of γ, μ by j, m.

IV. VECTOR MODEL FOR SU(1,1)

A direct orthonormal system of axes labeled by X^1, X^2, X^3 with origin O is shown in Fig. 1. In full analogy with the SU(2) case, one associates with a given state $|\gamma,\mu\rangle$ a set of three-dimensional vectors in Euclidean space as follows (remember that only positive discrete UIR's are considered, and therefore the x^3 components are positive). Equations (7) and (10) define a hyperboloid with rotational symmetry axis X^3 if $\gamma(\gamma+1)$ is positive, i.e., if γ is less than -1. The minimum distance of this hyperboloid to the origin O is equal to $(\gamma(\gamma+1))^{1/2}$, to be denoted by Γ from now on. Thus Eqs. (7) and (10) determine the whole set of vectors of origin O, the extremities of which are on the hyperboloid characterized by Γ. Equation (11) selects among these vectors those for which the projections on the X^3 axis are equal to μ. The latter vectors are assumed to be equiprobable as no more eigenvalues characterize the state $|\gamma,\mu\rangle$. The cases $\gamma=-1$ and $-\frac{1}{2}$ correspond to different geometrical pictures. For $\gamma=-1$, the hyperboloid becomes a cone. For $\gamma=-\frac{1}{2}$, the figure is generated by rotating around the X^3 axis the hyperbola $(x^3)^2-(x^1)^2=-0.25$. The results obtained from now on can be extended for these particular geometries.

A. Classical domains and algorithms for RF's

In the Lie algebra of the group SO(2,1), K^2 has the expression

$$K^2=i\left(x^3\frac{\partial}{\partial x^1}+x^1\frac{\partial}{\partial x^3}\right), \tag{104}$$

or, by using the parametric equation for the hyperboloid,

$$x^1=((\Gamma)^2+(x^2)^2)^{1/2}\sinh(t), \tag{105}$$

$$x^3=((\Gamma)^2+(x^2)^2)^{1/2}\cosh(t), \tag{106}$$

$$K^2=i\frac{d}{dt}. \tag{107}$$

A point on the hyperboloid Γ is thus determined by x^2 and t. Thus, for these points, the pseudo rotation operator has the following action in the configuration space:

$$(\Gamma,x^2,t\,|\exp(-i\beta K^2)|\Psi)=\Psi(\Gamma,x^2,t+\beta). \tag{108}$$

Otherwise stated, the operator $\exp(-i\beta K^2)$ acts on the state $|\Psi\rangle$ by transforming the points characterized by t into

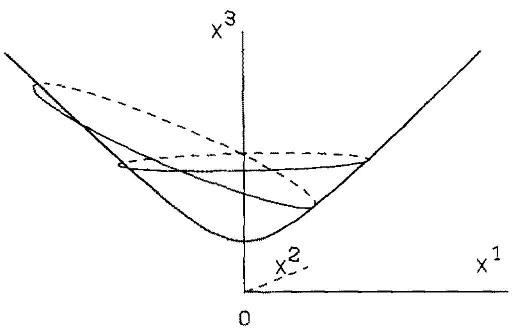


FIG. 1. Hyperboloid associated with a positive discrete UIR. The vectors with origin O and extremities on the horizontal circle characterized by $X^3=\mu$ are associated to the state $|\gamma,\mu\rangle$. After the action of the operator $\exp(-i\beta K^2)$, these vectors remain on the hyperboloid, but the plane developed by their extremities is no longer horizontal. (See text.)

the points characterized by $t-\beta$. Starting from an initial configuration associated with the state $|\gamma,\mu\rangle$, i.e., from the vectors whose extremities have for coordinates

$$x^1=((\mu+\Gamma)(\mu-\Gamma))^{1/2}\cos(\varphi)\equiv\Delta\cos(\varphi), \tag{109}$$

$$x^2=((\mu+\Gamma)(\mu-\Gamma))^{1/2}\sin(\varphi)\equiv\Delta\sin(\varphi), \tag{110}$$

$$x^3=\mu, \tag{111}$$

with a uniform probability density in φ between 0 and 2π, one obtains as a final configuration a set of vectors whose extremities have, for coordinates relative to the third axis,

$$\begin{aligned}x'^3&=((\Gamma)^2+(x^2)^2)^{1/2}\cosh(t-\beta)\\&=x^3\cosh(\beta)-x^1\sinh(\beta).\end{aligned} \tag{112}$$

The plane determined by the horizontal circle in Fig. 1 is thus rotated along the X^2 axis. The rotation angle is $\arctan(\tanh(\beta))$. It is clear from Eq. (112) that x'^3 is extremum when x^1 is extremum ($\varphi=0,\pi$). The classical domain for μ' is therefore determined, for $\beta>0$, by

$$\begin{aligned}&\mu\cosh(\beta)-((\mu)^2-\gamma(\gamma+1))^{1/2}\sinh(\beta)\\&\quad<\mu'<\mu\cosh(\beta)+((\mu)^2-\gamma(\gamma+1))^{1/2}\sinh(\beta).\end{aligned} \tag{113}$$

For SU(2), the same inequalities hold with the replacements $\gamma\to j$, $\mu\to m$, cosh→cos, sinh→sin, and the change of sign in the square root arguments. From Eq. (112) it is seen that μ' is a uniform function of φ in the interval $(0,\pi)$. The classical probability density $f(\mu')$ is then obtained according to the relation

$$f(\mu')d\mu'=d\varphi/\pi. \tag{114}$$

The result is

$$\begin{aligned}f(\mu')=1/[\pi(&-(\mu')^2-(\mu)^2-\gamma(\gamma+1)\sinh^2(\beta)\\&+2\mu\mu'\cosh(\beta))^{1/2}].\end{aligned} \tag{115}$$

Numerical stability for the three term recursion relations can only be *a priori* expected if the recursive evaluation proceeds from the classically forbidden region towards the classically allowed region.[18] The full domain of variation for μ' is between $-\gamma$ and infinity. For $\mu'=-\gamma$, Eq. (26) becomes a two term recursion relation. Therefore one starting value must be given, and is obtained from Eq. (20):

$$\begin{aligned}\delta^{\gamma}_{-\gamma,\mu}(\beta)&=((\mu-\gamma-1)!/((\gamma+\mu)!(-2\gamma-1)!))^{1/2}\\&\quad\times\sinh^{\gamma+\mu}(\beta/2)\cosh^{\gamma-\mu}(\beta/2).\end{aligned} \tag{116}$$

The recursion relation (26) is then used with increasing μ' values, provided μ' does not exceed the maximum classical value [Eq. (113)]. For μ' greater than this value, the recursion relation (26) must be used in the opposite direction, i.e., for decreasing μ' values. The problem now arises of how to initiate the recursion. The solution we propose is to use the relation (23) between SU(1,1) and SU(2) RF's. Indeed the full variation range of m' is bound from $-j$ to j. The recursion relation (27) becomes a two term recursion relation for these extermal values of m'. A recursion towards the classical SU(2) domain can therefore be made starting from the initial values

$$d^j_{-j,m}(\beta) = ((2j)!/((j+m)!(j-m)!))^{1/2} \times \cos^{j-m}(\beta/2)\sin^{j+m}(\beta/2), \quad (117)$$

$$d^j_{j,m}(\beta) = (-1)^{j-m}((2j)!/((j+m)!(j-m)!))^{1/2} \times \cos^{j+m}(\beta/2)\sin^{j-m}(\beta/2). \quad (118)$$

In fact these initial values are not even necessary. One can start the recursion (27) with arbitrary nonzero values and one matches the results inside the classical domain. The unitarity condition then provides the normalization. Thus only a phase must be specified. It is seen from Eq. (117) that the phase for $m' = -j$ and β between 0 and π is 1. Using the relation (23) for two successive values of μ', the recursion relation (26) can therefore be initiated for decreasing values of μ'. The determination of classical domains is particularly important for noncompact groups having infinite-dimensional UIR's. It provides a criterion for truncating infinite sums in practical calculations, as can be seen the following example:

$$\exp(-i\beta K^2)|\gamma,\mu) = \sum_{\mu'=-\gamma}^{\infty} \delta^{\gamma}_{\mu',\mu}(\beta)|\gamma,\mu'). \quad (119)$$

The square modulus of δ is reported in Fig. 2 for the case $\gamma = -2$, $\mu = 50$, and $\beta = 0.5$ radian. It is seen that outside the classical domain, the decrease of the square modulus of δ is very rapid. A numerical test is provided by the unitarity condition

$$\sum_{\mu'=-\gamma}^{\infty} |\delta^{\gamma}_{\mu',\mu}(\beta)|^2 = 1. \quad (120)$$

When the summation over μ' is restricted to the classical domain, the result is 0.96. When the summation is limited between $\mu' = 20$ and $\mu' = 100$, the result is already 0.999 999 9997.

Alternatively, a classical domain for γ can be determined from Eq. (113):

$$\gamma(\gamma+1) < (2\mu\mu' \cosh(\beta) - (\mu)^2 - (\mu')^2)/(\sinh^2(\beta)). \quad (121)$$

The recursion relation (63) therefore must be used in the direction of increasing values of γ. For γ minimum this recursion relation becomes a two term recursion relation. It can be initiated as before using the relation (23) between SU(2) and SU(1,1) RF's. It can also be initiated from the explicit expression [see Eq. (20)]

$$\delta^{-\mu'}_{\mu',\mu}(\beta) = ((\mu+\mu'-1)!/((\mu-\mu')!(2\mu'-1)!))^{1/2} \times \sinh^{\mu-\mu'}(\beta/2)\cosh^{-\mu-\mu'}(\beta/2), \quad (122)$$

$$\delta^{-\mu}_{\mu',\mu}(\beta) = (-1)^{\mu'-\mu}((\mu+\mu'-1)!/((\mu'-\mu)!(2\mu-1)!))^{1/2} \times \sinh^{\mu'-\mu}(\beta/2)\cosh^{-\mu-\mu'}(\beta/2). \quad (123)$$

The rapid decrease of RF's when entering the classically forbidden domain is illustrated in Fig. 3 for a particular case.

B. Classical domains and algorithms for CG coefficients

One first considers $\gamma_a,\mu_a,\gamma_b,\mu_b,\mu$ all fixed and one seeks the classical domain for γ. To the relation $K = K_a + K_b$ corresponds the diagram of Fig. 4, where O_aO_b represents a vector associated to $|\gamma_a,\mu_a)$. The azimuthal angle of O_aO_b has been taken arbitrarily equal to zero because of rotational invariance along the third axis. The third coordinate of O_b is equal to μ_a. To the state $|\gamma_b,\mu_b)$ are associated the classical vectors O_bM, where M is on the inner hyperboloid, and the component of O_bM along the third axis is equal to μ_b. The set of the points M describe the horizontal circle in Fig. 4, corresponding to a 2π variation of the azimuthal angle φ' of O_bM. The classical domain for γ is determined by the condition that M belongs to an hyperboloid with rotational symmetry axis OX^3. This condition yields

$$\begin{aligned}\gamma(\gamma+1) \equiv (\Gamma)^2 &= (\mu_a+\mu_b)^2 - (\Delta_a + \Delta_b\cos(\varphi'))^2 - (\Delta_b)^2\sin^2(\varphi') \\ &= (\Gamma_a)^2 + (\Gamma_b)^2 + 2(\mu_a\mu_b - \Delta_a\Delta_b\cos(\varphi')). \quad (124)\end{aligned}$$

From the above equation the classical domain for Γ is determined by the following inequalities:

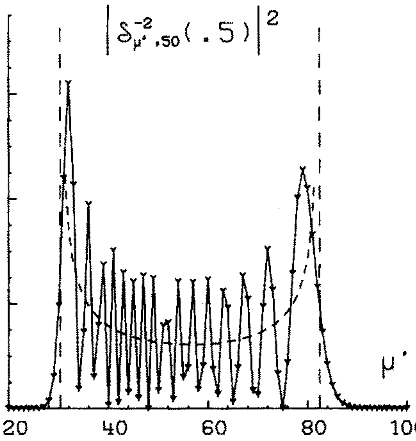


FIG. 2. Square modulus of the δ RF's as a function of μ' for $\gamma = -2$, $\mu = 50$, and $\beta = 0.5$. (See text.) The values are joined by straight lines. The vertical dashed lines correspond to the limits of the classical domain. The dashed line curve inside the classical domain represents the classical probability density.

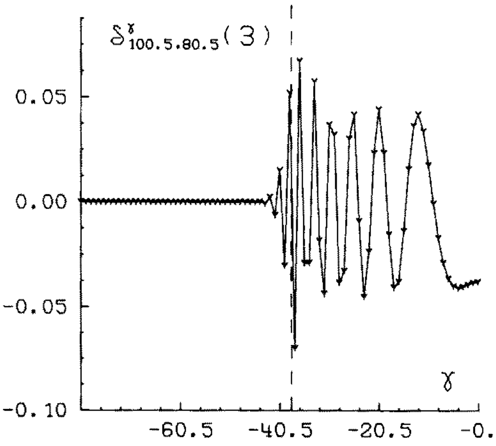


FIG. 3. The δ RF's as a function of γ for $\mu' = 100.5$, $\mu = 80.5$, and $\beta = 3$. The classical domain is at the right of the vertical dashed line.

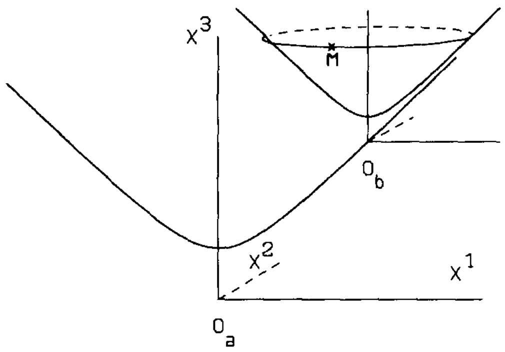


FIG. 4. Diagram corresponding to coupling of two positive discrete UIR's. (See text.)

$$(\Gamma_a)^2 + (\Gamma_b)^2 + 2(\mu_a\mu_b - \Delta_a\Delta_b) < (\Gamma)^2 < (\Gamma_a)^2 + (\Gamma_b)^2 + 2(\mu_a\mu_b + \Delta_a\Delta_b)\ . \quad (125)$$

From Eq. (124) it is seen that γ is a uniform function of φ' in the interval $(0,\pi)$. The classical probability density $f(\gamma)$ is then determined as previously from the uniform probability density of φ':

$$f(\gamma) = -(\gamma+\tfrac{1}{2})/[\pi((\Gamma_a\Gamma_b)^2 - (\mu_b\Gamma_a)^2 - (\mu_a\Gamma_b)^2 - E^2 + 2\mu_a\mu_b E)^{1/2}]\ , \quad (126)$$
$$E \equiv ((\Gamma)^2 - (\Gamma_a)^2 - (\Gamma_b)^2)/2\ .$$

Alternatively, Eq. (125) can be used to determine the classical domain for μ_a for fixed values of γ, γ_a, γ_b, and μ:

$$X - Y < \mu_a < X + Y\ , \quad (127)$$

with

$$X \equiv (\mu((\Gamma)^2 + (\Gamma_a)^2 - (\Gamma_b)^2)/2)/(\Gamma)^2\ , \quad (128)$$
$$Y \equiv \Delta[(((\Gamma)^2 + (\Gamma_a)^2 - (\Gamma_b)^2)/2)^2 - (\Gamma\Gamma_a)^2]^{1/2}/(\Gamma)^2\ . \quad (129)$$

Since the square modulus of a CG coefficient may be interpreted either as the probability to find the value γ for all other independent arguments fixed or as the probability to find the value μ_a for all other independent arguments fixed, the classical probability density $f(\mu_a)$ is also given by the right-hand side of Eq. (126).

The algorithms for the computation of SU(1,1) coefficients within a given UIR are now easily derived from the recursion relation (34) and the corresponding classical domain defined by the inequalities (127). For a recursion between different UIR's, the algorithms are derived from the recursion relation (101) and the corresponding classical domain (125). In both cases the full domain of variation is bound and the recursion relations become two term recursion relations at its limits. The unitarity conditions [Eqs. (30) and (31)] provide the normalization. The method has been exposed in detail for the SU(2) case in Ref. 18 and will not be repeated here. We only give the necessary phases according to the convention used in the present work:

$$\text{phase}((\gamma_a, -\gamma_a;\gamma_b,\mu_b|\gamma,\mu)) = 1\ , \quad (130)$$

$$\text{phase}((\gamma_a,\mu_a;\gamma_b,\mu_b|-\mu,\mu)) = (-1)^{\gamma_a+\mu_a}\ . \quad (131)$$

It can be seen that the above algorithms exactly correspond, in fact, to the SU(2) algorithms.[18] Indeed a comparison between the explicit expression given in the Appendix [Eq. (A11)] and the one given in Ref. 26 gives[10]

$$\begin{aligned}
&(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu)(-1)^{\gamma_a+\mu_a} = \langle j_a,m_a;j_b,m_b|j,m\rangle\ ,\\
&j_a = (\mu_a+\mu_b-\gamma_b+\gamma_a-1)/2\ ,\\
&j_b = (\mu_a+\mu_b-\gamma_a+\gamma_b-1)/2\ ,\\
&m_a = (\mu_b-\mu_a-\gamma_b-\gamma_a-1)/2\ ,\\
&m_b = (\mu_a-\mu_b-\gamma_b-\gamma_a-1)/2\ ,\\
&j = -\gamma-1\ .
\end{aligned} \quad (132)$$

From the above equations one observes that the variation of j for all other SU(2) arguments fixed also corresponds to the variation of γ for all other SU(1,1) arguments fixed. Similarly, the variation of m_a for m, j, j_a, and j_b fixed corresponds to the variation of μ_a for μ, γ, γ_a, and γ_b fixed. Therefore the computation of a whole set of SU(2) CG coefficients for successive values of j or m_a by means of the SU(2) algorithms described in Ref. 18 directly provides the values of a whole set of SU(1,1) CG coefficients for successive values of γ or μ_a according to the correspondence given by the Eqs. (132). It can also be verified that, within this correspondence, the classical SU(1,1) probability density given by Eq. (126) is the classical SU(2) probability density.[14,15,19] The behavior of CG coefficients inside and outside the classical domain is illustrated in Refs. 18 and 19, together with a comparison with the classical probability density.

V. CONCLUDING REMARKS

Symmetry properties of generalized CG coefficients or of RF's, or equivalently relations between SU(1,1) and SU(2) CG coefficients or RF's, allow one to interpret some special properties by group theory. For example, the orthogonality property of δ (see Ref. 21) can be expressed by using Eq. (23) as

$$\int_0^\pi d\beta \tan\left(\frac{\beta}{2}\right) d^j_{m,m+k}(\beta) d^j_{m',m'+k}(\beta) = \frac{\delta_{mm'}}{2m+k}\ , \quad (133)$$

for $2m+k$ and $2m'+k$ positive.

The latter relation seems difficult to explain only from the SU(2) group theoretical point of view.

Except in Sec. II, all the results have been obtained starting from the defining commutation relations for the Hermitic generators, for both SU(2) and SU(1,1). A more economical way would be to derive directly the SU(1,1) results by analytic continuation of the SU(2) results. We believe that the direct approach that has been used is also instructive.

Finally we emphasize the importance of the existence of a classical domain in the infinite-dimensional case. This allows one to see how an infinite expansion can be truncated in practical calculations. The δ RF occurs in the radial matrix elements of the Coulomb problem, and in the matrix elements between different Sturmian bases.[1,2] An approximate residual O(4) symmetry has been found for diexcited two electron atoms.[27,28] It would be of great interest to know if this could be generalized to include states of different energies within the framework of an o(4,2) noninvariance algebra.[1,2,29]

Note added in proof: We have recently learned that the relation between SU(1,1) RF's and SU(2) RF's [Eq. (23) of the present paper] was previously discovered by A. M. Perelomov, Rep. Math. Phys. **2**, 277 (1971), within different phase conventions.

ACKNOWLEDGMENTS

The author would like to thank J. Pascale, M. Poirier, F. Gounand, and B. Carré for helpful discussions and for a careful reading of the manuscript.

APPENDIX: PHASE CONVENTIONS FOR SU(1,1) CG COEFFICIENTS

The procedure briefly described now closely follows the one given in Ref. 26 for SU(2) CG coefficients. The first convention is given by

$$|(\gamma_a\gamma_b)\gamma_a+\gamma_b,-\gamma_a-\gamma_b)=|\gamma_a,-\gamma_a;\gamma_b,-\gamma_b)\,. \tag{A1}$$

[The symbols $(\gamma_a\gamma_b)$ will be dropped in the following for the sake of simplicity.] Equation (70) gives

$$(\gamma,\mu+1|K^+|\gamma,\mu)(\gamma,\mu|K_a^+|\gamma',\mu-1) - (\gamma,\mu+1|K_a^+|\gamma',\mu)(\gamma',\mu|K^+|\gamma',\mu-1)=0\,. \tag{A2}$$

Since the matrix elements of K^+ in the above equation are positive [see Eq. (15)], it follows that the matrix elements of K_a^+ between states of given γ and γ' have the same phase. Now, Eq. (69) gives

$$(\gamma,\mu+1|K_a^3|\gamma+1,\mu+1) = (c(\gamma,\mu)(\gamma,\mu|K_a^3|\gamma+1,\mu) + (\gamma,\mu+1|K_a^+|\gamma+1,\mu))/c(\gamma+1,\mu)\,, \tag{A3}$$

$$(\gamma,-\gamma|K_a^3|\gamma+1,-\gamma) = (\gamma,-\gamma|K_a^+|\gamma+1,-\gamma-1)/c(\gamma+1,-\gamma-1)\,. \tag{A4}$$

The two above equations show that the nondiagonal matrix elements of K_a^3 have the same phase as those of K_a^+.

The following phase convention can therefore be chosen: all the matrix elements of K_a^3 nondiagonal in γ are real and nonpositive.

The consequences of this convention on the phase of certain CG coefficients are now considered. Operating with K^+ or K^- on Eq. (28) gives

$$c(\gamma,\mu\pm1)(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu\pm1) = c(\gamma_a,\pm(\mu_a\mp1))(\gamma_a,\mu_a\mp1;\gamma_b,\mu_b|\gamma,\mu) + c(\gamma_b,\pm(\mu_b\mp1))(\gamma_a,\mu_a;\gamma_b,\mu_b\mp1|\gamma,\mu)\,. \tag{A5}$$

In particular, the action of K^- in the case $\mu=-\gamma$ shows that the signs of these CG coefficients alternate with μ_a:

$$\text{phase}((\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,-\gamma)) = (-1)^{\gamma_a+\mu_a}\text{phase}((\gamma_a,-\gamma_a;\gamma_b,\mu_b'|\gamma,-\gamma))\,. \tag{A6}$$

Equation (69) gives

$$\begin{aligned}0&>(\gamma+1,-\gamma-1|K^-|\gamma+1,-\gamma)\times(\gamma+1,-\gamma|K_a^3|\gamma,-\gamma)\\&=(\gamma+1,-\gamma-1|K_a^-|\gamma,-\gamma)\\&=\sum_{\mu_a,\mu_b}(\gamma+1,-\gamma-1|\gamma_a,\mu_a;\gamma_b,\mu_b)\times(\gamma_a,\mu_a|K_a^-|\gamma_a,\mu_a+1)(\gamma_a,\mu_a+1;\gamma_b,\mu_b|\gamma,-\gamma)\,.\end{aligned} \tag{A7}$$

Using Eq. (A6), one obtains

$$\text{phase}((\gamma+1,-\gamma-1|\gamma_a,-\gamma_a;\gamma_b,\mu_b))\times\text{phase}((\gamma_a,-\gamma_a;\gamma_b,\mu_b'|\gamma,-\gamma))>0\,. \tag{A8}$$

According to the first convention [Eq. (A1)], it follows that

$$\text{phase}((\gamma_a,-\gamma_a;\gamma_b,\mu_b|\gamma,-\gamma))=1\,, \tag{A9}$$

for any allowed value of γ. Finally Eq. (A5) yields, in the case $\mu_a=-\gamma_a$,

$$c(\gamma,\mu)(\gamma_a,-\gamma_a;\gamma_b,\mu_b|\gamma,\mu+1) = c(\gamma_b,\mu_b-1)(\gamma_a,-\gamma_a;\gamma_b,\mu_b-1|\gamma,\mu)\,. \tag{A10}$$

Therefore

$$\text{phase}((\gamma_a,-\gamma_a;\gamma_b,\mu_b|\gamma,\mu))=1\,.$$

The phase conventions of the present paper agree with the explicit expression given by Sannikov[30]:

$$\begin{aligned}(\gamma_a,\mu_a;\gamma_b,\mu_b|\gamma,\mu) &= [(-2\gamma-1)(-\gamma-\gamma_a-\gamma_b-2)!(-\gamma-\gamma_a+\gamma_b-1)!(-\gamma+\gamma_a-\gamma_b-1)!\\&\quad\times(-\gamma+\gamma_a+\gamma_b)!(\mu+\gamma)!(\mu_a+\gamma_a)!(\mu_a-\gamma_a-1)!(\mu_b+\gamma_b)!(\mu_b-\gamma_b-1)!/(\mu-\gamma-1)!]^{1/2}\\&\quad\times\sum_z\frac{(-1)^z}{[z!(\gamma_a+\gamma_b-\gamma-z)!(\gamma_a+\mu_a-z)!(-\gamma+\gamma_a-\gamma_b-1-z)!(-2\gamma_a-1+z)!(\gamma-\gamma_a+\mu_b+z)!]}\,.\end{aligned} \tag{A11}$$

[1]B. G. Adams, J. Cizek, and J. Paldus, *Dynamical Groups and Spectrum Generating Algebras*, edited by Y. Neeman, A. O. Barut, and A. Bohm (World Scientific, Singapore, 1986).

[2]M. J. Englefield, *Group Theory and the Coulomb Problem* (Wiley–Interscience, New York, 1972), and references therein.

[3]M. Bohm and G. Junker, J. Math. Phys. **28**, 1978 (1987).

[4]C. C. Gerry and J. Kiefer, Phys. Rev. A **37**, 665 (1988).

[5]J. Katriel, M. Rasetti, and A. I. Solomon, Phys. Rev. D **35**, 1248 (1987).

[6]G. Dattoli, F. Orsitto, and A. Torre, Phys. Rev. A **34**, 2466 (1986).

[7]S. K. Bose, J. Phys. A **18**, 903 (1985).

[8]T. Hori and K. Kamimura, Nucl. Phys. B **253**, 249 (1985).

[9]A. D'Adda, R. Dauria, and G. Ponzano, J. Math. Phys. **15**, 1543 (1974).

[10]W. Rasmussen, J. Phys. A **8**, 1038 (1975).
[11]E. Chacon, D. Levi, and M. Moshinsky, J. Math. Phys. **16**, 1876 (1975).
[12]W. J. Holman and L. C. Biedenharn, Ann. Phys. (NY) **39**, 1 (1966).
[13]W. J. Holman and L. C. Biedenharn, Ann. Phys. (NY) **47**, 205 (1968).
[14]E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[15]P. J. Brussaard and H. A. Tolhoek, Physica **23**, 955 (1957).
[16]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics: Theory and Application*, Vol. 8, *Encyclopedia of Mathematics and its Applications* (Addison–Wesley, Reading, MA, 1981).
[17]L. C. Biedenharn and J. D. Louck, *The Racah–Wigner Algebra in Quantum Theory*, Vol. 9, *Encyclopedia of Mathematics and its Application* (Addison–Wesley, Reading, MA, 1981).
[18]K. Schulten and R. G. Gordon, J. Math. Phys. **16**, 1961 (1975).
[19]K. Schulten and R. G. Gordon, J. Math. Phys. **16**, 1971 (1975).
[20]To avoid possible confusion between contravariant components and powers, parentheses will be included if the upper index is a power. For example, $(x)^2$ means xx, y^2 means the second contravariant component of y.
[21]V. Bargmann, Ann. Math. **48**, 568 (1947).
[22]A. O. Barut and E. C. Phillips, Commun. Math. Phys. **8**, 52 (1968).
[23]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), 5th ed.
[24]J. Schwinger, "On angular momentum" (1952), in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965), pp. 229–279.
[25]E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U.P., Cambridge, 1953).
[26]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1974).
[27]O. Sinanoglu and D. R. Herrick, J. Chem. Phys. **62**, 886 (1975).
[28]C. Wulfman, Chem. Phys. Lett. **23**, 370 (1973).
[29]A. O. Barut and H. Kleinert, Phys. Rev. A **156**, 1541 (1967).
[30]S. S. Sannikov, Sov. Phys Dokl. **11**, 1045 (1967).

Integral-spin fields on (3+ 2)-de Sitter space

Jean-Pierre Gazeau
Laboratoire de Physique Théorique et Mathématique, Université Paris 7, Equipe de Recherches Centre National de la Recherche Scientifique 177, 2, place Jussieu, 75251 Paris Cedex 05, France
Michel Hans
UAP, New Technology Division, 20 ter rue de Bezons, 92411 Courbevoie Cedex, France

(Received 26 October 1987; accepted for publication 8 June 1988)

Nowadays, (3 + 2)-de Sitter (or anti-de Sitter space) appears as a very attractive possibility at several levels of theoretical physics. The Wigner definition of an elementary system as associated to a unitary irreducible representation of the Poincaré group may be extended to the de Sitter group SO(3,2) [or $\overline{\mathrm{SO}(3,2)}$] without great difficulty. The constant curvature, as small as it can be, is a natural candidate to play the role of a regularization parameter with respect to the flat-space limit. Massless particles in (3 + 2)-de Sitter theory are composite (singletons). On the other hand, supergravity theories necessitate a (large) constant curvature. The content of this paper is group theoretical. It attempts to continue the "*à la* Wigner" program for SO(3,2), already largely broached by Fronsdal. Three recurrence formulas are presented. They permit one to build up the carrier states for representations with arbitrary integral spin. Two of them are valid for the "massive" representations whereas the third one is applicable to the indecomposable massless representations. In addition, other presumably indecomposable, though nonphysical, representations are studied, in relation to the existence of "generalized" gauge fields and divergences. The recurrence formulas also allow one to build up the invariant two-point functions or homogeneous propagators. Hence it becomes possible to examine the problems of light-cone propagation and "reverberation" into the light cone and to make the following assertion: for a certain choice of the gauge-fixing parameters, the massless states with arbitrary spin propagate only on the light cone and whatever gauge one chooses their *physical* parts propagate on the light cone.

I. INTRODUCTION

As Galilean relativity appears to be the limit $c\to+\infty$ of Poincaré relativity, the latter can be considered as the idealistic "flat" limit $\rho\to 0$ of two possible curved-space-time relativities of maximal symmetry. Indeed, a four-dimensional (pseudo-) Riemanian space may admit a continuous group of isometry (i.e., preserving the metric $g_{\mu\nu}$) with up to ten essential parameters. The maximum number is realized only for a space of constant curvature ρ, the curvature tensor then reading

$$R_{\mu\nu\lambda\rho}=\rho(g_{\mu\lambda}g_{\nu\rho}-g_{\mu\rho}g_{\nu\lambda}).$$

Those space-times that go to the flat Minkowski space as the curvature ρ tends to zero are the ordinary de Sitter spaces, of which there are two. The first one admits SO(4,1) as a group of motions. It is essentially finite in extension[1]: given any point P and any timelike direction in that point, the geodesics through P perpendicular to the chosen timelike direction are finite. On the other hand, the second one, commonly called anti-de Sitter space or (3 + 2)-de Sitter space, is infinite in extension; analogous geodesics have infinite lengths and are completely spacelike. Its group of motion is SO(3,2). The time is proportional to the rotation parameter associated with the subgroup SO(2). If this periodicity is seen as a difficulty, it is easy to circumvent it by dealing with the covering space, of motion group $\overline{\mathrm{SO}(3,2)}$: then the time is not bounded. As the Minkowski flat space-time is the limit $\rho\to 0$ of the ordinary de Sitter space-times, the Poincaré group, namely Lorentz $\otimes T_4$, can be seen as a contraction of SO(4,1) or SO(3,2). However, compared to the Galilean contraction of the Poincaré group, nothing, from a strict operational point of view, can be asserted about the de Sitter–Minkowski relationship. The former expresses in mathematical terms the historical emergence of a new physical theory supplanting the old one, relegated to the rank of an approximate, although honorable, framework. This process has been maturing for a long time through the experimental practice of the physicists. It is firmly supported by more and more precise measurements of what has acquired the respectable status of the "universal physical constant c."[2] No such accomplishment presently exists for the curvature parameter ρ or, equivalently, the Einstein cosmological constant Λ (it is not difficult to show the relationship $|\Lambda| = 3|\rho|$). Upper bounds can only be given for the value of $|\Lambda|$, these estimates being based on its relationship[3] with the stress-energy tensor $T^{(\mathrm{vac})}_{\mu\nu}$ associated with the vacuum ("vacuum polarization"),

$$T^{(\mathrm{vac})}_{\mu\nu}=-(\Lambda/4\pi)g_{\mu\nu}.$$

One can place the limit $|\Lambda|\sim 10^{-58}\,\mathrm{cm}^{-2}$ on the cosmological constant. This number is negligibly small. But *no* sound argument allows us to state that the cosmological constant is zero.

On the other hand, a reasonably speculative attitude is to consider ρ as associated to a new degree of freedom, in the sense given by Fronsdal in 1965[4]: "A physical theory that treats space-time as Minskovskian flat must be obtainable as a well-defined limit of a more general theory, for which the

 0022-2488/88/122533-20$02.50

assumption of flatness is not essential."

Interest in constant curvature is thus not purely gratuitous: a legitimate curiosity presses mathematical physicists to continuously deform commonly accepted structures and analyze what ensues in all its physical implications. However, deforming in this manner the Poincaré group towards the $(4+1)$-de Sitter group causes some bothersome, even unacceptable, features to appear: finiteness in extension, already mentioned,[1] nonexistence of a lower bound for the energy spectrum,[5] and absence of microcausality in a quantum field theory.[6] By contrast, the opposite choice does not disturb too much the canons of physics. Besides those geometrical aspects of the $(3+2)$-de Sitter space favored by Wigner, a causal structure exists, discovered by Castell.[7] There also exists a set of unitary irreducible representations of SO(3,2) to which elementary particles can be associated: for each of them, the spectrum of the Hamiltonian has a minimum, and the subspace of states having this minimal energy as an eigenvalue carries a well-defined angular momentum, which allows a natural definition of the spin. A reformulation of the Wigner program for elementary systems[8–10] can therefore be launched within the $(3+2)$-de Sitter framework. A large part of this task, including quantization of free fields, has been accomplished by Fronsdal and his collaborators throughout a series of comprehensive papers,[4] where they have successively treated the spinless representations,[11] the free Dirac fields,[12] and the massless integral-spin or half-integer-spin fields.[13] In particular, the concept of masslessness on $(3+2)$-de Sitter space is now firmly established by controlling the validity of several criteria like conformal extension and Poincaré contraction[14] and gauge structure and light-cone propagation.[15,16] A feature of considerable interest is also the composite nature of the massless particles in $(3+2)$-de Sitter space. At the lower bound of unitarity two representations discovered by Dirac take place[17]: the spinless "Rac" and the spin-$\frac{1}{2}$ "Di." The nonobservability of these "singletons" holds for purely kinematical reasons. Furthermore, their remarkable role as constituents of massless particles was proved by Flato and Fronsdal[18] through the tensor reduction

$$\text{singleton} \otimes \text{singleton} = \oplus \text{ massless particles},$$

an equation that loses any sort of meaning in the flat-space limit.

The $(3+2)$-de Sitter space is endowed with other advantages. For instance, the curvature ρ implies a sort of universal confinement for free particles, well put in evidence by examining the "nonrelativistic" contraction of SO(3,2),[19,20] obtained through rescaling $\rho \to \rho c^{-2}$, and going to the limit $c \to +\infty$. An intermediate group is obtained, describing a world with finite curvature but infinitely fast signal propagation. The one-particle representative Hamiltonian is then given by

$$H = p^2/2m + (m/2)\rho q^2.$$

Here ρ appears as an external-harmonic-oscillator coupling constant, responsible for the discretization of the free-particle energy. The same procedure applied to SO(4,1) yields an opposite sign and explains the nonexistence of a lower bound for the energy spectrum. Regularization of the infrared region of $(3+2)$-de Sitter covariant field theories by the introduction of a small but nonzero constant curvature hence becomes a very interesting opportunity.

It was tempting to exploit this idea of geometrical confinement at a totally different order of magnitude. Salam and Strathdee[21] proposed the closed SO(3,2) symmetric de Sitter universe as the best candidate for strong curvature. For instance, a model of hadrons, in which quarks, antiquarks, and gluons move inside a "finite spherical" and strongly curved anti-de Sitter universe, is discussed by van Beveren, Dullemond, and Rijken in a series of papers quoted in Ref. 22.

Motivation for $(3+2)$-de Sitter space is also provided by its appearance in the maximally supersymmetric classical solutions of supergravity theories.[23,24]

Some interpretative difficulties subsist, however. What is the effect of the information (Cauchy data) entering and leaving the space-time through its timelike spatial infinity, even though a consistent quantization scheme has been devised by Avis, Isham, and Storey[25]? The loss of the concept of helicity for massless fields on $(3+2)$-de Sitter space, in spite of the existence of two independent Gupta–Bleuler triplets, merits more research by carefully examining the flat-space limit.[15] Besides, this problem is intimately connected to the previous one and to the question of separate domains of self-adjointness for the corresponding Hamiltonians.

This paper attempts to tell more about the $(3+2)$-de Sitter theory. Its content is mainly group-theoretical and answers several technical questions like explicit construction of states and related homogeneous propagators for arbitrary integral-spin representations. In particular, a demonstration is given for the light-cone propagation of the massless fields, completing the Appendix of Ref. 16. The methods are mostly inductive. Their field of application is large enough to include some of the nonunitary and/or indecomposable representations. One thus achieves a better understanding of the place occupied by the "physical sector" within a large set of de Sitter representations.

The organization of this paper is as follows. In Sec. II, we briefly review the definitions and properties of the $(3+2)$-de Sitter space, its motion group, and the representations of the latter that are physically interesting. Two recurrence formulas are given in Sec. III. They permit one to give the general solutions of wave equations and hence the associated representation spaces. The massless case is, however, excluded.

The possible occurrence of invariant subspaces of solutions naturally leads to the Weyl equivalence between group representations, a concept explained in Sec. IV. These solutions have the general form

$$\tilde{k} = \mathscr{P}_{s',s}\zeta$$

and are called "generalized gauge fields." Here, ζ is a tensor of rank s whereas k is of rank $s' > s$; $\mathscr{P}_{s',s}$ is a differential operator of order $s'-s$. Its definition, some of its remarkable properties, and its explicit expression are listed in Sec. V.

In Sec. VI, the Weyl equivalence, and the operator $\mathscr{P}_{s',s}$ and its "dual" $\mathscr{P}^*_{s',s}$ are shown to be the cornerstones for the construction of certain possible indecomposable representations of the $(3+2)$-de Sitter group. Also $\mathscr{P}^*_{s,s'}$ is a differential operator of order $s'-s$ and generalizes the divergence. Together with $\mathscr{P}_{s',s}$, it realizes a factorization of central elements of the enveloping algebra:

$$\mathscr{P}^*_{s,s'}\mathscr{P}_{s',s}\zeta \propto \prod_{i=1}^{s'-s}(Q-\lambda_i)\zeta,$$

where Q is the second-order Casimir operator. The lowest case, i.e., $s'=s+1$, is precisely the massless case. Herein, $\mathscr{P}_{s',s}$ and $\mathscr{P}^*_{s,s'}$ are, respectively, reduced to the analog of the gradient and the divergence in the five-dimensional de Sitter formalism. This case was partially treated in Ref. 26, where it was demonstrated that the Gupta–Bleuler minimal structure is reached for the spin-s-dependent gauge fixing $c_s = 2/(2s+1)$. This study is continued in Sec. VII, where the third recurrence formula, necessary to the construction of the massless states, is restated and completed. Very convincing indications in favor of true indecomposibility exist. They are based on simple checking in the lower-spin cases. However, no rigorously complete proof can presently be given. This explains the presence in Secs. VI and VII of a certain number of conjectural assertions.

Section VIII finally deals with the propagation problem through the construction of invariant two-point functions or homogeneous propagators. Besides the purely light-cone propagation of the massless fields for a particular gauge, i.e., when the gauge-fixing parameter is adjusted to the "good" value $c_s = 2/(2s+1)$, a striking feature for the *massive* fields of integral energy emerges: the homogeneous propagator of the corresponding unitary irreducible representation $\mathscr{U}$ admits a decomposition in two terms. The first one is responsible for a propagation confined to the light cone whereas the other one, describing the propagation *into* the light cone, has as an analytic factor the polynomial propagator for the unique finite irreducible representation Weyl-equivalent to $\mathscr{U}$. This remarkable duality certainly deserves further investigation.

II. (3+2)-DE SITTER SPACE, ALGEBRA, GROUP, AND REPRESENTATIONS

The $(3+2)$-de Sitter space[11] is most easily described as embedded in $\mathbb{R}^5$ provided with the metric: $\delta_{\alpha\beta} = \mathrm{diag}(1,-1,-1,-1,1)$. Here, the Greek letters take the values 0,1,2,3,5. The missing number 4 is left apart for a possible extension to conformal theories. Points in $\mathbb{R}^5$ are thus denoted by

$$y=(y_0,y_1,y_2,y_3,y_5)\equiv(y_0,\mathbf{y},y_5)\equiv(y_\alpha),$$

and the de Sitter space can be visualized as (the covering space of) the connected hyperboloid

$$y^2\equiv\delta_{\alpha\beta}y^\alpha y^\beta = y_0^2-\mathbf{y}^2+y_5^2 = 1/\rho,$$

ρ being the (positive) curvature. Integral spin fields in de Sitter space are represented by using symmetric tensor fields on the hyperboloid:

$$y\to k(y)\equiv(k_{\alpha_1\cdots\alpha_s}(y)),$$

where s is the rank of k.

Some definitions become necessary here.

(a) *Symmetrizer:* The symmetrizer of the tensor product of two symmetric tensors ζ and η of rank p and $s-p$, respectively, where $p\leqslant[s/2]$, is denoted by Σ_p. The components of the symmetrized tensor product are given by

$$(\Sigma_p\zeta\eta)_{\alpha_1\cdots\alpha_s} = \sum_{i_1<i_2<\cdots<i_p}\zeta_{\alpha_{i_1}\alpha_{i_1}\cdots\alpha_{i_p}}\eta_{\alpha_1\cdots\not\alpha_{i_1}\cdots\not\alpha_{i_2}\cdots\not\alpha_{i_p}\cdots\alpha_s}. \tag{2.1}$$

(b) *Transversality:* A symmetric tensor field k is said to be transverse if

$$y^{\alpha_s}k_{\alpha_1\cdots\alpha_s} = (y\cdot k)_{\alpha_1\cdots\alpha_{s-1}} = 0,$$

or more concisely $y\cdot k=0$. The transverse projector for tensors of rank 1 is defined by the symmetric rank-2 tensor $\Theta=(\Theta_{\alpha\beta})$:

$$\Theta_{\alpha\beta}=\delta_{\alpha\beta}-\rho y_\alpha y_\beta. \tag{2.2}$$

(Important note: hereafter, y^2 is denoted by $1/\rho$, whether it is a constant or not.)

Precisely, the transverse projection Tk of the symmetric tensor of rank s, $k=(k_{\alpha_1\cdots\alpha_s})$, has the components

$$(Tk)_{\alpha_1\cdots\alpha_s} = \left(\prod_{i=1}^{s}\Theta_{\alpha_i}{}^{\beta_i}\right)k_{\beta_1\cdots\beta_s}. \tag{2.3}$$

(c) *Tracelessness:* A symmetric tensor field k is traceless if

$$\delta^{\alpha_{s-1}\alpha_s}k_{\alpha_1\cdots\alpha_{s-2}\alpha_{s-1}\alpha_s}=0.$$

The trace of an arbitrary tensor k is denoted by k'. It is a symmetric tensor of rank $s-2$, the components of which are given by

$$(k')_{\alpha_1\cdots\alpha_{s-2}} = \delta^{\alpha_{s-1}\alpha_s}k_{\alpha_1\cdots\alpha_{s-2}\alpha_{s-1}\alpha_s}. \tag{2.4}$$

More generally, the nth trace, $n\leqslant[s/2]$, of an s-rank tensor is denoted by $k^{(n)}$:

$$k^{(n)}\equiv\overbrace{(\cdots(k')'\cdots)'}^{n\text{ times}}.$$

The group of isometries of (the universal covering of) de Sitter space is the pseudo-orthogonal group $[\,\overline{SO(3,2)}\,]$ SO(3,2). Let us describe its action on symmetric tensor fields of rank s. The infinitesimal generators in this vector space are denoted by $L^{(s)}_{\alpha\beta} = -L^{(s)}_{\beta\alpha}$. We stress, when it seems necessary, the importance of the rank s by making it explicit in the various symbols introduced in the text. Note that the rank s should generally be dissociated from the value s of the spin. The generator representatives $L^{(s)}_{\alpha\beta}$ are defined by

$$L^{(s)}_{\alpha\beta} = M_{\alpha\beta}+S^{(s)}_{\alpha\beta}, \tag{2.5}$$

where $M_{\alpha\beta}$ is the "orbital part"

$$M_{\alpha\beta}=i(y_\alpha\,\partial_\beta - y_\beta\,\partial_\alpha),\quad \partial_\alpha=\frac{\partial}{\partial y^\alpha}, \tag{2.6}$$

and $S^{(s)}_{\alpha\beta}$ is the "spinorial part"

$$S^{(s)}_{\alpha\beta}k_{\alpha_1\cdots\alpha_s} = i\sum_i(\delta_{\alpha\alpha_i}k_{\alpha_1\cdots\not\alpha_i^{\,\beta}\cdots\alpha_s} - \delta_{\beta\alpha_i}k_{\alpha_1\cdots\not\alpha_i^{\,\alpha}\cdots\alpha_s}). \tag{2.7}$$

The second-order Casimir operator representative is denoted by Q_s:

$$Q_s \equiv \tfrac{1}{2} L^{(s)}_{\alpha\beta} L^{(s)\alpha\beta}.$$

The following contraction formula enables one to express it in terms of $\bar{\partial}$:

$$\tfrac{1}{2} M^{\alpha\beta} M_{\alpha\beta} = Q_0 = -\rho^{-1} \bar{\partial}^\alpha \bar{\partial}_\alpha \equiv -\rho^{-1} \bar{\partial}^2. \quad (2.8)$$

Herein, $\bar{\partial}$ designates the transverse projection of the gradient ∂ and can be considered as the tangentional derivative:

$$\bar{\partial}_\alpha = \Theta_{\alpha\beta} \partial^\beta = \partial_\alpha - \rho y_\alpha y \cdot \partial, \quad (2.9)$$

$$M^{\alpha\beta} S^{(s)}_{\alpha\beta} k = 2\Sigma_1 \partial y \cdot k - 2\Sigma_1 y \, \partial \cdot k - 2sk, \quad (2.10)$$

$$\tfrac{1}{2} S^{(s)\alpha\beta} S^{(s)}_{\alpha\beta} k = s(s+3)k - 2\Sigma_2 \delta k'. \quad (2.11)$$

The expression for Q_s follows:

$$Q_s k = Q_0 k + 2\Sigma_1 \partial y \cdot k - 2\Sigma_1 y \, \partial \cdot k - 2\Sigma_2 \delta k' + s(s+1)k. \quad (2.12)$$

It is clear that the operators $\bar{\partial}$, $L^{(s)}_{\alpha\beta}$, and Q_s commute with y^2: they are intrinsically defined on the hyperboloid $y^2 = \text{const}$.

III. $D(E_0,s)$-CARRIER STATES: TWO RECURRENCE FORMULAS

The group SO(3,2) is a real form of the complex Lie algebra named B_2 in the Cartan classification.[27] The finite irreducible representations are labeled by two positives integers (k_1,k_2) (dominant weight) but we use (E_0,s) instead:

$$E_0 = k_1 - k_2/2, \quad s = k_2/2,$$

since they are more adapted to physical representations. In what follows, E_0 is allowed to take arbitrary real values whereas s takes integral or half-integral positive values. Here E_0 is the lowest among the eigenvalues E (energy) of L_{50} and the spin s is the angular momentum of the lowest energy space. The corresponding representations are denoted $D(E_0,s)$. If $D(E_0,s)$ is irreducible, the second-order Casimir operator takes the constant value

$$\langle Q_s^{E_0} \rangle = E_0(E_0 - 3) + s(s+1). \quad (3.1)$$

The representations $D(E_0,s)$ are irreducible and unitary if

$$s = 0 \quad \text{and} \quad E_0 > \tfrac{1}{2},$$
$$s = \tfrac{1}{2} \quad \text{and} \quad E_0 > 1, \quad (3.2)$$
$$s \geqslant 1 \quad \text{and} \quad E_0 > s + 1.$$

At the lower limit of unitarity, i.e., $D(\tfrac{1}{2},0)$, $D(1,\tfrac{1}{2})$, and $D(s+1,s)$ for $s \geqslant \tfrac{1}{2}$, invariant subspaces exist. Unitarity and irreducibility are restored by considering quotient spaces.

We are now in a position to say more about the representations $D(E_0,s)$ and their carrier states. The most familiar way to characterize a carrier space is to appeal to the solutions of some differential equations. Let us consider the following "wave equation":

$$(Q_s - \langle Q_s^{E_0} \rangle) k = 0, \quad (3.3)$$

supplemented with the auxiliary conditions: homogeneity,

$$(\hat{N} - N) k = 0, \quad \hat{N} = y \cdot \partial \quad (3.4a)$$

(N is an arbitrarily fixed complex number); transversality,

$$y \cdot k = 0; \quad (3.4b)$$

and divergencelessness,

$$\partial \cdot k = 0. \quad (3.4c)$$

Note that the two last conditions imply the tracelessness $k' = 0$. Note, too, that the auxiliary conditions permit one to simplify considerably Eq. (3.3):

$$(Q_0 - E_0(E_0 - 3)) k = 0.$$

An ambiguity arises when we try to characterize $D(E_0,s)$ spaces with the aid of Eqs. (3.3) and (3.4): the latter admit as well a space of solutions carrying the representation $D(3 - E_0,s)$, since $\langle Q_s^{E_0} \rangle = \langle Q_s^{3-E_0} \rangle$ is trivially verified. Besides the explicit action of L_{50}, a way to distinguish between both is to examine the behavior of the states at spatial infinity.[28] It will be possible to achieve this once given the recurrence formulas. The latter permit one to build up the carrier states and make both possibilities appear.

The recurrence formulas make use of constant polarization five-vectors $Z = (Z_\alpha)$ that carry the five-dimensional representation $D(-1,0)$. An orthonormal set of five such vectors is presented in Table I: they are classified as to which energy 0, ± 1 and angular momentum 0, ± 1 they carry.

Moreover, expressing tensors of rank s in terms of tensors of rank $s-1$ involves operators that obey commutation/intertwining rules with the generators $L^{(s)}_{\alpha\beta}$ and the Casimir operator Q_s.

The operator $\Sigma_1 \Theta \cdot Z$: The contraction of the transverse projector Θ with a polarization vector Z permits one to define an operator that makes a symmetric transverse tensor field k of rank s from a symmetric transverse tensor field ζ of rank $s-1$:

$$k = \Sigma_1 \Theta \cdot Z \zeta. \quad (3.5)$$

Such an expression will be a key piece in reducing the tensor product $D(E_0,s-1) \otimes D(-1,0)$. Indeed, the linear span of $\{\Theta \cdot Z\}$, Z taking the values of Table I, is the carrier space of $D(-1,0)$ in terms of transverse tensor fields of rank 1.

The operator $\Sigma_1 \eta(Z,Z')$: Let Z and Z' be two polarization five-vectors. They serve to build up the transverse five-vector field:

$$\eta(Z,Z') = \rho(y \cdot Z Z' - y \cdot Z' Z). \quad (3.6)$$

The linear span of $\{\eta(Z,Z')\}$, Z and Z' taking the values of Table I, is the carrier space of the ten-dimensional representation of SO(3,2): $D(-1,1)$. Table II is a classification of its basis elements according to their energy and angular momentum. The operator $\Sigma_1 \eta(Z,Z')$ makes a symmetric transverse tensor field k of rank s from a symmetric transverse tensor field ζ of rank $s-1$:

$$k = \Sigma_1 \eta(Z,Z') \zeta.$$

This construction will be important in reducing the tensor products

$$D(E_0 \pm 1, s-1) \otimes D(-1,1).$$

The operators $\omega(Z,Z')$ *and* $\Omega_s(Z,Z')$: The five-vectors Z,Z' equally serve to build up the five-component transverse differential operator

$$\bar{\omega}(Z,Z') = \Theta \cdot Z' Z \cdot \bar{\partial} - \Theta \cdot Z Z' \cdot \bar{\partial}. \quad (3.7)$$

We then define $\Omega_s(Z,Z') \equiv T \Sigma_1 \bar{\omega}(Z,Z')$, which makes a symmetric transverse tensor field k of rank s from a symmetric transverse tensor field ζ of rank $s-1$:

TABLE I. Set of five complex five-vectors $Z^a = (Z^a_\alpha)$, where a takes the values $+, -, k+, k-, k$, (i,j,k) being an even permutation of (1,2,3). They span the tangent space of $\mathbb{R}^5$; the latter is provided with the metric $\delta_{\alpha\beta} = (+,-,-,-,+)$. They can be used as orthonormal-basis elements of the carrier space for the five-dimensional fundamental representation $D(-1,0)$. As eigenvectors of $L^{(1)}_{50}$ and $L^{(1)}_{ij}$, they are classified according to the respective eigenvalues.

	Z^+	Z^-	Z^{k+}	Z^{k-}	Z^k
Polarization vector $Z^a = (Z^a_\alpha)$	$Z^+_\alpha = \frac{\delta_{\alpha 5} + i\delta_{\alpha 0}}{\sqrt{2}}$	$Z^-_\alpha = \frac{\delta_{\alpha 5} - i\delta_{\alpha 0}}{\sqrt{2}}$	$Z^{k+}_\alpha = \frac{\delta_{\alpha i} + i\delta_{\alpha j}}{\sqrt{2}}$	$Z^{k-}_\alpha = \frac{\delta_{\alpha i} - i\delta_{\alpha j}}{\sqrt{2}}$	$Z^k_\alpha = \delta_{k\alpha}$
Energy as eigenvalue of $L^{(1)}_{50}$	-1	$+1$	0	0	0
Angular momentum as eigenvalue of $L^{(1)}_{ij}$, (i,j,k) being some even permutation of (1,2,3)	0	0	-1	$+1$	0

$$k = \Omega_s(Z,Z')\zeta. \tag{3.8}$$

The operator $\Sigma_2\Theta$*:* It makes a symmetric transverse tensor field k of rank s from a symmetric transverse tensor field ζ of rank $s-2$:

$$k = \Sigma_2\Theta\zeta.$$

The generalized or "purified" gradient D_s*:* The operator D_s makes a symmetric transverse tensor field k of rank s from a symmetric transverse tensor field ζ of rank $s-1$:

$$k = D_s\zeta \equiv \rho^{-1} T\Sigma_1 \partial\zeta$$
$$= \rho^{-1}\Sigma_1(\bar{\partial} + \rho(s-1)y)\zeta. \tag{3.9}$$

The transverse divergence $\partial^T\cdot$*:* The simple divergence $\partial\cdot k$ of a symmetric transverse field of rank s is not transverse. We thus define

$$\partial^T\cdot k \equiv T\partial\cdot k = \partial\cdot k + \rho\Sigma_1 yk'. \tag{3.10}$$

The following equivalence between conditions should be mentioned: $\partial\cdot k = 0$ if and only if $\partial^T\cdot k = 0$ and $k' = 0$. The features common to the operators Θ, D_s, and ∂^T are their intertwining properties:

$$L^{(s)}_{\alpha\beta}\Sigma_2\Theta\eta = \Sigma_2\Theta L^{(s-2)}_{\alpha\beta}\eta, \tag{3.11a}$$
$$L^{(s)}_{\alpha\beta}D_s\zeta = D_s L^{(s-1)}_{\alpha\beta}\zeta, \tag{3.11b}$$
$$L^{(s-1)}_{\alpha\beta}\partial^T\cdot k = \partial^T\cdot(L^{(s)}_{\alpha\beta}k), \tag{3.11c}$$

where η, ζ, and k are symmetric transverse tensors of rank $s-2$, $s-1$, and s, respectively.

An immediate corollary holds:

$$Q_s\Sigma_2\Theta\eta = \Sigma_2\Theta Q_{s-2}\eta,$$
$$Q_s D_s\zeta = D_s Q_{s-1}\zeta, \tag{3.11d}$$
$$Q_{s-1}\,\partial^T\cdot k = \partial^T\cdot(Q_s k).$$

Moreover, we have the commutation/contraction rules between $\Sigma_2\Theta$, D_s, and $\partial^T\cdot$:

$$D_{s+1}\Sigma_2\Theta\eta = \Sigma_2\Theta D_{s-1}\eta, \tag{3.12}$$
$$\partial^T\cdot\Sigma_2\Theta\eta = \Sigma_2\Theta\,\partial^T\cdot\eta + \rho D_{s-1}\eta, \tag{3.13}$$
$$\partial^T\cdot D_s\zeta = -(Q_{s-1} - \langle Q^{s+1}_s\rangle)\zeta$$
$$-4\Sigma_2\Theta\zeta' + D_{s-1}\,\partial^T\cdot\zeta. \tag{3.14}$$

The intertwining rules obeyed by the polarized operators $\Sigma_1\Theta\cdot Z$, $\Sigma_1\eta(Z,Z')$, and $\Omega_s(Z,Z')$ are more complicated. For this reason, they are given in Appendix A. It is sufficient to note that $\Sigma_1\Theta\cdot Z$, $\Sigma_2\Theta$, and D_s form a closed set with respect to them, as do $\Sigma_1\eta(Z,Z')$, $\Omega_s(Z,Z')$, $\Sigma_2\Theta$, and D_s. These rules are the necessary elements of the (mainly technical) proof of the following recurrence formulas.

Proposition 1 (first recurrence formula)*:* Let k be a carrier state for the representation $D(E_0,s)$ [resp. $D(3-E_0,s)$], with E_0 different from $s+1$ and $2-s$. As a solution of Eqs. (3.3) and (3.4), it is given in terms of tensors of rank $s-1$ and $s-2$ through the recurrence formula

$$k = \Sigma_1\Theta\cdot Z\zeta_1 + \Sigma_2\Theta\zeta_2 + D_s\zeta_3. \tag{3.15}$$

Here, ζ_1 is a carrier state for $D(E_0,s-1)$ [resp. $D(3-E_0,s-1)$] and thus obeys

$$(Q_{s-1} - \langle Q^{E_0}_{s-1}\rangle)\zeta_1 = 0,$$
$$y\cdot\zeta_1 = 0,\quad \partial\cdot\zeta_1 = 0,\quad (\hat{N} - N)\zeta_1 = 0. \tag{3.16}$$

In Eq. (3.15) ζ_2 and ζ_3 are completely determined by ζ_1, $\zeta_2 = -[2/(2s-1)]Z\cdot\zeta_1$ carries $D(E_0,s-2)$ [resp. $D(3-E_0,s-2)$], and

$$\zeta_3 = \frac{1}{(E_0 - s - 1)(E_0 - 2 + s)}$$
$$\times\left[\nabla_Z\zeta_1 - (s+1)\rho y\cdot Z\zeta_1 - \frac{2\rho}{2s-1}D_{s-1}Z\cdot\zeta_1\right].$$

Here,

$$\nabla_Z\zeta = Z\cdot\bar{\partial}\zeta + \rho\Sigma_1 yZ\cdot\zeta \equiv TZ\cdot\bar{\partial}\zeta.$$

TABLE II. The $D(-1,1)$ states $\eta(Z^a,Z^{a'}) = -\eta(Z^{a'},Z^a)$ classified according to their energy and angular momentum. These ten states form an orthogonal basis of the $D(-1,1)$-carrier space.

$Z^{a'}$ \ Z^a	Z^+	Z^-	Z^{k+}	Z^{k-}	Z^k
Z^+		$E=0$, $s=0$	$E=-1$, $s=-1$	$E=-1$, $s=1$	$E=1$, $s=0$
Z^-			$E=1$, $s=-1$	$E=1$, $s=1$	$E=1$, $s=0$
Z^{k+}				$E=0$, $s=0$	$E=0$, $s=-1$
Z^{k-}					$E=0$, $s=1$
Z^k					

The terms making up the expression of ζ_3 can be rearranged (except if $E_0=\frac{3}{2}$) in order to display their group-theoretical meaning:

$$\zeta_3=\frac{1}{2E_0-3}\left[\frac{\eta_-}{E_0+s-2}+\frac{\eta_+}{E_0-s-1}\right]-\frac{2\rho D_{s-1}Z\cdot\zeta_1}{(2s-1)(E_0-s-1)(E_0-2+s)}. \quad (3.17)$$

Here,

$$\eta_-=\nabla_Z\zeta_1+\rho(E_0-3)y\cdot Z\zeta_1+\frac{\rho D_{s-1}Z\cdot\zeta_1}{E_0-s-1}, \quad (3.18)$$

η_- is a carrier state for $D(E_0-1,s-1)$ [resp. $D(4-E_0,s-1)$]; and

$$\eta_+=\nabla_Z\zeta_1-\rho E_0 y\cdot Z\zeta_1-\frac{\rho D_{s-1}Z\cdot\zeta_1}{E_0+s-2}, \quad (3.19)$$

η_+ is a carrier state for $D(E_0+1,s-1)$ [resp. $D(2-E_0,s-1)$].

The reason for the exclusion of $E_0=s+1$ (that is, the lower bound of unitarity) and $E_0=2-s$ is manifest. We shall return to this question later. Let us remark that a degeneracy case appears at $E_0=\frac{3}{2}$, i.e., $E_0=3-E_0$: η_+ and η_- are then identical and the only valid expression for ζ_3 is the first one.

The first recurrence formula simply illustrates the tensor-product reduction:

$$D(E_0,s-1)\otimes D(-1,0)=D(E_0,s)\oplus D(E_0,s-1)\oplus D(E_0,s-2)\oplus D(E_0-1,s-1)\oplus D(E_0+1,s-1) \quad (3.20)$$

[resp. $D(3-E_0,s-1)\otimes D(-1,0)$].

Proposition 2 (second recurrence formula): Let k be a carrier state for the representation $D(E_0,s)$ [resp. $D(3-E_0,s)$], with E_0 different from $s-1$, s, $s+1$, $2-s$, $3-s$, $4-s$. As a solution of Eqs. (3.3) and (3.4), it is given in terms of tensors of rank $s-1$ and $s-2$ through the recurrence formula

$$k=\Omega_s(Z,Z')\zeta_1+\Sigma_1\eta(Z,Z')\zeta_2+\Sigma_2\Theta\zeta_3+D_s\zeta_4. \quad (3.21)$$

Here, ζ_1 is an arbitrary linear combination of two tensors $\zeta_\pm$: $\zeta_1=\mu_+\zeta_++\mu_-\zeta_-$, and should be homogeneous of degree $N+1$. Now, ζ_+ is a carrier state for $D(E_0+1,s-1)$ [resp. $D(2-E_0,s-1)$], ζ_- is a carrier state for $D(E_0-1,s-1)$ [resp. $D(4-E_0,s-1)$], and ζ_2, ζ_3, and ζ_4 are completely determined by $\zeta_\pm$.

First,

$$\zeta_2=\mu_+(E_0-2)\zeta_++\mu_-(1-E_0)\zeta_-.$$

Next, to find ζ_3 and ζ_4, we introduce eigenstates of Q_{s-2} built from $\zeta_\pm$ through contraction with η and ϖ. For convenience we do not specify the (Z,Z') dependence of $\eta(Z,Z')$ and $\varpi(Z,Z')$. Now, $\zeta_{+-}=(E_0-1)\eta\cdot\zeta_++\varpi\cdot\zeta_+$, ζ_{+-} carries $D(E_0,s-2)$ [resp. $D(3-E_0,s-2)$]; $\zeta_{++}=-E_0\eta\cdot\zeta_++\varpi\cdot\zeta_+$, ζ_{++} carries $D(E_0+2,s-2)$ [resp. $D(1-E_0,s-2)$]; $\zeta_{--}=(E_0-3)\eta\cdot\zeta_-+\varpi\cdot\zeta_-$, ζ_{--} carries $D(E_0-2,s-2)$ [resp. $D(5-E_0,s-2)$]; and $\zeta_{-+}=(2-E_0)\eta\cdot\zeta_-+\varpi\cdot\zeta_-$, ζ_{-+} carries $D(E_0,s-2)$ [resp. $D(3-E_0,s-2)$]. We then have

$$\zeta_3=\frac{2\mu_+}{2E_0-1}\left[-\frac{2E_0-3}{2s-1}\zeta_{+-}+\frac{\zeta_{++}}{E_0-s}\right]+\frac{2\mu_-}{2E_0-5}\left[\frac{\zeta_{--}}{E_0+s-3}-\frac{2E_0-3}{2s-1}\zeta_{-+}\right].$$

Another differential operator, denoted by $\mathscr{E}\equiv\mathscr{E}(Z,Z')$, is involved in the expression of ζ_4:

$$\mathscr{E}\zeta\equiv T\eta\cdot\bar\partial\zeta+\rho\Sigma_1(\Theta\cdot Z'Z\cdot\zeta-\Theta\cdot ZZ'\cdot\zeta). \quad (3.22)$$

Further eigenstates of Q_{s-1} are then introduced:

$$\chi_+=\mathscr{E}\zeta_++\frac{2\rho}{(E_0-s+1)(E_0+s-2)}\times D_{s-1}((s-1)\eta\cdot\zeta_+-\varpi\cdot\zeta_+),$$

χ_+ carries $D(E_0+1,s-1)$ [resp. $D(2-E_0,s-1)$]; and

$$\chi_-=\mathscr{E}\zeta_-+\frac{2\rho}{(E_0-s-1)(E_0+s-4)}\times D_{s-1}((s-1)\eta\cdot\zeta_--\varpi\cdot\zeta_-),$$

χ_- carries $D(E_0-1,s-1)$ [resp. $D(4-E_0,s-1)$]. With these definitions, ζ_4 is given by

$$\zeta_4=\mu_+\left[\frac{\chi_+}{E_0-s-1}-\frac{2\rho D_{s-1}\zeta_{+-}}{(2s-1)(2E_0-1)(E_0+s-2)}+\frac{\rho D_{s-1}\zeta_{++}}{(2E_0-1)(E_0-s)(E_0-s+1)}\right]-\mu_-\left[\frac{\chi_-}{E_0+s-2}+\frac{2\rho D_{s-1}\zeta_{-+}}{(2s-1)(2E_0-5)(E_0-s-1)}+\frac{\rho D_{s-1}\zeta_{--}}{(2E_0-5)(E_0+s-3)(E_0+s-4)}\right].$$

The second recurrence formula partially illustrates tensor-product reductions involving the ten-dimensional representation $D(-1,1)$:

$$D(E_0+1,s-1)\otimes D(-1,1)=D(E_0,s)\oplus D(E_0+1,s-1)\oplus D(E_0+2,s-2)\oplus D(E_0,s-2)\oplus[D(E_0+1,s)\oplus D(E_0,s-1)\oplus D(E_0+1,s-2)\oplus D(E_0+2,s)\oplus D(E_0+2,s-1)] \quad (3.23)$$

[resp. $D(E_0-1,s-1)\otimes D(-1,1)$].

The expression between brackets in Eq. (3.23) has no corresponding element in the recurrence formula. The reason for all exclusion cases concerning E_0 is manifest. Degeneracy phenomena occur when $E_0=\frac{1}{2},\frac{3}{2},\frac{5}{2}$: ζ_{+-} and ζ_{++} are identical for $E_0=\frac{1}{2}$ whereas $\zeta_{--}=\zeta_{-+}$ for $E_0=\frac{5}{2}$. The corresponding factors are simply canceled. On the other hand, the recurrence formula can be considered as valid when $E_0=s$ or $s-1$: it is then necessary to put $\mu_+=0$ and the $D(E_0,s)$

states only are used to build up the $D(E_0,s)$ states. Symmetrically, we can put $\mu_- = 0$ when $E_0 = 3 - s$ or $4 - s$. Only $D(E_0 + 1,s - 1)$ states are then used.

The ground state for $D(E_0,0)$,[11] i.e., $y_+^{-E_0} = (y_5 + iy_0)^{-E_0}$ and the recurrence formulas permit one to get the expression of the absolute ground state for $D(E_0,s)$ (described as a symmetric transverse tensor of rank s). First, we have the relationship between the $D(E_0,s-1)$ ground state $g_{E_0}^{s-1}(y)$ and the $D(E_0,s)$ ground state $g_{E_0}^{s}(y)$:

$$g^{s}_{E_0}(y) \propto \Sigma_1 \eta(Z^{k-},Z^{+}) y_+^{-1} g^{s-1}_{E_0}(y),$$

where

$$\eta(Z^{k-},Z^{+}) = \rho(y\cdot Z^{k-}Z^{+} - y\cdot Z^{+}Z^{k-}),$$

i.e.,

$$\eta_\alpha(Z^{k-},Z^{+}) = (\rho/\sqrt{2})((y_i - iy_j)(\delta_{\alpha 5} + i\delta_{\alpha 0}) - y_+(\delta_{\alpha i} - i\delta_{\alpha j})),$$

is the absolute ground state for the ten-dimensional representation $D(-1,1)$. Repeating the induction procedure leads to

$$g^{s}_{E_0}(y) \propto [\eta(Z^{k-},Z^{+})]^{\otimes s} y_+^{-(E_0+s)}, \tag{3.24}$$

where the notation $[\ ]^{\otimes s}$ abusively stands for the symmetrized sth-tensor power of $\eta(Z^{k-},Z^{+})$.

Note that the asymptotic behavior of $g^{s}_{E_0}$ at spatial infinity is independent of s; it is simply determined by the factor $y_+^{-E_0}$. Finally, the whole carrier space of $D(E_0,s)$ is built up through the usual action of energy and angular momentum-shifting operators:

$L^{(s)}_{5i} \mp iL^{(s)}_{0i}$: energy-$\substack{\text{raising}\\ \text{lowering}}$ operators,

$L^{(s)}_{ij} \mp iL^{(s)}_{jk}$: angular momentum-$\substack{\text{raising}\\ \text{lowering}}$ operators,

where (i,j,k) represents an even permutation of $(1,2,3)$.

IV. WEYL EQUIVALENCE BETWEEN REPRESENTATIONS

Let us go back to the difficulty that arose in Sec. III about the determination of the carrier space from the wave equation (3.3) and the auxiliary conditions (3.4). Two different types of solutions are distinguished by their behavior at spatial infinity. The first type that carries $D(E_0,s)$ behaves like $|\mathbf{y}|^{-E_0}$, whereas the second one behaves like $|\mathbf{y}|^{E_0-3}$ and therefore carries $D(3-E_0,s)$. These representations are Weyl equivalent, i.e., they admit the same values for all Casimir operators

$$D(E_0,s) \underset{\mathrm{W}}{\equiv} D(3-E_0,s).$$

These W-equivalent representations may be both unitary [cf. (3.2)] but generally only one of them (the one with greatest E_0) is.

The situation is more complicated when E_0 is integer or half-integer since more representations admit the same value $\langle Q_s^{E_0}\rangle$. The external weights that characterize them are in the W orbit $\Theta_{[E_0,s]}$ of the weight $[E_0,s]$.[27] This W designates the Weyl group of the algebra B_2:

$$\begin{aligned}\Theta_{[E_0,s]} = \{&[E_0,s],[E_0,-s-1],[1-s,1-E_0],\\ &[1-s,E_0-2],[3-E_0,s],\\ &[3-E_0,-s-1],[s+2,1-E_0],\\ &[s+2,E_0-2]\}.\end{aligned}$$

They are obtained from two substitutions:

$$\begin{cases}E_0 \to 3-E_0,\\ s\to s,\end{cases}\quad \begin{cases}E_0\to s+2,\\ s\to E_0-2,\end{cases}$$

corresponding to reflections about the Weyl axis in Fig. 1. Negative spin weights are not possible, and this reduces the number of Weyl-equivalent representations to at most four. Several cases are possible according to the value of

$$\sigma = E_0 - s.$$

The generic case occurs for $\sigma > 3$, with four W-equivalent representations, one of them being unitary and another finite.

$$\begin{array}{ccccc} & & D(s+2,s+\sigma-2) & & \\ D(s+\sigma,s) & \underset{\mathrm{W}}{\equiv} & & \underset{\mathrm{W}}{\equiv} & D(3-s-\sigma,s)\\ \text{unitary} & \underset{\mathrm{W}}{\equiv} & D(1-s,s+\sigma-2) & \underset{\mathrm{W}}{\equiv} & \text{finite}\end{array}$$

(see Fig. 1). In fact, finite irreducible representations occur if

$$E_0 = -s-\sigma,\quad \sigma = 0,1,2,\ldots .$$

Here $D(1-s,s-1)$ is of particular interest: it is Weyl equivalent to a massless representation $D(s,s-1)$ (limit of unitary $\sigma = 1$ or 3, see Fig. 1). There are other interesting cases:

$\sigma = 1,3$ (massless case),

$\sigma = \frac{3}{2},\frac{5}{2}$ (supersymmetric case),

$\left.\begin{array}{l}\sigma = 2\\ E_0 = \frac{3}{2}\end{array}\right\}$ (degenerate cases),

the Dirac multiplet (see Fig. 1 and Refs. 29 and 30).

In the remainder of this paper, only integral values of spin and energy will be considered. We shall make use of the abbreviations Urrep (Firrep) for unitary (finite) irreducible representations.

V. THE INTERTWINING OPERATOR $\mathscr{P}_{s',s}$

The occurrence of the two (generally) nonunitary representations $D(s+2,s+\sigma-2)$ and $D(1-s,s+\sigma-2)$ Weyl-equivalent to the Urrep $D(s+\sigma,s)$ and the Firrep $D(3-s-\sigma,s)$ is intimately related to the existence of a differential operator $\mathscr{P}_{s+\sigma-2,s}$, of order $\sigma - 2$, that intertwines representations carried by tensors of rank s and $s+\sigma-2$, respectively. More precisely, we have the following proposition.

Proposition 3: To any Firrep $D(-s-\sigma,s)$, $\sigma = 0,1,2,\ldots$, one can associate a differential operator $\mathscr{P}_{s+\sigma+1,s}$, of order $\sigma+1$. The action of this operator is defined on transverse symmetric tensor fields of rank s and transforms them into transverse symmetric tensor fields of rank $s+\sigma+1$. Then $\mathscr{P}_{s+\sigma+1,s}$ possesses the following properties.

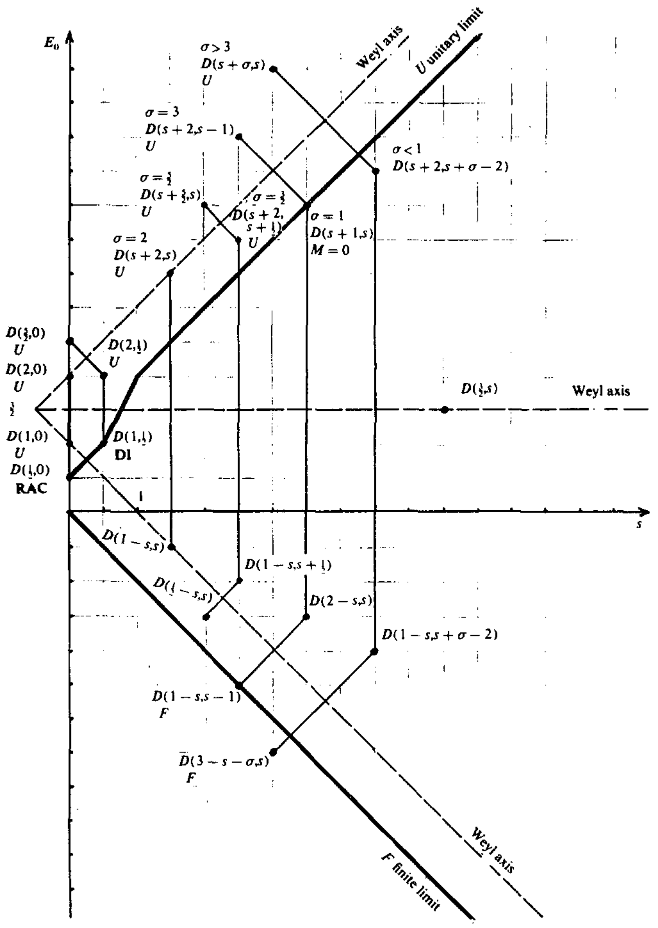


FIG. 1. Weyl equivalence is shown. The generic case (four representations, one unitary and one finite) occurs for $\sigma > 3$. For $\sigma = 3$, there are still four W-equivalent representations but besides the unitary and the finite representations there is a massless representation $D(s+1,s)$. Here $\sigma = \frac{3}{2}$ or $\frac{5}{2}$ corresponds to a "supersymmetric multiplet" since it mixes representations with spin s and $s+\frac{1}{2}$, two representations are unitary and there is no finite representation. Here $\sigma = 2$ is a degenerate case with only two W-equivalent representations (one of them is unitary). The most degenerate case is $D(3/2,s)$ since there is no W-equivalent representation to it. The multiplets with $s < 1$ have special properties: for instance, $\sigma = \frac{3}{2}$, $s = 0$ is the Dirac multiplet. Besides the two unitary representations it contains the two peculiar representations: Di and Rac (unitarity limit).

(i) The operator $\mathscr{P}_{s+\sigma+1,s}$ cancels the space of $D(-s-\sigma,s)$:

$$\mathscr{P}_{s+\sigma+1,s}\zeta_{/} = 0, \quad \text{for all } D(-s-\sigma,s)\text{-state } \zeta_{/}. \tag{5.1}$$

For this reason, we shall sometimes call $\mathscr{P}_{s+\sigma+1,s}$ the "annihilator" of $D(-s-\sigma,s)$.

(ii) The operator $\mathscr{P}_{s+\sigma+1,s}$ obeys the intertwining rules:

$$\mathscr{P}_{s+\sigma+1,s} L^{(s)}_{\alpha\beta} = L^{(s+\sigma+1)}_{\alpha\beta} \mathscr{P}_{s+\sigma+1,s}.$$

As a trivial consequence,

$$\mathscr{P}_{s+\sigma+1,s} Q_s = Q_{s+\sigma+1} \mathscr{P}_{s+\sigma+1,s}. \tag{5.2}$$

(iii) The building blocks of the operator $\mathscr{P}_{s+\sigma+1,s}$ are the generalized gradient D_s, and the operator $\Sigma_2\Theta$, i.e., $\mathscr{P}_{s+\sigma+1,s}$ is a polynomial in $D_{s'}$ and $\Sigma_2\Theta$, and has leading term

$$D_{s+\sigma+1} D_{s+\sigma} \cdots D_{s+1}.$$

Details are given in Appendix B.

The simplest example that comes to us is the transverse gradient $D_1 = \rho^{-1}\bar{\partial} \equiv \mathscr{P}_{1,0}$. It is precisely the annihilator of the constants and the latter can be considered as carrying the trivial representation $D(0,0)$!

We are now in a better position to understand the question of the Weyl equivalence between representations in the general situation described by Fig. 1.

The carrier spaces of the (nonunitary) representations $D(s+2,s+\sigma-2)$ and $D(1-s,s+\sigma-2)$, $\sigma > 3$, are made up of the solutions of the wave equation $(Q_{s+\sigma-2} - \langle Q_s^{s+\sigma}\rangle)k = 0$ supplemented with the auxiliary conditions (3.4). Because of the intertwining rules (ii) and the property (i) stated in Proposition 3, invariant subspaces of solutions might exist that are precisely formed by what we call a "generalized gauge field":

$$\tilde{k} = \mathscr{P}_{s+\sigma-2,s}\zeta, \tag{5.3}$$

provided that ζ obeys the inhomogeneous wave equation,

$$(Q_s - \langle Q_s^{s+\sigma}\rangle)\zeta = \zeta_{/}, \tag{5.4}$$

where $\zeta_{/}$ carries the Firrep $D(3-s-\sigma,s)$. The solution ζ should be of homogeneity degree $N-(\sigma-2)$ in order that k be of degree N and the divergencelessness has to be ensured. The result stated below helps one to answer this question.

Proposition 4: Let $\tilde{k}$ be defined by (5.3). Its trace $\tilde{k}'$ and its traceless divergence $\partial^T\cdot\tilde{k}$ involve six operators $\mathscr{R}^{(i)}_{s_1,s_2}$, $0 \leqslant i \leqslant 6$, $s_1 \geqslant s_2$, that intertwine the generators $L^{(s_1)}_{\alpha\beta}$ and $L^{(s_2)}_{\alpha\beta}$, i.e.,

$$L^{(s_1)}_{\alpha\beta}\mathscr{R}^{(i)}_{s_1,s_2} = \mathscr{R}^{(i)}_{s_1,s_2} L^{(s_2)}_{\alpha\beta}, \tag{5.5}$$

namely,

$$\tilde{k}' = \mathscr{R}^{(1)}_{s+\sigma-4,s}(Q_s - \langle Q_s^{s+\sigma}\rangle)\zeta + \mathscr{R}^{(2)}_{s+\sigma-4,s-2}\zeta' + \mathscr{R}^{(3)}_{s+\sigma-4,s-1}\partial^T\cdot\zeta, \tag{5.6a}$$

$$\partial^T\cdot\tilde{k} = \mathscr{R}^{(4)}_{s+\sigma-3,s}(Q_s - \langle Q_s^{s+\sigma}\rangle)\zeta + \mathscr{R}^{(5)}_{s+\sigma-3,s-2}\zeta' + \mathscr{R}^{(6)}_{s+\sigma-3,s-1}\partial^T\cdot\zeta. \tag{5.6b}$$

The building blocks of the operators $\mathscr{R}^{(i)}_{s_1,s_2}$ are the operators $D_{s'}$ and $\Sigma_2\Theta$.

The proof uses recurrence techniques, intertwining/commutation rules and the following formulas concerning traces:

$$(L^{(s)}_{\alpha\beta}k)' = L^{(s-2)}_{\alpha\beta}k', \tag{5.7a}$$

$$(D_s\zeta)' = D_{s-2}\zeta' + 2\rho^{-1}\partial^T\cdot\zeta, \tag{5.7b}$$

$$(\Sigma_2\Theta\eta)' = \Sigma_2\Theta\eta' + 2s\eta. \tag{5.7c}$$

Here, ζ and η are of rank $s-1$ and $s-2$, respectively.

Let us now assume that the s-rank tensor ζ obeys Eq. (5.4) with $\zeta_{/} = 0$ and is divergenceless, which implies $\zeta' = 0$, $\partial^T\cdot\zeta = 0$. Equations (5.6a) and (5.6b) of Proposition 4 state that $\tilde{k}' = 0$, $\partial^T\cdot\tilde{k} = 0$, i.e., $\tilde{k}$ is also divergenceless. Consequently, there really exists an invariant subspace of solutions of the wave equation (3.1) and its auxiliary conditions (3.2) associated to the representations $D(s+2,s+\sigma-2)$ and $D(1-s,s+\sigma-2)$. What is the representation carried by this subspace? A first candidate is reasonably the Urrep $D(s+\sigma,s)$. Hypothetical indecomposable representations are thus suggested:

$$D(s+2,s+\sigma-2)\rightarrow D(s+\sigma,s)$$

or

$$D(1-s,s+\sigma-2)\rightarrow D(s+\sigma,s)?$$

The arrow means that successive actions of energy-raising and energy-lowering operators onto the fundamental states $g^{s+\sigma-2}_{s+2}$ or $g^{s+\sigma-2}_{1-s}$ make the state $\mathscr{P}_{s+\sigma-2,s}\,g^{s}_{s+\sigma}$ appear in the resulting expression. The second possibility should, however, be excluded because of another remarkable feature of the operator $\mathscr{P}_{s+\sigma-2,s}$.

Proposition 5: All the $(s+\sigma-2)$-rank carrier states of the representation $D(1-s,s+\sigma-2)$ are generalized gauge fields.

Indeed, two realizations of $D(1-s,s+\sigma-2)$-carrier space exist. One of them is made up with tensors of rank $s+\sigma-2$ whereas the other is formed of tensors of rank s up to the *addition* of $D(3-s-\sigma,s)$ states. The corresponding intertwining operator is precisely $\mathscr{P}_{s+\sigma-2,s}$. This fact is put in evidence by the connection between respective ground states:

$$g^{s+\sigma-2}_{1-s}=\mathscr{P}_{s+\sigma-2,s}h^{s+\sigma-2}_{1-s}, \tag{5.8}$$

where

$$h^{s+\sigma-2}_{1-s}\propto[\eta(Z^{k_-},Z^{+})]^{\otimes s}(y\cdot Z^{k_-})^{\sigma-2}y_+^{-1}. \tag{5.9}$$

Moreover, there exists a $D(s+2,s+\sigma-2)$ state $\not{f}^{s+\sigma-2}_{s+2}$ that should leak to the ground state (5.8) when acted on $(2s+1)$ times by a *lowering*-energy operator. This state reads

$$\not{f}^{s+\sigma-2}_{s+2}\propto\Sigma_{\min(s,\sigma-2)}[\eta(Z^{k_-},Z^{+})]^{\otimes s}\times[\eta(Z^{k_-},Z^{k})]^{\otimes\sigma-2}y_+^{-2}. \tag{5.10}$$

The conclusion of this discussion is a first conjecture (verified by hand in the lower-spin cases).

Conjecture 1: The space of solutions of the wave equation $(Q_{s+\sigma-2}-\langle Q^{s+\sigma}_{s}\rangle)k=0$, $\sigma\geqslant 3$, supplemented with the conditions

$$\partial\cdot k=0,\quad(\hat{N}-N)k=0,$$

carries the direct sum of the two indecomposable representations:

$$D(s+2,s+\sigma-2)\rightarrow D(s+\sigma,s), \tag{5.11}$$

$$D(s+2,s+\sigma-2)\rightarrow D(1-s,s+\sigma-2). \tag{5.12}$$

Here $g^{s+\sigma-2}_{s+2}$ is the absolute ground state for the first one whereas $\not{f}^{s+\sigma-2}_{s+2}$ is a cyclic state for the second doublet.

Describing another doublet in which the representation $D(1-s,s+\sigma-2)$ could be involved necessitates a modification of the subsidiary conditions accompanying the wave equation (3.3). For this purpose, let us consider the space of solutions of $(Q_{s+\sigma-2}-\langle Q^{s+\sigma}_{s}\rangle)k=0$ supplemented by

$$(\hat{N}-N)k=0, \tag{5.13a}$$

$$k'=\mathscr{R}^{(1)}_{s+\sigma-3,s}\zeta_{/}, \tag{5.13b}$$

$$\partial^{T}\cdot k=\mathscr{R}^{(4)}_{s+\sigma-4,s}\zeta_{/}, \tag{5.13c}$$

where $\zeta_{/}$ can be *any* state carrying the Firrep $D(3-s-\sigma,s)$. Let us now consider the generalized gauge fields $\tilde{k}=\mathscr{P}_{s+\sigma-2,s}\zeta$ for which ζ is solution of the equation (5.4) *with* the second member $\zeta_{/}$ different from zero. Here ζ and $\zeta_{/}$ are supposed to be of homogeneity degree $N-(\sigma-2)$ and are divergenceless. A recurrence formula, adapted from Proposition 1, can be set up for such solutions ζ and would lead to the following expression for the lowest-energy state:

$$g_{/}=\{1/[2(s+\sigma)-3]\}g^{s}_{3-s-\sigma}\ln\sqrt{\rho}y_{+}. \tag{5.14}$$

This logarithmic ground state satisfies

$$(Q_s-\langle Q^{s+\sigma}_{s}\rangle)g_{/}=g^{s}_{3-s-\sigma}. \tag{5.15}$$

A simple checking leads us to affirm that $g_{/}$ is divergenceless. All the divergenceless, homogeneous solutions of $(Q_s-\langle Q^{s+\sigma}_{s}\rangle)\zeta=\zeta_{/}$ are generated by (5.14) and produce generalized gauge fields $\tilde{k}$ that are solutions of (3.3) and (5.13).

Next, starting from the ground state $g_{/}$ and applying onto it the raising-energy operator $(\sigma-2)$ times permit one to reach the absolute ground state (s-rank version) $h^{s+\sigma-2}_{1-s}$ for $D(1-s,s+\sigma-2)$ *modulo* logarithmic states and $D(3-s-\sigma,s)$ states. If we go higher by applying a number of times raising-energy operators equal to $2(s+\sigma-3)+1$, the maximal length of a ladder in the weight diagram of the Firrep $D(3-s-\sigma,s)$, we see the logarithmic states disappear and only the $D(1-s,s+\sigma-2)$ states, modulo $D(3-s-\sigma,s)$ states, stand in the final expression.

On the other hand, the state $h^{s+\sigma-2}_{1-s}$ leaks to $g^{s}_{3-s-\sigma}$ when acted on a number of times equal to $2-\sigma$ by lowering-energy operators. We conclude that $g_{/}$ is an absolute ground state for the following indecomposable representation:

$$D(3-s-\sigma,s)\rightarrow D(1-s,s+\sigma-2)\rightarrow D(3-s-\sigma,s). \tag{5.16}$$

Now, the expression of the generalized gauge fields contains neither the logarithmic state nor the $D(3-s-\sigma,s)$ state, since

(i) the operator $\mathscr{P}_{s+\sigma-2,s}$ annihilates the latter,

(ii) $\mathscr{P}_{s+\sigma-2,s}\zeta_{/}\ln\sqrt{\rho}y_{+}$

$$=(\mathscr{P}_{s+\sigma-2,s}\zeta_{/})\ln\sqrt{\rho}y_{+}+\sum_{i,j}\mathscr{D}_{i}\zeta_{/}\mathscr{D}_{j}\ln\sqrt{\rho}y_{+},$$

where the symbols $\mathscr{D}_i$ designate derivations. We finally reach the conclusion that the fields $\tilde{k}=\mathscr{P}_{s+\sigma-2,s}\zeta$, where ζ carries the triplet (5.16), carry the indecomposable doublet

$$D(3-s-\sigma,s)\rightarrow D(1-s,s+\sigma-2). \tag{5.17}$$

VI. THE DUAL INTERTWINING OPERATOR $\mathscr{P}^{*}_{s,s'}$ AND THE GENERALIZED GUPTA–BLEULER TRIPLETS

It is obviously possible to go further into the search for more complex indecomposable representations involving the four representations of Fig. 1. However, they will be associated to higher-order wave equations supplemented by less constraining subsidiary conditions. A central role is then played by the operator $\mathscr{P}_{s+\sigma-2,s}$ and its "dual" $\mathscr{P}^*_{s,s+\sigma-2}$, defined as follows.

Definition: The operator $\mathscr{P}^*_{s,s'}$, $s'>s$, is defined on transverse symmetric tensor fields k of rank s' and transforms them into transverse symmetric tensor fields of rank s. It obeys the intertwining rules

$$\mathscr{P}^*_{s,s'}L^{(s')}_{\alpha\beta}=L^{(s)}_{\alpha\beta}\mathscr{P}^*_{s,s'}. \tag{6.1}$$

Its building blocks are

(i) the transverse divergence $\partial^T\cdot$,
(ii) the second-order Casimir operator Q_s,
(iii) the trace,
(iv) the symmetrizer $\Sigma_2\Theta$,
(v) the purified gradient D_s.

Its leading term is

$$\mathscr{P}^*_{s,s'}k=\underbrace{\partial^T\cdot(\partial^T\cdot(\cdots\partial^T\cdot k)\cdots)}_{s'-s\text{ times}}+\cdots.$$

The properties that completely determine its expression proceed from a generalization of the contraction formula (3.14). Simple linear analysis leads to the following statement.

Proposition 6: The dual operator $\mathscr{P}^*_{s,s'}$ is entirely determined by the following conditions.

(i) The tensor field $\mathscr{P}^*_{s,s'}$ is differential of order $s'-s$.

(ii) The tensor field $\mathscr{P}^*_{s,s'}k$ is traceless, whenever k is $(s'-s+1)$th traceless:

$$k^{\overbrace{\cdots}^{s'-s+1}}=0.$$

(iii) When applied to a generalized gauge field $\tilde{k}=\mathscr{P}_{s',s}\xi$ where ξ is traceless, a polynomial expression in the powers of Q_s is obtained:

$$\mathscr{P}^*_{s,s'}\mathscr{P}_{s',s}\xi=\beta(Q_s-\langle Q^{s'+2}_s\rangle)\prod_{i=1}^{s'-s-1}(Q_s-\alpha_i)\xi, \tag{6.2}$$

where the α_i's are strictly positive functions of s and s' and different from $\langle Q^{s'+2}_s\rangle$ for any i, and β is a certain constant.

Some examples of operators $\mathscr{P}_{s',s}$ and $\mathscr{P}^*_{s,s'}$ are given in Appendix B.

Let us now consider the space of solutions of the $2(\sigma-2)$th-order wave equation:

$$\begin{aligned}&\left[-\beta\prod_{i=1}^{\sigma-3}(Q_{s+\sigma-2}-\alpha_i(s,\sigma))\right]\\&\quad\times(Q_{s+\sigma-2}-\langle Q^{s+\sigma}_s\rangle)k\\&\quad+c\mathscr{P}_{s+\sigma-2,s}\mathscr{P}^*_{s,s+\sigma-2}k=0,\end{aligned} \tag{6.3}$$

supplemented by the auxiliary conditions

$$k^{\overbrace{\cdots}^{\sigma-1}}=0,\quad(\hat{N}-N)k=0, \tag{6.4}$$

where c is a constant called the "gauge fixing parameter" for reasons that will appear later.

The generalized gauge fields $\tilde{k}=\mathscr{P}_{s+\sigma-2,s}\xi$, with

$$\xi'=0,\quad[\hat{N}-(N-\sigma+2)]\xi=0, \tag{6.5}$$

clearly form a subspace of solutions if ξ satisfies the inhomogeneous equation

$$(c-1)\beta\left[\prod_{i=1}^{\sigma-3}(Q_s-\alpha_i)\right](Q_s-\langle Q^{s+\sigma}_s\rangle)\xi=\xi_{\not},\tag{6.6}$$

where $\xi_{\not}$ carries the Firrep $D(3-s-\sigma,s)$.

On the other hand, applying the dual operator $\mathscr{P}^*_{s,s+\sigma-2}$ to the left member of Eq. (6.3) leads to the homogeneous equation

$$\begin{aligned}&(c-1)\beta\left[\prod_{i=1}^{\sigma-3}(Q_s-\alpha_i)\right]\\&\quad\times(Q_s-\langle Q^{s+\sigma}_s\rangle)\mathscr{P}^*_{s,s+\sigma-2}k=0.\end{aligned} \tag{6.7}$$

We see that the wave equation (6.3) becomes fully "gauge invariant" if we put $c=1$. For other choices of c, a structure analogous to that of the Gupta–Bleuler triplets of Minkowski QED or of de Sitter QED[15,31] could appear in the space of solutions.

Hereafter, we shall exclude the case $c=1$. Since the divergencelessness $\partial\cdot k=0$ implies $\mathscr{P}^*_{s,s+\sigma-2}k=0$ (each term of the latter contains at least either a trace or a transverse divergence $\partial^T\cdot$), Eqs. (6.3) and (6.4) possess as particular solutions those previously described, i.e., carrying the possible indecomposable representations

$$D(s+2,s+\sigma-2)\to D(s+\sigma,s)$$

and

$$D(s+2,s+\sigma-2)\to D(1-s,s+\sigma-2).$$

Next, let us examine the states k whose "generalized divergences" $\mathscr{P}^*_{s+\sigma-2,s}k$ carry $D(s+\sigma,s)$. They will be called "dual states." For instance, let us seek a state $G^s_{s+\sigma}$ dual of the ground state $g^s_{s+\sigma}$ of $D(s+\sigma,s)$ and solution of Eq. (6.3):

$$\mathscr{P}^*_{s,s+\sigma-2}G^s_{s+\sigma}=g^s_{s+\sigma}. \tag{6.8}$$

Because of the intertwining properties of $\mathscr{P}^*$, the $(s+\sigma-2)$th-rank tensor $G^s_{s+\sigma}$ carries an energy $s+\sigma$ and an angular momentum s, which justifies the notations. The tensor $G^s_{s+\sigma}$ can be written as follows:

$$G^s_{s+\sigma}=\Gamma_0+\Gamma_1,$$

where the expression of the leading term Γ_0 is found to be

$$\Gamma_0\propto\Sigma_{\mathrm{Min}(s,\sigma-2)}[\eta(Z^+,Z^-)]^{\otimes\sigma-2}g^s_{s+\sigma}. \tag{6.9}$$

The second term Γ_1 lies in the linear span of "subgauge terms" like

$$\Theta^{\otimes r}D^{\otimes(2r+t+1)}_{s+\sigma-2-2r-t}\zeta_{r,t},$$

$0\leqslant 2r+t\leqslant\sigma-2$ and $\zeta_{r,t}$ is of rank $s+\sigma-3-2r-t$. The same linear span contains $(Q_{s+\sigma-2}-\langle Q^{s+\sigma}_s\rangle)\Gamma_0$, or equivalently $(Q_{s+\sigma-2}-\langle Q^{s+\sigma}_s\rangle)G^s_{s+\sigma}$.

When acted on by an energy-lowering operator $(\sigma-2)$ times, the state $G^s_{s+\sigma}$ could leak to $g^{s+\sigma-2}_{s+2}$, the first-doublet ground state. Similarly, a state $H^{s+\sigma-2}_{1-s}$, the solution of Eq. (6.3) and dual of $h^{s+\sigma-2}_{1-s}$, introduced in Eq. (5.9), exists. It can be written

$$H^{s+\sigma-2}_{1-s}=\mathscr{H}_0+\mathscr{H}_1,$$

where

$$\mathscr{H}_0\propto\Sigma_{\mathrm{Min}(s,\sigma-2)}[\eta(Z^{k_-},Z^+)]^{\otimes s}(\Theta\cdot Z^{k_-})^{\otimes\sigma-2}y^{-1}_{+} \tag{6.10}$$

and $\mathscr{H}_1$ lies in the linear space of subgauge terms like

$$\Theta^{\otimes r}D^{\otimes(2r+t+1)}_{s+\sigma-2-2r-t}\eta_{r,t}.$$

The same linear span contains $(Q_{s+\sigma-2}-\langle Q^{s+\sigma}_s\rangle)\mathscr{H}_0$ or equivalently $(Q_{s+\sigma-2} - \langle Q_s^{s+\sigma}\rangle)H_{1-s}^{s+\sigma-2}$. When acted on by energy-raising operators $(\sigma-2)$ times, the state $H_{1-s}^{s+\sigma-2}$ could leak to the cyclic state $\not{\ell}_{s+2}^{s+\sigma-2}$ of the second doublet, given by Eq. (5.10) *modulo* generalized gauge fields carrying the doublet (5.17).

Generally Γ_1 and $\mathscr{H}_1$ contain logarithms, except for one value of the gauge-fixing parameter c. This value is conjectured to be

$$c = 2(\sigma-2)/[2(s+\sigma-2)+1]. \tag{6.11}$$

Section VII shows that this result is true for the lowest value of σ: $\sigma = 3$. It has also been verified for the cases $s = 0$, $\sigma = 4$, and $\sigma = 5$.

For such a value of c, the involved indecomposable SO(3,2) representations should reach their simplest ("minimal") triplet structure. A (sub-) space of solutions of Eq. (6.3) and (6.4) carries the direct sum of the following triplets:

$$D(s+\sigma,s) \to D(s+2,s+\sigma-2) \to D(s+\sigma,s), \tag{6.12}$$

$$\begin{array}{c} D(1-s,s+\sigma-2) \to D(s+2,s+\sigma-2) \to D(1-s,s+\sigma-2). \\ \searrow \quad \oplus \quad \nearrow \\ D(3-s-\sigma,s) \end{array} \tag{6.13}$$

VII. MASSLESS INDECOMPOSABLE REPRESENTATION: THIRD RECURRENCE FORMULA

The limit case $\sigma = 3$ combines indecomposability with unitarity. For a sake of convenience, we shall consider here $D(s+1,s)$ rather than $D(s+2,s+1)$. It is presently a well-known result that the representation $D(s+1,s)$ can be considered through quotient carrier spaces as unitary and irreducible. This fact is directly proved from conformal-group considerations[14]: it is the restriction to SO(3,2) of massless representations $\mathscr{T}_n^+$, $s = |n|$, of SO(4,2), the latter being unique extensions of Poincaré massless representations with helicity $\pm s$. The notation is borrowed from Ref. 14, where this result is given. It should be remarked that the restriction of $\mathscr{T}_0^+$ is the sum of the two scalar SO(3,2) Urreps $D(1,0)$ and $D(2,0)$. This de Sitter "dichotomy" takes place at each value of the spin s through the existence of two indecomposable representations where $D(s+1,s)$ occupies the central part:

$$D(s+2,s-1) \to D(s+1,s) \to D(s+2,s-1), \tag{7.1}$$

$$\begin{array}{c} D(2-s,s) \to D(s+1,s) \to D(2-s,s) \\ \searrow \quad \oplus \quad \nearrow \\ D(1-s,s-1) \end{array} \quad . \tag{7.2}$$

The third recurrence formula will give more details of this alternative. It replaces the two previous recurrence formulas at their limits of validity: $E_0 = s+1$ and $E_0 = 2-s$. These singularities are related to the fact that they are reduction points of the minimal weight representations.

The wave equation (6.3) reads, in the present case,

$$(Q_s - \langle Q_s^{s+1}\rangle)k + cD_s\,\partial_s\cdot k = 0, \tag{7.3}$$

where

$$\partial_s\cdot k \equiv \mathscr{P}^*_{s-1,s}k = \partial^T\cdot k - (\rho/2)D_{s-1}k', \tag{7.4}$$

and the subsidiary conditions are

$$y\cdot k = 0, \quad k'' = 0, \quad (\hat{N} - N)k = 0. \tag{7.5}$$

Proposition 7 (third recurrence formula[26]): Let c be different from 1. Let k be a solution of (7.3) and (7.4). Moreover, let us suppose that k is traceless: $k' = 0$. Then k can admit the representation

$$k = \Sigma_1\Theta\cdot Z\zeta_1 + \Sigma_2\Theta\zeta_2 + \rho D_s D_{s-1}\zeta_3 + D_s(\zeta_4 + \zeta_g). \tag{7.6}$$

Here, ζ_1 is a carrier state for the Urrep $D(s+1,s-1)$ or the nonunitary representation $D(2-s,s-1)$. It is divergenceless and homogeneous of degree N. Next, ζ_2 and ζ_3 are given in terms of ζ_1 by

$$\zeta_2 = -\frac{2}{2s-1}Z\cdot\zeta_1 \quad \text{and} \quad \zeta_3 = -\frac{2}{(2s-1)^3}Z\cdot\zeta_1 \tag{7.7}$$

and carry $D(s+1,s-2)$ or $D(3-s,s-1)$. Then ζ_4 is given in terms of ζ_1 by

$$\zeta_4 = [1/(2s-1)^2](\nabla_Z\zeta_1 + \rho(s-2)y\cdot Z\zeta_1) \tag{7.8}$$

(where $\nabla_Z \equiv TZ\cdot\bar{\partial}$) and satisfies

$$[Q_{s-1} - \langle Q_{s-1}^s\rangle]\zeta_4 + [2/(2s-1)]D_{s-1}\,\partial^T\cdot\zeta_4 = 0. \tag{7.9}$$

Finally, ζ_g is given by

$$\zeta_g = \Lambda + \{[c(2s+1)-2]/(1-c)\}\Gamma. \tag{7.10}$$

The tensor field Λ is divergenceless and satisfies

$$[Q_{s-1} - \langle Q_s^{s+1}\rangle]\Lambda = \eta_{/}, \tag{7.11}$$

where $\eta_{/}$ is an arbitrary carrier state for the Firrep $D(1-s,s-1)$. The general solution of (7.11) has the form

$$\Lambda = [Q_{s-1} - \langle Q_s^{s+1}\rangle]^{-1}\eta_{/} \sim \eta_{/}\ln(\sqrt{\rho}y_+) \tag{7.12}$$

[modulo $D(s+2,s-1) \oplus D(2-s,s)$ states].

We here insist on the fact that the presence of a Firrep state in the rhs of (7.11) is due to the canceling property of D_s:

$$D_s\eta_{/} = 0,$$

for any carrier state of $D(1-s,s-1)$. The tensor field Γ satisfies the "dipole equation"

$$[Q_{s-1} - \langle Q_s^{s+1}\rangle]^2\Gamma = 0, \tag{7.13}$$

but is a *particular* solution of that equation:

$$\Gamma = [-1/(2s-1)][Q_{s-1} - \langle Q_s^{s+1}\rangle]^{-1} \times [\nabla_Z\zeta_1 - [1/(2s-1)]\rho D_{s-1}Z\cdot\zeta_1 - (s+1)\rho y\cdot Z\zeta_1]. \tag{7.14}$$

It contributes to logarithms not canceled by the action of D_s.

Proof: The proof of this proposition rests on a few technical relations given in Appendix A. Because of the commutation rules between Q_s and $\partial_s\cdot$, and between $\Sigma_1\Theta\cdot Z$, $\Sigma_2\Theta$, and D_s the general solution of (7.3) is obtained by setting

$$k=\Sigma_1\Theta\cdot Z\zeta_1+\Sigma_2\Theta\zeta_2+D_s\Phi, \tag{7.6'}$$

where ζ_1, ζ_2, and Φ have to obey

$$[Q_{s-1}-\langle Q^{s+1}_{s-1}\rangle]\zeta_1+cD_{s-1}\partial_{s-1}\cdot\zeta_1=0, \tag{7.15}$$

$$[Q_{s-2}-\langle Q^{s+1}_{s-1}\rangle]\zeta_2+cD_{s-2}\partial_{s-2}\cdot\zeta_2=4Z\cdot\zeta_1, \tag{7.16}$$

$$\begin{aligned}(1-c)[Q_{s-1}-\langle Q^{s+1}_s\rangle]\Phi &=\rho((s+3)c-2)y\cdot Z\zeta_1\\ &\quad+c[\rho D_{s-1}(Z\cdot\zeta_1+(s-1)\zeta_2)-TZ\cdot\bar\partial\zeta_1\\ &\quad+\Sigma_2\Theta(4\Phi'-y\cdot Z\zeta_1')\\ &\quad+(\rho/2)D_{s-1}D_{s-2}\Phi']+\eta_{/}.\end{aligned} \tag{7.17}$$

The tracelessness condition on k reads

$$\begin{aligned}k'=0&=\Sigma_1\Theta\cdot Z\zeta_1'+\Sigma_2\Theta\zeta_2'+D_{s-2}\Phi'\\ &\quad+2(Z\cdot\zeta_1+s\zeta_2+\rho^{-1}\partial^T\cdot\Phi).\end{aligned} \tag{7.18}$$

This condition is satisfied if the tracelessness of ζ_1, ζ_2, and Φ ($\zeta_1'=0, \zeta_2'=0, \Phi'=0$) is combined with the equation

$$Z\cdot\zeta_1+s\zeta_2+\rho^{-1}\partial^T\cdot\Phi=0. \tag{7.19}$$

Next, the expression (7.6') has to possess a group-theoretical meaning. It actually displays the reduction, through the leading term $\Sigma_1\Theta\cdot Z\zeta_1$, of the tensor product $D(-1,0)\otimes[D(s+1,s-1)\oplus D(2-s,s-1)]$ that certainly contains the indecomposable representations (7.1) and (7.2). Hence ζ_1 has to be a carrier state for the above direct sum of representations, which imposes its divergencelessness $\partial\cdot\zeta_1=0$. This is consistent with Eq. (7.15), which is reduced to

$$(Q_{s-1}-\langle Q^{s+1}_{s-1}\rangle)\zeta_1=0.$$

Similarly, Eq. (7.16) acquires a group-theoretical meaning if ζ_2 is divergenceless. Its value can be chosen to be

$$[-2/(2s-1)]Z\cdot\zeta_1.$$

The fact that it carries $D(s+1,s-2)$ or $D(3-s,s-1)$ is easily understood through the identities

$$\begin{aligned}(Q_{s-2}-\langle Q^{s+1}_{s-2}\rangle)\zeta_2 &=(Q_0-(s+1)(s-2))\zeta_2\\ &=[-2/(2s-1)](Q_0-(s+1)(s-2))Z\cdot\zeta_1=0.\end{aligned}$$

The inhomogeneous equation (7.17) can be rearranged:

$$\Phi=\tilde\Phi+\Lambda,\quad \Lambda=(Q_{s-1}-\langle Q^{s+1}_s\rangle)^{-1}\eta_{/},$$

$$\begin{aligned}(Q_{s-1}-\langle Q^{s+1}_s\rangle)\tilde\Phi &=[1/(1-c)][-c,c/(2s-1),c(s+3)-2],\end{aligned} \tag{7.20}$$

where $[a,b,c]=a\nabla_Z\zeta_1+b\rho D_{s-1}Z\cdot\zeta_1+c\rho y\cdot Z\zeta_1$ is an element of the three-dimensional space E generated by the three basic functions

$$\nabla_Z\zeta_1\equiv TZ\cdot\bar\partial\zeta_1,\quad \rho D_{s-1}Z\cdot\zeta_1,\quad \rho y\cdot Z\zeta_1. \tag{7.21}$$

However, one can compute the action of $Q_{s-1}-\langle Q^{s+1}_s\rangle$ on each of them. Now, $Q_{s-1}-\langle Q^{s+1}_s\rangle$ leaves E invariant:

$$\begin{aligned}&(Q_{s-1}-\langle Q^{s+1}_s\rangle)\nabla_Z\zeta_1\\ &\quad=-2[s+1,1,(s+1)(s-2)],\\ &(Q_{s-1}-\langle Q^{s+1}_s\rangle)\rho D_{s-1}Z\cdot\zeta_1=-2[0,2s-1,0],\\ &(Q_{s-1}-\langle Q^{s+1}_s\rangle)\rho y\cdot Z\zeta_1=-2[1,0,s-2].\end{aligned} \tag{7.22}$$

Let us, therefore, look for a solution of (7.20) inside E:

$$\tilde\Phi=\Phi_1=[x,y,z]. \tag{7.23}$$

We have to solve a 3×3 system

$$\begin{pmatrix}s+1 & 0 & 1\\ 1 & 2s-1 & 0\\ (s+1)(s-2) & 0 & s-2\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix} =\frac{1}{2(c-1)}\begin{pmatrix}-c\\ c/(2s-1)\\ c(s+3)-2\end{pmatrix}. \tag{7.24}$$

Obviously, the matrix determinant is zero. The first consequence is the existence of a function Φ_0 inside E that satisfies

$$(Q_{s-1}-\langle Q^{s+1}_s\rangle)\Phi_0=0.$$

It is given, up to a multiplicative constant, by

$$\Phi_0=[1,-1/(2s-1),-(s+1)]. \tag{7.25}$$

In fact, Φ_0 is nothing but the $\eta_\pm$ function of the first recurrence formula (RF) and $\partial\cdot k$ is proportional to it. The second consequence is that (7.23) cannot be a solution of (7.20) unless

$$\frac{c(s+3)-2}{1-c}=(s-2)\frac{c}{c-1}\Leftrightarrow c=c_s\equiv\frac{2}{2s+1}. \tag{7.26}$$

In that case, (7.24) can be solved:

$$\Phi_1=[0,-1/(2s-1)^3,1/(2s-1)]+\kappa\Phi_0, \tag{7.27}$$

where κ is an arbitrary constant. Of course, one could drop this $\kappa\Phi_0$ term because of the presence of the arbitrary Λ function inside Φ but it is useful in order to find a group-theoretical meaning to Φ_1. Let us find the equation satisfied by Φ_1. First of all, Φ_0 is divergenceless but Φ_1 is not:

$$\partial^T\cdot\Phi_0=0,\quad D_{s-1}\,\partial^T\cdot\Phi_1=[0,1/(2s-1),0]. \tag{7.28}$$

This last equation is, by the way, compatible with (7.19). Second, using (7.22) we find

$$(Q_{s-1}-\langle Q^s_{s-1}\rangle)\rho D_{s-1}Z\cdot\zeta_1=0, \tag{7.29}$$

and, since this function is divergenceless, that means that inside k, the $\rho D_sD_{s-1}Z\cdot\zeta_1$ term carries the same representation as ζ_2 (since $\langle Q^s_{s-1}\rangle=\langle Q^{s+1}_{s-2}\rangle$). We denote this term by ζ_3 and call ζ_4 what is left from Φ_1.

What can we say about the action of $Q_{s-1}-\langle Q^s_{s-1}\rangle$ on ζ_4? One computes

$$\begin{aligned}(Q_{s-1}-\langle Q^s_{s-1}\rangle)\zeta_4 &=\Bigg[2(2s-1)\left(\kappa-\frac{1}{(2s-1)^2}\right),-2\kappa,\\ &\quad-2(s+1)(2s-1)\left(\kappa-\frac{1}{(2s-1)^2}\right)\Bigg].\end{aligned} \tag{7.30}$$

Since ζ_4 is not divergenceless, we have to compare (7.30) with (7.28)

$$D_{s-1}\,\partial^T\cdot\zeta_4=[0,1/(2s-1),0].$$

We see that ζ_4 satisfies

$$(Q_{s-1}-\langle Q^s_{s-1}\rangle)\zeta_4+[2/(2s-1)]D_{s-1}\,\partial^T\cdot\zeta_4=0, \tag{7.31}$$

if we choose

$$\kappa=1/(2s-1)^2.$$

Its group-theoretical meaning is then obvious and so is the recurrence. So, if $c=2/(2s+1)$, we have $\Phi=\Lambda+\zeta_3+\zeta_4$, with

$$\zeta_3=[0,-2/(2s-1)^3,0],$$
$$\zeta_4=[1/(2s-1)^2,0,(s-2)/(2s-1)^2],$$

or

$$k=\Sigma_1\Theta\cdot Z\zeta_1+\Sigma_2\Theta\zeta_2+\frac{1}{(2s-1)^2}\rho D_sD_{s-1}\zeta_2+D_s(\zeta_4+\Lambda). \tag{7.6$''$}$$

Now, if $c\neq 2/(2s+1)$, one cannot find a solution $\tilde\Phi$ of (7.20) inside E. Thus Φ_1 is not sufficient. Let us call Φ_2 what is necessary in order to have a solution of (7.20):

$$\tilde\Phi=\Phi_1+\Phi_2.$$

This Φ_2 has to satisfy

$$(Q_{s-1}-\langle Q^{s+1}_s\rangle)\Phi_2=\{[(2s+1)c-2]/(2s-1)(c-1)\}\Phi_0, \tag{7.32}$$

where Φ_0 is that particular solution (7.25) of $(Q_{s-1}-\langle Q^{s+1}_s\rangle)\Phi_0=0$. We see therefore that Φ_2 does indeed satisfy a dipole equation

$$(Q_{s-1}-\langle Q^{s+1}_s\rangle)^2\Phi_2=0, \tag{7.33}$$

where Φ_2 is not equal to the general solution of (7.33) but to the general solution of (7.32):

$$\begin{aligned}\Phi_2&=\{[(2s+1)c-2]/(1-c)\}\Gamma,\\ \Gamma&=-[1/(2s-1)](Q_{s-1}-\langle Q^{s+1}_s\rangle)^{-1}\Phi_0.\end{aligned} \tag{7.34}$$

In conclusion

$$k=\Sigma_1\Theta\cdot Z\zeta_1+\Sigma_2\Theta\zeta_2+[1/(2s-1)^2]\rho D_sD_{s-1}\zeta_2+D_s(\zeta_4+\zeta_g), \tag{7.35}$$

with

$$\zeta_g=\Lambda+\{[c(2s+1)-2]/(1-c)\}\Gamma,$$
$$\zeta_2=-[2/(2s-1)]Z\cdot\zeta_1,$$
$$\zeta_4=[1/(2s-1)^2](\nabla_Z\zeta_1+(s-2)\rho y\cdot Z\zeta_1),$$
$$\Gamma=(Q_{s-1}-\langle Q^{s+1}_s\rangle)^{-1}\times\left\{\frac{-1}{2s-1}\nabla_Z\zeta_1+\frac{\rho D_{s-1}Z\cdot\zeta_1}{(2s-1)^2}+\frac{s+1}{2s-1}\rho y\cdot Z\zeta_1\right\}.$$

The tracelessness of k is easily verified. □

When $c\neq 2/(2s+1)$, solutions of (7.3)–(7.5), represented by (7.6), are carrier states of the direct sum of two indecomposable representations. The first one is obtained with the choice that ζ_1 carries $D(s+1,s-1)$:

$$\begin{array}{c} D(s+2,s-1)\to D(s+1,s)\to D(s+2,s-1)\\ \nearrow^{c}\\ D(s+2,s-1)\end{array} \tag{7.36}$$

The second one corresponds to the choice that ζ_1 carries $D(2-s,s-1)$:

$$\begin{array}{c} D(2-s,s)\to D(s+1,s)\to D(2-s,s)\\ \searrow \quad\oplus\quad \uparrow\\ D(2-s,s)\xrightarrow{c}D(1-s,s-1)\end{array} \tag{7.37}$$

The arrows accompanied by the gauge-fixing parameter c indicate the presence in the solution of the term $D_s\Gamma$, where Γ is the solution of the homogeneous dipole equation

$$(Q_{s-1}-\langle Q^{s+1}_s\rangle)^2\Gamma=0.$$

Such solutions can carry the doublet

$$D(s+2,s-1)\to D(s+2,s-1) \tag{7.38}$$

or the triplet

$$D(2-s,s)\to D(1-s,s-1)\to D(2-s,s). \tag{7.39}$$

(Note here the necessary presence of the operator D_s.)

If c is equal to the value $2/(2s+1)$, the indecomposable representations (7.36) and (7.37) reach their simplest structures (7.1) and (7.2) or "minimal" Gupta–Bleuler triplets, thus named by reference to the Poincaré massless representations.[15,31] The third recurrence formula then corresponds to a certain tensor-product reduction, easily revealed through the rewriting of the representation formula:

$$\Sigma_1\Theta Z\zeta_1=k+[2/(2s-1)]\Sigma_2\Theta Z\cdot\zeta_1+[2\rho/(2s-1)^3]D_sD_{s-1}Z\cdot\zeta_1-D_s\zeta_4-D_s\Lambda. \tag{7.40}$$

Let us recall that ζ_4 obeys Eq. (7.31); therefore, it carries the massless indecomposable representations with spin $s-1$ and gauge-fixing parameter $c=2/(2s-1)$. The latter precisely corresponds to the minimal structure. It is then apparent that formula (7.40) illustrates both tensor-product reductions:

$$\begin{aligned}D(-1,0)\otimes D(s+1,s-1)&=D(s+1,s-1)\oplus[D(s+2,s-1)\to D(s+1,s)\to D(s+2,s-1)]\\&\quad\oplus[D(s+1,s-2)\to D(s,s-1)\to D(s+1,s-2)],\end{aligned} \tag{7.41}$$

$$D(-1,0)\otimes D(2-s,s-1)=D(2-s,s-1)\oplus\left(\begin{array}{c}D(2-s,s)\to D(s+1,s)\to D(2-s,s)\\ \searrow\quad\oplus\quad\nearrow\\ D(1-s,s-1)\end{array}\right)$$

$$\oplus\begin{pmatrix} D(3-s,s-1)\rightarrow D(s,s-1)\rightarrow D(3-s,s-1) \\ \searrow \quad \oplus \quad \nearrow \\ D(2-s,s-2) \end{pmatrix}. \tag{7.42}$$

The triplets (7.1) and (7.2), corresponding to $c = 2/(2s+1) \equiv c_s$, are associated with a chain of invariant subspaces $V_g^{c_s} \subset V^{c_s} \subset V'^{c_s}$ in the space of solutions of (7.3)–(7.5). More precisely, we designate by V'^{c_s} an invariant subspace of solutions k square-integrable with respect to a certain invariant indefinite form[32]:

$$(k_1,k_2)_{c_s} = \int \frac{d^3y}{\rho(1/\rho + y^2)} k_1^* \cdot A_{c_s} k_2, \tag{7.43}$$

where A_{c_s} is a certain c_s-dependent matrix differential operator. The solutions that carry the representation (7.1) belong to such a subspace since all representations involved in the triplet are unitary. We now designate by $V^{c_s} \subset V'^{c_s}$ the invariant subspace of tensor fields that are divergenceless, $\partial \cdot k = 0$. For such states, the form is positive semidefinite even for solutions that carry the representation (7.2). On the other hand, each element of the invariant quotient space V'^{c_s}/V^{c_s} can be put in one-to-one correspondence with the nonzero divergence $\partial \cdot k$ of any state k lying in the complement of V^{c_s} in V'^{c_s}. Because of the intertwining rules, the divergence of any solution of Eq. (7.3)–(7.5) obeys

$$(1-c_s)(Q_{s-1} - \langle Q_s^{s+1}\rangle)\partial \cdot k = 0 \tag{7.44}$$

and is thus a candidate to carry $D(s+2,s-1)$ [resp. $D(2-s,s)$]. In the second case, k is not square-integrable since the representation is not unitary. The double divergencelessness $\partial \cdot (\partial \cdot k) = 0$ has been used. It is a consequence of $k' = 0$ and Eq. (7.3). For $s = 1$, the fields just considered should be reminiscent of the scalar photons of Minkowski QED.

In a totally symmetric way, the gauge fields $\tilde{k} = D_s \Lambda$ obey

$$(1-c_s)D_s(Q_{s-1} - \langle Q_s^{s+1}\rangle)\Lambda = 0. \tag{7.45}$$

They carry $D(s+2,s-1)$ [resp. $D(1-s,s-1) \rightarrow D(2-s,s)$] as well. They form an invariant subspace $V_g^{c_s}$ of V^{c_s} made up with the zero-norm states,

$$(\tilde{k},\tilde{k})_{c_s} = 0 \tag{7.46}$$

and, for $s = 1$, such fields should be reminiscent of longitudinal photons.

Finally, the "physical states," i.e., those carrying the actual Urrep $D(s+1,s)$, are put in one-to-one correspondence with the elements of the quotient space $V^{c_s}/V_g^{c_s}$. The form (7.43) is there positive definite and defines a norm separately on each triplet.

It is interesting to characterize the invariant subspace V^{c_s} by a condition other than the divergencelessness $\partial \cdot k = 0$. Applying the divergence operator to the expression (7.6) of k gives the result

$$\partial \cdot k = \frac{1}{1-c_s}\Big[\nabla_Z \zeta_1 - \rho(s+1) y \cdot Z\zeta_1 - \frac{\rho}{2s-1} D_{s-1} Z \cdot \zeta\Big] \equiv \frac{\eta_+}{1-c_s} \equiv \frac{\eta_-}{1-c_s}. \tag{7.47}$$

The $\eta_\pm$ were precisely introduced in the first recurrence formula, Eq. (3.19). The present case, $E_0 = s+1$, makes a degeneracy appear: η_+ becomes identical to η_-. Depending on whether ζ_1 carries $D(s+1,s-1)$ or $D(2-s,s-1)$, η_+ now carries $D(s+2,s-1)$ or $D(2-s,s)$. The subspace V^{c_s} is thus defined by the condition

$$\eta_+ = \eta_- = 0. \tag{7.48}$$

The state k itself can be put into a form that makes apparent its "physical" content besides its "gauge" and "scalar" parts [see Eqs. (7.25) and (7.27)]:

$$\begin{aligned} k = {} & \Sigma_1 \Theta \cdot Z\zeta_1 - [2/(2s-1)]\Sigma_2 \Theta Z \cdot \zeta_1 \\ & + [\rho/(2s-1)] D_s\, y \cdot Z\zeta_1 \\ & - [\rho/(2s-1)^3] D_s D_{s-1} Z \cdot \zeta_1 \\ & + [(1-c_s)/(2s-1)^2] D_s\, \partial \cdot k \\ \equiv {} & [\text{physical part}] + [(1-c_s)/(2s-1)^2] D_s\, \partial \cdot k. \end{aligned} \tag{7.49}$$

Therefore the gauge part only appears coupled to the scalar part.

If c is different from $c_s = 2/(2s+1)$ but still not equal to 1, there exists a corresponding chain $V_g^c \subset V^c \subset V'^c$. The latter is provided with a certain invariant form $(k_1,k_2)_c$ and can be put in one-to-one correspondence with the minimal chain, $c = c_s$, previously described. If k^{c_s} is a square-integrable solution of Eqs. (7.3)–(7.5) for $c = c_s$, a similar solution of (7.3)–(7.5) for $c \neq c_s$ can be built up from it as follows:

$$\begin{aligned} k^c = {} & k^{c_s} + [(c_s - c)/(1-c)] D_s \\ & \times (Q_{s-1} - \langle Q_s^{s+1}\rangle)^{-1} \partial \cdot k^{c_s} \\ \equiv {} & k^{c_s} + [(c_s - c)/(1-c)] D_s \Gamma \end{aligned} \tag{7.50}$$

[see Eq. (7.34)]. Since defined up to the addition of a gauge field, the application (7.50) can be considered as the identity when restricted to the subspace V^{c_s}: $V^c = V^{c_s}$. On the quotient spaces V'^c/V^c and V'^{c_s}/V^{c_s}, Eq. (7.50) induces the relationship

$$\partial \cdot k^c = [(1-c_s)/(1-c)] \partial \cdot k^{c_s}. \tag{7.51}$$

The first indecomposable representation pictured by (7.36) is built up from the scalar massless Urrep $D(2,0)$. Indeed, the general carrier state k is constructed with ζ_1, which carries $D(s+1,s-1)$. Now, the latter can be obtained by applying the second recurrence formula $(s-1)$ times. It is necessary here to put the coefficient μ_+ equal to zero at each step of the induction process.

The same method is used to show that the second indecomposable representation (7.37) is built up from the scalar massless Urrep $D(1,0)$. There ζ_1 carries $D(2-s,s-1)$, the second recurrence formula is applied $(s-1)$ times with $\mu_-=0$ at each step of the construction, and thus $D(1,0)$ is reached.

Let us finally examine the case $c=1$. Equations (7.3) and (7.5) are now "fully" gauge invariant, since $\tilde{k}=D_s\zeta$ is the solution of it for any traceless tensor ζ homogeneous of degree $N-1$. Here, the double-tracelessness condition $k''=0$ acquires its entire significance and k displays a slightly modified representation in terms of tensors of lower rank.

Proposition 8: If c is equal to 1, a solution k of Eqs. (7.3) and (7.5) can have the representation

$$k=\Sigma_1\Theta\cdot Z\zeta_1+\Sigma_2\Theta\zeta_2+D_s\zeta_g, \tag{7.52}$$

where ζ_1 carries $D(s+1,s-1)$ [resp. $D(2-s,s-1)$] and $\zeta_2=-[2/(2s-1)]Z\cdot\zeta_1$ carries $D(s+1,s-2)$ [resp. $D(3-s,s-1)$]. Moreover, ζ_1 has to obey the auxiliary condition

$$\eta_+\equiv Z\cdot\partial^T\zeta_1-\rho(s+1)y\cdot Z\zeta_1 - [\rho/(2s-1)]D_{s-1}Z\cdot\zeta_1=0. \tag{7.53}$$

Now ζ_g only has to be traceless and homogeneous of degree $N-1$. With such a representation, trace and divergence are given by

$$k'=\zeta_2+2\rho^{-1}\partial^T\cdot\zeta_g, \tag{7.54}$$

$$\partial_s\cdot k=-2\rho y\cdot Z\zeta_1-(Q_{s-1}-\langle Q_s^{s+1}\rangle)\zeta_g. \tag{7.55}$$

Putting it another way,

$$k=k_{\text{Phys.}}+D_s\zeta_g, \tag{7.56}$$

with

$$\zeta'_g=0,\quad \partial\cdot k_{\text{Phys.}}=0.$$

VIII. TWO-POINT FUNCTIONS AND THE PROPAGATION PROBLEM

Some of the main features of the irreducible or supposed indecomposable representations previously described are well understood in terms of the invariant two-point functions or homogeneous propagators in biunivocal association with them. The two-point invariant function is an analytic tensor function K of the variable $z=\rho y\cdot y'$ and the quantum field propagators are determined in terms of some of its boundary values.[11,16] The separation between two points y and y' is spacelike if $|z|>1$, timelike if $|z|<1$, and lightlike if $|z|=1$

We understand that propagation is confined to the light cone if and only if K is meromorphic with poles at $z=\pm 1$, because, in this case, the vacuum expectation value of the quantum-field commutator vanishes except on the light cone. On the other hand, branch cuts at $z=\pm 1$ imply "reverberations," that is, propagation in the interior of the light cone.

For a given irreducible representation $D(E_0,s)$, the function $K_{E_0}^{(s)}(z)$ is a solution of the corresponding wave equations (3.3) and (3.4), expressed in terms of the variable y or y'. It can be defined as follows:

$$K_{E_0}^{(s)}(z)=\sum_I k_I(y)k_I^*(y'), \tag{8.1}$$

where $\{k_I\}$ is a set of solutions of the wave equations that is complete for the $D(E_0,s)$-carrier space. It will thus be built up by using the same recurrence formulas as those given in the previous sections: since all the calculations can be drawn back to the scalar case $s=0$, we shall review the main results for the latter situation.

A. Scalar representations—general case

The scalar propagators $K_{E_0}^{(0)}\equiv K_{E_0}$ obey the eigenvalue equation,

$$Q_0K_{E_0}=E_0(E_0-3)K_{E_0},$$

or, in terms of the variable z, are solutions of the differential equation

$$\left[(z^2-1)\frac{d^2}{dz^2}+4z\frac{d}{dz}-E_0(E_0-3)\right]K_{E_0}(z)=0. \tag{8.2}$$

The Gegenbauer functions $C^{3/2}_{-E_0}(z)$ and $C^{3/2}_{E_0-3}(z)$,[28] are the solutions of the above equation for any E_0. They can be expressed as Legendre functions of the first kind:

$$K_{E_0}(z)\propto(E_0-\tfrac{3}{2})(z^2-1)^{-1/2}\mathfrak{P}^1_{E_0-2}(z)\exp(-2i\pi nE_0). \tag{8.3}$$

For a complete study of their singularities, we refer to the discussion of Fronsdal in Ref. 11. Note the existence of the two solutions in relationship with the Weyl equivalence between the two representations $D(E_0,0)$ and $D(3-E_0,0)$.

B. Scalar representations—integer case

Now let us suppose that E_0 is a negative integer $-\sigma$. The representation $D(-\sigma,0)$ is finite and irreducible and the corresponding two-point function is the Gegenbauer polynomial of degree σ:

$$K_{-\sigma}(z)\propto C^{3/2}_\sigma(z). \tag{8.4}$$

For a positive integral value $E_0=\sigma$, the classical result[28] about special function analysis exists. The two-point functions $K_\sigma(z)$ possess simple poles at $z=\pm 1$ for any $\sigma\geqslant 0$. Moreover, they have logarithmic branch cuts at $z=\pm 1$ for any $\sigma\geqslant 3$.

Poles and singularities are displayed by the following formulas:

$$K_1(z)=z(z^2-1)^{-1} \tag{8.5}$$

is the homogeneous propagator for the massless Urrep $D(1,0)$,

$$K_2(z)=(z^2-1)^{-1} \tag{8.6}$$

is the homogeneous propagator for the massless Urrep $D(2,0)$,

$$K_\sigma(z)=\frac{Az+B}{z^2-1}+D_{\sigma-4}(z)+\frac{1}{2}K_{3-\sigma}\ln\frac{z-1}{z+1}, \tag{8.7}$$

for $\sigma\geqslant 3$, is the homogeneous propagator for the scalar Urrep $D(\sigma,0)$. Here, $A=0$ and $B=1$ if σ is even, whereas $A=1$ and $B=0$ if σ is odd. The polynomial D_q is given by

$$D_q(z) = \sum_{k=0}^{[q/2]} \frac{(2q+3-4k)(2k^2-(2q+3)k+(q+2)^2)}{(2k+1)(q+2-k)(q+1-2k)(q+2-2k)} C^{3/2}_{q-2k}(z). \tag{8.8}$$

We remark the role played in Eq. (8.7) by $K_{3-\sigma} \propto C^{3/2}_{\sigma-3}$, propagator of the unique Firrep $D(3-\sigma,0)$ Weyl-equivalent to $D(\sigma,0)$.

C. Representations with nonzero integral spin s—general case

Applying within its domain of validity the first recurrence formula a finite number of times enables us to express the propagator $K^{(s)}_{E_0}(z)$ in terms of $K_{E_0}(z)$. Therefore, the latter will alone carry the possible poles and singularities of $K^{(s)}_{E_0}(z)$.

D. Representations with nonzero integral spin s and integral energy E_0 out of the range $2-s \leqslant E_0 \leqslant s+1$

When E_0 is a negative integer smaller or equal to $-s$, i.e., $E_0 = -\sigma - s$, $\sigma \geqslant 0$, the propagator of the corresponding Firrep is polynomial. The first recurrence formula makes its calculation possible in terms of Gegenbauer polynomials $C^{3/2}_{s+\sigma}(z)$.

When E_0 is a positive integer greater or equal to $s+2$, i.e., $E_0 = s+\sigma$, $\sigma > 1$, the corresponding representation $D(s+\sigma,s)$ is unitary and irreducible. Its propagator can also be determined, through the recurrence formula, from the propagator for $D(s+\sigma,0)$. In particular, poles and (logarithmic) singularities appearing for $D(s+\sigma,s)$ are the same as those encountered in the $D(s+\sigma,0)$ case and specified by Eq. (8.7). The result below follows.

Proposition 9: The homogeneous propagator $K^{(s)}_{s+\sigma}(z)$ for the Urrep $D(s+\sigma,s)$, σ integer $\geqslant 2$, possesses poles at $z = \pm 1$ for any $\sigma \geqslant 1$. It has logarithmic singularities at $z = \pm 1$ for any $\sigma \geqslant 3$. Poles and singularities are displayed by the formulas

$$K^{(s)}_{s+2}(z) = S^{(s)}_2(z)/(z^2-1)^{2s+1}, \tag{8.9}$$

$$K^{(s)}_{s+\sigma}(z) = \frac{S^{(s)}_\sigma(z)}{(z^2-1)^{2s+1}} + \frac{1}{2} K^{(s)}_{3-s-\sigma}(z) \ln \frac{z-1}{z+1}, \tag{8.10}$$

for $\sigma \geqslant 3$. Here $S^{(s)}_\sigma(z)$ is a tensor polynomial in the z variable and $K^{(s)}_{3-s-\sigma}$ is the propagator of the *unique* Firrep representation $D(3-s-\sigma,s)$ Weyl-equivalent to $D(s+\sigma,s)$.

We insist on the fact that, if $\sigma \geqslant 3$, there exist two other representations Weyl-equivalent to the latter, namely $D(s+2,s+\sigma-2)$ and $D(1-s,s+\sigma-2)$, which are infinite and nonunitary. In this context, the case $\sigma = 2$ is remarkable since there exists only one representation, namely $D(1-s,s)$, Weyl-equivalent to $D(s+2,s)$. This fact prevents the appearance of the logarithmic term in the expression of the propagator $K^{(s)}_{s+2}(z)$, a feature obviously shared by the propagator $K^{(s)}_{1-s}(z)$ for $D(1-s,s)$.

Another way to demonstrate this property is to make use of the second recurrence formula repeated s times with $\mu_+ = 0$ for $D(s+2,s)$ and $\mu_- = 0$ for $D(1-s,s)$. We thus reach the $D(2,0)$ propagator for the former and the $D(1,0)$ propagator for the latter, and these functions possess single poles at $z = \pm 1$ only.

A further remark is important for the sequel.

If we consider the generalized gauge fields $\tilde{k} = \mathscr{P}_{s+\sigma-2,s}\zeta$, where ζ carries $D(s+\sigma,s)$, their corresponding propagator is meromorphic at $z = \pm 1$. Indeed, it can be written

$$\begin{aligned}\tilde{K}^{(s+\sigma-2)} &\equiv \mathscr{P}_{s+\sigma-2,s} \mathscr{P}'_{s+\sigma-2,s} K^{(s)}_{s+\sigma}(z) \\ &= \tilde{S}(z)/(z^2-1)^{2(s+\sigma-2)+1},\end{aligned} \tag{8.11}$$

where $\tilde{S}(z)$ is a certain tensor polynomial in the z variable. The disappearance of the logarithmic singularity is due to the specific annihilating property of the differential operator $\mathscr{P}_{s+\sigma-2,s}$:

$$\mathscr{P}_{s+\sigma-2,s} K^{(s)}_{3-s-\sigma}(z) = 0. \tag{8.12}$$

E. The representations $D(1-s,s+\sigma-2)$, $\sigma \geqslant 3$, $s \geqslant 0$

Here, we enter into the "forbidden" range for E_0. In Sec. V, we saw that the carrier states of the indecomposable representation

$$D(3-s-\sigma,s) \to D(1-s,s+\sigma-2), \tag{5.17}$$

$\sigma \geqslant 3$, are the generalized gauge fields $\tilde{k} = \mathscr{P}_{s+\sigma-2,s}\zeta$, ζ being itself a solution of the inhomogeneous equation

$$(Q - \langle Q^{s+\sigma}_s \rangle)\zeta = \text{“}D(3-s-\sigma,s)\text{ state”} \tag{8.13}$$

supplemented by $\partial \cdot \zeta = 0$, $(\hat{N} - (N-\sigma+2))\zeta = 0$. These states ζ carry the triplet

$$\begin{aligned}D(3-s-\sigma,s) &\to D(1-s,s+\sigma-2) \\ &\to D(3-s-\sigma,s).\end{aligned} \tag{5.16}$$

Now, the general solution of (8.13) is given in terms of lower-rank tensors through a recurrence formula adapted from the first one. The corresponding propagator, denoted by $I^{(s)}_{3-s-\sigma}$, is also determined through this recurrence procedure. When applied s times, the latter leads to the resolution of the nonhomogeneous equation satisfied by $I^{(0)}_{3-\sigma} \equiv I_{3-\sigma}$,

$$[Q_0 + \sigma(\sigma-3)] I_{3-\sigma} = K_{3-\sigma}, \tag{8.14}$$

or in terms of the z variable,

$$\left[(z^2-1)\frac{d^2}{dz^2} + 4z\frac{d}{dz} - q(q+3)\right]\varphi(z) = C^{3/2}_q(z). \tag{8.15}$$

Here, we have put $q = \sigma - 3$ and $I_{3-\sigma}(z) = \varphi(z)$. A particular solution of (8.15) is

$$\begin{aligned}\varphi(z) &= (Az+B)/(z^2-1) + E_{q-2}(z) \\ &\quad + \tfrac{1}{2} C^{3/2}_q(z)\ln(z^2-1),\end{aligned} \tag{8.16}$$

where $A = 0$ and $B = 1$ if q is even whereas $A = 1$ and $B = 0$ if q is odd. The polynomial $E_q(z)$ is given by

$$E_q(z) = \sum_{k=0}^{[q/2]} \frac{(2q+3-4k)(2k^2-(2q+3)k+q^2+5q+7)}{(k+1)(2q+5-2k)(q+2-2k)(q+1-2k)} C^{3/2}_{q-2k}(z). \tag{8.17}$$

From the above expression, we can conclude the following.

Proposition 10: The invariant two-point function of the indecomposable representation,

$$D(3-s-\sigma,s) \to D(1-s,s+\sigma-2) \to D(3-s-\sigma,s),$$

σ an integer $\geqslant 3$, possesses poles and logarithmic singularities at $z = \pm 1$. Poles and singularities are displayed by the following formula:

$$I^{(s)}_{3-s-\sigma}(z) = S^{(s)}(z)/(z^2-1)^{2s+1} + \tfrac{1}{2}K^{(s)}_{3-s-\sigma}(z)\ln(z^2-1), \tag{8.18}$$

where $S^{(s)}(z)$ is a tensor polynomial in the z variable.

Therefore, it can be said that the propagator for the generalized gauge fields $\tilde{k} = \mathscr{P}_{s+\sigma-2,s}\zeta$ does not itself contain any logarithmic singularity, since $\mathscr{P}_{s+\sigma-2,s}\mathscr{P}'_{s+\sigma-2,s}$ cancels the Firrep propagator $K^{(s)}_{3-s-\sigma}(z)$:

$$\tilde{I}^{(s)}_{3-s-\sigma}(z) \equiv \mathscr{P}_{s+\sigma-2,s}\mathscr{P}'_{s+\sigma-2,s}I^{(s)}_{3-s-\sigma}(z) = \tilde{S}^{(s)}(z)/(z^2-1)^{2s+1}, \tag{8.19}$$

where $\tilde{S}^{(s)}(z)$ is a tensor polynomial in the z variable.

F. The representations $D(s+2,s+\sigma-2)$, $\sigma \geqslant 3$, $s \geqslant 0$

We saw in Sec. V that the corresponding wave equation is satisfied by the possibly indecomposable representations

$$D(s+2,s+\sigma-2) \to D(s+\sigma,s) \tag{5.11}$$

or

$$D(s+2,s+\sigma-2) \to D(1-s,s+\sigma-2). \tag{5.12}$$

Applying the first recurrence formula $\sigma - 3$ times leads down to the carrier states for the indecomposable representations

$$D(s+2,s+1) \to D(s+3,s)$$

or

$$D(s+2,s+1) \to D(1-s,s+1).$$

Hence the singularities of the propagators for the former doublets are the same as those carried by the latter representations. We are thus led to examine the following case.

G. The massless representation $D(s+1,s)$

The third recurrence formula can be used as well to build up the propagators for the indecomposable representations (7.36) and (7.37) corresponding to a certain choice of the gauge-fixing parameter c different from the value 1. The formula (7.6) expresses them in terms of the meromorphic propagators of the representations $D(s+1,s-1)$ and $D(2-s,s-1)$, up to the addition of gauge-field propagators corresponding to the terms $D_s\Lambda$ and $D_s\Gamma$ [see Eqs. (7.12) and (7.13)]. Here, we reach the ultimate proposition of this paper.

Proposition 11: The homogeneous propagators $K^{(s)}_{\mathrm{I},s+1}(z)$ and $K^{(s)}_{\mathrm{II},s+1}(z)$ corresponding to the indecomposable representations (7.36) and (7.37), respectively, possess poles of multiplicity $2s+1$ at $z = \pm 1$. They have logarithmic singularities at $z = \pm 1$, afforded by their coupled "scalar-gauge" parts, except if the gauge-fixing parameter c is put equal to $c_s = 2/(2s+1)$. More precisely,

$$K^{(s)}_{\mathrm{I},s+1}(z) = P^{(s)}_{\mathrm{I}}(z)/(z^2-1)^{2s+1} + D_sD'_sK^{(s-1)}_{s+2}(z) + \{[c(2s+1)-2]/(1-c)\}D_sD'_sL^{(s-1)}_{\mathrm{I}}(z), \tag{8.20}$$

$$K^{(s)}_{\mathrm{II},s+1}(z) = P^{(s)}_{\mathrm{II}}(z)/(z^2-1)^{2s+1} + D_sD'_sI^{(s-1)}_{1-s}(z) + \{[c(2s+1)-2]/(1-c)\}D_sD'_sL^{(s-1)}_{\mathrm{II}}(z). \tag{8.21}$$

Here, $P^{(s)}_{\mathrm{I}}$ and $P^{(s)}_{\mathrm{II}}$ are polynomial in the z variable; $K^{(s-1)}_{s+2}$ and $I^{(s-1)}_{1-s}$ are given by (8.9) and (8.18), respectively: their own logarithmic singularities are eliminated under the action of $D_sD'_s$.

The propagators $L^{(s-1)}_{\mathrm{I}}$ and $L^{(s-1)}_{\mathrm{II}}$ are the ones corresponding to the "dipole" equation (7.13): their own logarithmic singularities are not eliminated under the action of $D_sD'_s$.

From the above expressions, it can be understood that, at least for one precise choice of the gauge-fixing parameter, the states propagate on the light cone only. Now, the divergenceless states do not depend on the gauge fixing. They propagate on the light cone only, whatever the choice of c. Since, on their own, they carry the indecomposable representations

$$D(s+1,s) \to D(s+2,s-1)$$

or

$$D(s+1,s) \to D(2-s,s),$$

their corresponding propagators are meromorphic. We can then answer the question asked in Sec. VIII F. The propagators for the supposed indecomposable representations (5.11) and (5.12) are meromorphic in the z variable and do not show any logarithmic singularity.

It would be interesting to extend this study to the most general indecomposable representation (6.12) and (6.13). For this purpose, further recurrence formulas are needed, which do not enter into the scope of this paper.

IX. CONCLUSION

Physics on de Sitter space has been truly rich with possibilities during the past decade. One of the last but not the least examples being the remarkable continuity linking the $\mathrm{Osp}(N|4)$ supersymmetry to the Poincaré one through contraction.[33]

It seems to us to be a good opportunity to make more complete the de Sitter glossary concerning the free states and related propagators. However, it should be stressed that the advantages proper to the de Sitter geometry (semisimple kinematic group, discretization of the continuum) have to be moderated by the complexity of writing down the states (compare with the simplicity of the exponential functions of the Poincaré kinematics!).

This work should be extended very soon by considering the half-integer spin case,[34] and by deepening the conformal invariance subjacent to the massless theories,[32,35] in continuation of previous works treating the lower-spin cases.[15,36] In particular, a conformal interpretation of the value $c_s = 2/(2s+1)$ will be given. In this context, we study the role played by the nonunitary representations $D(\sigma,s)$, $2-s \leqslant \sigma \leqslant s$, lying vertically in the Weyl diagram (Fig. 1) between the massless representation $D(s+1,s)$ and its Weyl-equivalent $D(2-s,s)$.

On the other hand, the present paper has to be considered as one of these preparing to deal with really physical problems, like the de Sitter version of the nonlinear Maxwell–Dirac system.

ACKNOWLEDGMENTS

We thank Moshe Flato and Guy Rideau for useful discussions.

APPENDIX A: SOME COMMUTATION/INTERTWINING RULES

All involved tensor fields are symmetric and transverse.

1. Commutation rules involving the operator $\Sigma_1\Theta\cdot Z$

The expression $\Sigma_1\Theta\cdot Z\zeta$, ζ of rank $s-1$, is defined by Eq. (3.5):

(i) trace,

$$(\Sigma_1\Theta\cdot Z\zeta)' = \Sigma_1\Theta\cdot Z\zeta' + 2Z\cdot\zeta;$$

(ii) divergence,

$$\partial\cdot(\Sigma_1\Theta\cdot Z\zeta) = \Sigma_1\Theta\cdot Z\,\partial\cdot\zeta - \rho\Sigma_1\, yZ\cdot\zeta + Z\cdot(\bar\partial - \rho(s+3)y)\zeta;$$

(iii) generator $L^{(s)}_{\alpha\beta}$,

$$L^{(s)}_{\alpha\beta}(\Sigma_1\Theta\cdot Z\zeta) = \Sigma_1\Theta\cdot Z(L^{(s-1)}_{\alpha\beta}\zeta) + i\Sigma_1(Z_\beta\Theta_\alpha - Z_\alpha\Theta_\beta)\zeta;$$

(iv) operator Q_s,

$$Q_s\Sigma_1\Theta\cdot Z\zeta = \Sigma_1\Theta\cdot Z(Q_{s-1} + 2s)\zeta + 2\rho D_s\, y\cdot Z\zeta - 4\Sigma_2\Theta Z\cdot\zeta;$$

(v) operator D_{s+1},

$$D_{s+1}\Sigma_1\Theta\cdot Z\zeta = \Sigma_1\Theta\cdot ZD_s\zeta - 2y\cdot Z\Sigma_2\Theta\zeta.$$

2. Commutation rules involving the operator $\Sigma_1\eta(Z,Z')$

The expression $\Sigma_1\eta(Z,Z')\zeta$, ζ of rank $s-1$, is defined by Eq. (3.6). These rules also involve the operator $\overline{\omega}(Z,Z')$, defined by Eq. (3.7), the operator $\Omega_s(Z,Z') \equiv T\Sigma_1\overline{\omega}(Z,Z')$, and the operator $\mathscr{C}(Z,Z')$, defined by Eq. (3.22). Here, we shall shorten the notation by dropping the (Z,Z') dependence. We have

(i) trace,

$$(\Sigma_1\eta\zeta)' = \Sigma_1\eta\zeta' + 2\eta\cdot\zeta;$$

(ii) divergence,

$$\partial\cdot(\Sigma_1\eta\zeta) = \Sigma_1\eta\,\partial\cdot\zeta + \eta\cdot\bar\partial\zeta + \rho\Sigma_1(Z'Z\cdot\zeta - ZZ'\cdot\zeta);$$

(iii) generator $L^{(s)}_{\alpha\beta}$,

$$L^{(s)}_{\alpha\beta}(\Sigma_1\eta\zeta) = \Sigma_1\eta(L^{(s-1)}_{\alpha\beta}\zeta) + i[(y_\alpha Z_\beta - y_\beta Z_\alpha)Z' - (y_\alpha Z'_\beta - y_\beta Z'_\alpha)Z + (\delta_\alpha\eta_\beta - \delta_\beta\eta_\alpha)]\zeta;$$

(iv) operator Q_s,

$$Q_s\Sigma_1\eta\zeta = \Sigma_1\eta(Q_{s-1} + (2s+4))\zeta - 2\Omega_s\zeta - 4\Sigma_2\Theta\eta\cdot\zeta;$$

(v) operator D_{s+1},

$$D_{s+1}\Sigma_1\eta\zeta = \Sigma_1\eta D_s\zeta.$$

3. Commutation rules involving the operator $\Omega_s(Z,Z')$

The rules are

(i) trace,

$$(\Omega_s\zeta)' = \Omega_{s-2}\zeta' + 2\overline{\omega}\cdot\zeta;$$

(ii) divergence,

$$\partial\cdot(\Omega_s\zeta) = \Omega_{s-1}\,\partial\cdot\zeta - \Sigma_1\eta\partial\cdot\zeta - 2\rho\Sigma_1\, y\overline{\omega}\cdot\zeta + 2\rho(s+2)\Sigma_1\, y\eta\cdot\zeta + (s+2)\eta\cdot\bar\partial\zeta + \rho(s+2)\times\Sigma_1(Z'Z\cdot\zeta - ZZ'\cdot\zeta);$$

(iii) generator $L^{(s)}_{\alpha\beta}$,

$$L^{(s)}_{\alpha\beta}(\Omega_s\zeta) = \Omega_s(L^{(s-1)}_{\alpha\beta}\zeta) - \Sigma_1\eta(L^{(s-1)}_{\alpha\beta}\zeta) + \Sigma_1\eta\Theta_{\alpha\beta}\zeta + i\Sigma_1(y_\alpha\delta_\beta - y_\beta\delta_\alpha)\times(\eta\cdot\bar\partial\zeta + \rho\Sigma_1\, y\eta\cdot\zeta) + iT\Sigma_1(\Theta_\alpha\overline{\omega}_\beta - \Theta_\beta\overline{\omega}_\alpha)\zeta;$$

(iv) operator Q_s,

$$Q_s\Omega_s\zeta = \Omega_s(Q_{s-1} + 2s - 2)\zeta - 2\Sigma_1\eta(Q_{s-1} - s(s-1))\zeta - 2\Sigma_1\eta(\Sigma_2\Theta\zeta') - 4\Sigma_2\Theta\overline{\omega}\cdot\zeta + 4\Sigma_2\Theta\eta\cdot\zeta - 2D_s\mathscr{C}\zeta;$$

(v) operator D_{s+1},

$$D_{s+1}\Omega_s\zeta = \Omega_s D_s\zeta - \Sigma_1\eta D_s\zeta + 2\Sigma_2\Theta\mathscr{C}\zeta.$$

4. Miscellaneous commutation rules involving the second-order Casimir operator Q_s

The rules are

(i) operator $y\cdot Z$,

$$Q_s\, y\cdot Zk = y\cdot Z(Q_s + 4)k - 2\rho^{-1}TZ\cdot\bar\partial k;$$

(ii) operator $TZ\cdot\bar{\partial}\equiv\nabla_Z$,

$$Q_s\nabla_Z k = \nabla_Z(Q_s-2)k - 2\rho y\cdot Z(Q_s - s(s+1))k - 2\rho D_s Z\cdot k - 4\rho y\cdot Z\Sigma_2\Theta k' + 2T\Sigma_1\Theta\cdot Z\,\partial\cdot k;$$

(iii) operator $Z\cdot$,

$$Q_{s-1}Z\cdot k = Z\cdot(Q_s-2s)k + 2y\cdot Z\,\partial\cdot k + 2\Sigma_1 Zk';$$

(iv) operator $\eta(Z,Z')\cdot$,

$$Q_{s-1}\eta\cdot k = \eta\cdot(Q_s - 2(s-1))k - 2\bar{\omega}\cdot k + 2\Sigma_1\eta k';$$

(v) operator $\bar{\omega}(Z,Z')\cdot$,

$$\begin{aligned}Q_{s-1}\bar{\omega}\cdot k &= \bar{\omega}\cdot(Q_s-2s)k \\ &\quad - 2\eta\cdot(Q_s - s(s+1)+2)k \\ &\quad - 6\Sigma_1\eta k' + 2\Sigma_1(\nabla_Z Z' - \nabla_{Z'}Z)k' \\ &\quad - 4\Sigma_2\Theta\eta\cdot k' + 2(y\cdot Z'\nabla_Z - y\cdot Z\nabla_{Z'})\partial\cdot k \\ &\quad + 2T\Sigma_1(\Theta\cdot ZZ'\cdot(\partial\cdot k) - \Theta\cdot Z'Z\cdot(\partial\cdot k));\end{aligned}$$

(vi) operator $\mathscr{C}$,

$$Q_s\mathscr{C}k = \mathscr{C}Q_s k - 4\rho D_s\eta\cdot k.$$

5. Commutation rules involving ∇_Z

The rules are

$$\begin{aligned}\partial^T\cdot(\nabla_Z k) &= \nabla_Z(\partial^T\cdot k) + \rho(s+2)Z\cdot k \\ &\quad - \rho y\cdot Z\,\partial^T\cdot k - \rho\Sigma_1\Theta\cdot Zk',\\ D_s\nabla_Z\zeta &= \nabla_Z D_s\zeta - \rho y\cdot ZD_s\zeta \\ &\quad - (s-1)\Sigma_1\Theta\cdot Z\zeta + 2\Sigma_2\Theta Z\cdot\zeta.\end{aligned}$$

APPENDIX B: THE OPERATORS $\mathscr{P}$ AND $\mathscr{P}^*$

1. Construction of $\mathscr{P}$

The following results make precise the entire structure of $\mathscr{P}_{s',s}$. As a first step, let us consider the scalar finite irreducible representation $D(-\sigma,0)$ and its absolute ground state y_+^σ. Any combination of sums and tensor compositions of $D_{s'}$ and $\Sigma_2\Theta$ whose action cancels y_+^σ will also annihilate the entire carrier space because of the intertwining rules (3.11a) and (3.11b). For any positive integer $r\leqslant\sigma+1$, let us recursively define the polynomial operator $\mathscr{P}_{r,0}(\sigma)$ through the equation

$$\mathscr{P}_{r,0}^{(\sigma)}y_+^\sigma = [r!\sigma!/(\sigma-r)!]\rho^{-r}\Theta_+^{\otimes r}y_+^{\sigma-r}, \tag{B1}$$

where Θ_+ designates the one-rank tensor with components,

$$\Theta_{+\alpha} = \Theta_{5\alpha} + i\Theta_{0\alpha}$$

and $\Theta_+^{\otimes r}$ stands for its r th-tensor power.

The right-hand side of Eq. (B1) is clearly a symmetric transverse tensor field of rank r: $\mathscr{P}_{r,0}(\sigma)$ merely realizes the isomorphism between two different carrier spaces of $D(-\sigma,0)$: scalar fields and r-rank tensors. And the annihilator of the carrier space of $D(-\sigma,0)$ is clearly obtained by putting $r=\sigma+1$:

$$\mathscr{P}_{\sigma+1,0} = \mathscr{P}_{\sigma+1,0}(\sigma). \tag{B2}$$

Proposition A: The differential operator $\mathscr{P}_{r,0}(\sigma)$ is polynomial in the successive symmetrized tensor powers

$$D_{2t+1}^{\otimes(r-2t)}\Theta^{\otimes t} = \Theta^{\otimes t}D_1^{\otimes(r-2t)},\quad 0\leqslant t\leqslant[r/2], \tag{B3}$$

where

$$\Theta^{\otimes t}\zeta = \underbrace{\Sigma_2\Theta(\Sigma_2\Theta\cdots(\Sigma_2\Theta}_{t}\zeta)\cdots)$$

and $D_s^{\otimes p}$ is an (abusive) abbreviated notation for

$$D_{s+p-1}D_{s+p-2}\cdots D_s.$$

Equation (B3) results from the commutation rule (3.12). The polynomial coefficients $S_t^{(r)}(\sigma)$, defined by

$$\mathscr{P}_{r,0}(\sigma) = \sum_{0\leqslant t\leqslant[r/2]}\rho^{-t}S_t^{(r)}(\sigma)\Theta^{\otimes t}D_1^{\otimes(r-2t)}, \tag{B4}$$

are themselves polynomials of degree t in the variable σ. They are determined from the recurrence relationship

$$S_t^{(r+1)}(\sigma) = S_t^{(r)}(\sigma) + 2r(\sigma-r+1)S_{t-1}^{(r-1)}(\sigma), \tag{B5}$$

with the initial conditions

$$S_t^{(r)} = 0,\quad \text{if } t>[r/2],\qquad S_0^{(r)} = 1. \tag{B6}$$

The general expression of these coefficients is not immediate:

$$\begin{aligned}S_t^{(r)}(\sigma) &= 2^{3t}\frac{((r-1)/2)!(\sigma/2)!}{(-\frac{1}{2})!((\sigma-r)/2)!}\\ &\quad\times\sum_{i_1=0}^{t}\sum_{i_2=i_1}^{t}\cdots\sum_{i_n=i_{n-1}}^{t}\prod_{k=1}^{n\equiv r-2t}\frac{((\sigma-r+k-1)/2+i_k)!((r-k-1)/2-i_k)!}{((\sigma-r+k)/2+i_k)!((r-k)/2-i_k)!},\quad \text{if } n\equiv r-2t\geqslant 1,\\ S_{r/2}^{r}(\sigma) &= 2^{3r/2}\frac{((r-1)/2)!(\sigma/2)!}{(-\frac{1}{2})!((\sigma-r)/2)!},\quad \text{if } r=2t.\end{aligned} \tag{B7}$$

Here, $x!$ stands for $\Gamma(x+1)$.

The fact of putting $r=\sigma+1$ does not bring noticeable simplification except in the following cases:

$$S_{(\sigma+1)/2}^{(\sigma+1)}(\sigma) = 2^{3(\sigma+1)/2}\left(\frac{(\sigma/2)!}{(-\frac{1}{2})!}\right)^2,\quad \text{if } \sigma \text{ is odd},$$

$$S_{\sigma/2}^{(\sigma+1)}(\sigma) = 2^{3\sigma/2}[(\sigma/2)!]^2,\quad \text{if } \sigma \text{ is even}.$$

Next, the knowledge of the $D(-\sigma,0)$ annihilator $\mathscr{P}_{\sigma+1,0}$ allows one to obtain straightforwardly the $D(-s-\sigma,s)$ annihilator $\mathscr{P}_{s+\sigma+1,s}$.

Proposition B: The differential operator $\mathscr{P}_{s+\sigma+1,s}$ is polynomial in the successive symmetrized tensor powers

$$\Theta^{\otimes t} D_{s+1}^{\otimes(\sigma+1-2t)}, \quad 0 \leqslant t \leqslant [(\sigma+1)/2],$$

with coefficients equal to those of $\mathscr{P}_{\sigma+1,0}$:

$$\mathscr{P}_{s+\sigma+1,s} = \sum_{0 \leqslant t \leqslant [(\sigma+1)/2]} \rho^{-t} S_t^{(\sigma+1)}(\sigma) \times \Theta^{\otimes t} D_{s+1}^{\otimes(\sigma+1-2t)}. \quad \text{(B8)}$$

The proof is mainly based on the fact that D_{s+1} *is* precisely the $D(-s,s)$ annihilator $\mathscr{P}_{s+1,s}$. To show this, it is sufficient to check the equation

$$D_2 g^1_{-1} \propto \rho^{-1} \Sigma_1 (\bar{\partial} + \rho y) \eta(Z^{k-}, Z^+) = 0,$$

where $g^1_{-1} \propto \eta(Z^{k-}, Z^+)$ is given by (3.24). A simple recurrence argument then leads to

$$D_{s+1}\, g^s_{-s} \propto \rho^{-1} \Sigma_1 (\bar{\partial} + \rho s y) [\eta(Z^{k-}, Z^+)]^{\otimes s} = 0, \quad \text{(B9)}$$

and to the more general commutation rules

$$D_{s+1}^{\otimes p}\, g^s_{-s-\sigma} \propto D_{s+1}^{\otimes p}\, g^s_{-s}\, y_+^\sigma = [\mathrm{Min}(p,s)]!\Sigma_{\mathrm{Min}(p,s)}\, g^s_{-s} D_1^{\otimes p} y_+^\sigma .$$

It follows that

$$\mathscr{P}_{s+\sigma+1,s}\, g^s_{-s-\sigma} \propto \Sigma_{\mathrm{Min}(s,\sigma+1)}\, g^s_{-s} \mathscr{P}_{\sigma+1,0}\, y_+^\sigma = 0. \quad \square$$

2. Examples of operators $\mathscr{P}_{s+\sigma,s}$ and $\mathscr{P}^*_{s,s+\sigma}$

The tensor ζ is a symmetric, transverse, traceless tensor of rank s, and k is a symmetric, transverse, $(\sigma+1)$th-traceless tensor of rank $s+\sigma$.

For the case $\sigma = 1$ ($\zeta' = 0$, $k'' = 0$),

(i) $\mathscr{P}_{s+1,s}\zeta \equiv D_{s+1}\zeta$,

(ii) $\mathscr{P}^*_{s,s+1} k \equiv \partial^T \cdot k - (\rho/2) D_s k'$,

(iii) $\mathscr{P}^*_{s,s+1} \mathscr{P}_{s+1,s}\zeta = -(Q_s - \langle Q_s^{s+3}\rangle)\zeta$.

For the case $\sigma = 2$ ($\zeta' = 0$, $k''' = 0$),

(i) $\mathscr{P}_{s+2,s}\zeta \equiv D_{s+2} D_{s+1}\zeta + 2\rho^{-1}\Sigma_2\Theta\zeta$,

(ii) $$\mathscr{P}^*_{s,s+2} k \equiv \partial^T \cdot \partial^T \cdot k + (\rho/4)(Q_s - 2s(s+5))k' - (\rho/2) D_s\, \partial^T \cdot k' + (\rho^2/8) D_s D_{s-1} k'' + \tfrac{3}{4}\rho\Sigma\Theta k'',$$

(iii) $$\mathscr{P}^*_{s,s+2} \mathscr{P}_{s+2,s}\zeta = \tfrac{3}{2}(Q_s - \langle Q_s^{s+4}\rangle) \times (Q_s - \langle Q_{s-1}^{s+3}\rangle)\zeta.$$

For the case $\sigma = 3$ ($\zeta' = 0$, $k''' = 0$),

(i) $\mathscr{P}_{s+3,s}\zeta \equiv D_{s+3} D_{s+2} D_{s+1}\zeta + (8/\rho)\Sigma_2\Theta D_{s+1}\zeta$,

(ii) $$\begin{aligned}\mathscr{P}^*_{s,s+3} k \equiv{}& \partial^T \cdot \partial^T \cdot \partial^T \cdot k + (\rho/2) \\ &\times (Q_s - 2(s^2+6s+1))\partial^T \cdot k' \\ &- (\rho^2/8)(Q_s - 2(s^2+4s+2)) D_s k'' \\ &- (\rho/2) D_s \partial^T \cdot \partial^T \cdot k' \\ &+ (\rho^2/8) D_s D_{s-1}\, \partial^T \cdot k'' \\ &+ \tfrac{5}{2}\rho\Sigma_2\Theta\partial^T \cdot k'' - \tfrac{11}{12}\rho^2\Sigma_2\Theta D_{s-2} k'' \\ &- (\rho^3/48) D_s D_{s-1} D_{s-2} k''',\end{aligned}$$

(iii) $$\begin{aligned}\mathscr{P}^*_{s,s+3} \mathscr{P}_{s+3,s}\zeta ={}& -3(Q_s - \langle Q_s^{s+5}\rangle) \\ &\times (Q_s - \langle Q_{s+1}^{s+2}\rangle) \\ &\times (Q_s - 2(s^2+2))\zeta.\end{aligned}$$

[1]E. P. Wigner, Proc. Natl. Acad. Sci. USA **36**, 184 (1950).

[2]J. M. Levy-Leblond, Riv. Nuovo Cimento **7**, 187 (1977); Rend. Sc. Int. Fis. Enrico Fermi **72**, 237 (1979).

[3]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973); Y. B. Zel'dovich, Zh. Eksp. Teor. Fiz. Pis'ma **6**, 883 (1967) [Sov. Phys. JETP Lett. **6**, 316 (1967)]; Y. B. Zel'dovich, Usp. Fiz. Nauk. **95**, 209 (1968) [Sov. Phys. Usp. **11**, 381 (1968)].

[4]C. Fronsdal, Rev. Mod. Phys. **37**, 221 (1965).

[5]F. Gürsey, in *Istanbul Summer School 1962* (Gordon and Breach, New York, 1963); F. Gürsey, in *Relativity, Groups and Topology*, Les Houches Summer School 1963 (Gordon and Breach, New York, 1963); F. Gürsey and T. D. Lee, Proc. Natl. Acad. Sci. USA **49**, 179 (1963).

[6]L. Castell, Nuovo Cimento A **61**, 585 (1969).

[7]L. Castell, Nucl. Phys. B **4**, 343 (1967); **5**, 601 (1968).

[8]E. P. Wigner, Ann. Math. **40**, 149 (1939).

[9]V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1948).

[10]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific, Warsaw, 1977).

[11]C. Fronsdal, Phys. Rev. D **10**, 589 (1974).

[12]C. Fronsdal and R. B. Haugen, Phys. Rev. D **12**, 3810 (1975).

[13]C. Fronsdal, Phys. Rev. D **12**, 3819 (1975); **20**, 848 (1979); J. Fang and C. Fronsdal, *ibid.* **22**, 1361 (1980).

[14]E. Angelopoulos, M. Flato, C. Fronsdal, and D. Sternheimer, Phys. Rev. D **23**, 1278 (1981).

[15]B. Binegar, C. Fronsdal, and W. Heidenreich, Ann. Phys. (NY) **149**, 254 (1983).

[16]M. Flato, C. Fronsdal, and J. P. Gazeau, Phys. Rev. D **33**, 415 (1986). In this paper prior claims by S. Deser and R. Nepomechie, Ann. Phys. (NY) **154**, 396 (1984) concerning the "reverberating" propagation of massless fields in anti-de Sitter space were discussed. We refer to the arguments presented therein. Let us just insist on the fact that, as proved throughout the present paper, only gauge modes (through their coupling to the scalar modes) can contribute to reverberation inside the light cone (they do not for a special choice of the gauge-fixing parameter: $2/2s+1$, the anti-de Sitter Feynman gauge). Equations (8.20) and (8.21) clearly exhibit this property. And they show that the physical part of the massless fields [the one that carries the URREP $D(s+1,s)$] cannot reverberate. Our arguments are also supported by conformal-invariance considerations [J. P. Gazeau and M. Hans, to appear in *Proceedings of the XVIth International Colloquium on Group Theoretical Methods in Physics*, Varna, Bulgaria, 1987 (Springer, Berlin, 1988)].

[17]P. A. M. Dirac, J. Math. Phys. **4**, 901 (1963).

[18]M. Flato and C. Fronsdal, Lett. Math. Phys. **2**, 421 (1978). For recent developments, see M. Flato and C. Fronsdal, Phys. Lett. B **172**, 412 (1986); preprint UCLA/87/TEP/41.

[19]H. Bacry and J. M. Levy-Leblond, J. Math. Phys. **9**, 1605 (1968); A. Sudbery, Nucl. Phys. B **44**, 520 (1972).

[20]G. Burdet and M. Perrin, Nuovo Cimento A **25**, 181 (1975).

[21]A. Salam and J. Strathdee, Phys. Lett. B **67**, 429 (1977).

[22]E. Van Beveren, T. A. Rijken, and C. Dullemond, J. Math. Phys. **27**, 1411 (1986).

[23]H. P. Nilles, Phys. Rep. C **110**, 1 (1984).

[24]See, for instance, the lectures by C. de Wit, P. Fré, and P. van Nieuwenhuizen in *Supersymmetry and Supergravity 84*, edited by B. de Wit, P. Fayet, and P. van Nieuwenhuizen (World Scientific, Singapore, 1984).

[25]S. J. Avis, C. J. Isham, and D. Storey, Phys. Rev. D **18**, 3565 (1978).

[26]J. P. Gazeau, Lett. Math. Phys. **8**, 507 (1984).

[27]N. Bourbaki, "Groupes et algèbres de Lie," Chaps. 7 and 8 in *Eléments de Mathématique* (Hermann, Paris, 1975).

[28]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, Berlin, 1966).

[29]C. Fronsdal, Phys. Rev. D **26**, 1988 (1982).

[30]L. Castell and W. Heidenreich, Phys. Rev. D **24**, 371 (1981).

[31]J. P. Gazeau, J. Math. Phys. **26**, 1847 (1985).

[32]J. P. Gazeau, M. Hans, and R. Murenzi, "Invariant bilinear forms on 3 + 2-de Sitter space," to be published.

[33]P. G. O. Freund, *Introduction to Supersymmetry* (Cambridge U. P., Cambridge, 1986).

[34]M. Lesimple, Lett. Math. Phys. **15**, 143 (1988).

[35]See Ref. 16, J. P. Gazeau and M. Hans, Varna, Bulgaria, 1987.

[36]B. Binegar, C. Fronsdal, and W. Heidenreich, J. Math. Phys. **24**, 2828 (1983); Phys. Rev. D **27**, 2249 (1983).

Wigner operator and Racah operator of the Lie superalgebra OSP(1,2)

Gao-Jian Zeng and Xiou-Hua Yuan
Department of Physics, Hunan Normal University, Hunan, People's Republic of China

(Received 15 March 1988; accepted for publication 27 July 1988)

Starting with the Wigner–Eckart theorem about the OSP(1,2) irreducible tensor [J. Phys. A: Math Gen. **20**, 5423 (1987)], the OSP(1,2) Wigner operator and Racah operator are studied, their general definitions, orthogonality properties, and coupling laws are given, and the connections between them and the corresponding operators of the SO(3) algebra are established. The results obtained in this paper are a natural development of the authors' theory on the OSP(1,2) irreducible tensor.

I. INTRODUCTION

For the Lie superalgebra OSP(1,2), we have studied the coupling laws of two and three irreps, calculated and discussed the Wigner coefficients and Racah coefficients,[1] put forward a general definition for the irreducible tensor, strictly proved the Wigner–Eckart theorem about the matrix element of an irreducible tensor, and presented methods for calculating the reduced matrix elements of the irreducible tensor.[2]

We intend to study the OSP(1,2) Wigner operator and Racah operator in this paper. This is a significant work. On one hand, by using the Wigner operator, we can give a fuller description of the irreducible tensor. For instance, we can expand the action of any irreducible tensor on a basis in terms of the Wigner operators; the coefficients in the expansion are just the reduced matrix elements of the irreducible tensor. This result shows that the action of an irreducible tensor on a basis is uniquely determined by its reduced matrix elements. On the other hand, by using the Wigner operators and Racah operators, we may express the theory on coupling of representations in the operator form totally.

The contents of the paper are arranged as follows: In Sec. II, we give a general definition for the OSP(1,2) Wigner operator, decide its orthogonalities and coupling laws, study the relations between the OSP(1,2) Wigner operator and the SO(3) Wigner operator, and establish the concept of the OSP(1,2) Racah invariant. In Sec. III, we give a general definition for the OSP(1,2) Racah operator, decide its orthogonalities and coupling laws, and discuss the relations between the OSP(1,2) Racah operator and Wigner operator. One can see that the theories on the OSP(1,2) Wigner operator and Racah operator in this paper are in agreement with the basic ideas and results in Refs. 1 and 2.

Before studying the OSP(1,2) Wigner operator and Racah operator, we list some fundamental results about the OSP(1,2) Wigner coefficient, Racah coefficient, and irreducible tensor in Refs. 1 and 2. We are to use them in this paper.

(1) The OSP(1,2) Wigner coefficients $\left(\begin{smallmatrix}2J_1 & 2J_2 & 2J\\ M_1 & M_2 & M\end{smallmatrix}\right)$ are defined as follows:

$$|2J,M\rangle = \sum_{M_1M_2} \begin{pmatrix}2J_1 & 2J_2 & 2J\\ M_1 & M_2 & M\end{pmatrix} |2J_1,M_1\rangle|2J_2,M_2\rangle, \quad (1.1a)$$

$$(-1)^{(2J_1-M_1)(2J_2-M_2)}|2J_1,M_1\rangle|2J_2,M_2\rangle = \sum_{2J}(-1)^{2(J_1+J_2-J)(2J-M)} \times \begin{pmatrix}2J_1 & 2J_2 & 2J\\ M_1 & M_2 & M\end{pmatrix}|2J,M\rangle, \quad (1.1b)$$

where $M=M_1+M_2$. They have the orthogonalities

$$\sum_{M_1M_2}(-1)^{(2J_1-M_1)(2J_2-M_2)}\begin{pmatrix}2J_1 & 2J_2 & 2J\\ M_1 & M_2 & M\end{pmatrix} \times \begin{pmatrix}2J_1 & 2J_2 & 2J'\\ M_1 & M_2 & M'\end{pmatrix} = (-1)^{2(J_1+J_2-J)(2J-M)}\delta_{JJ'}\delta_{MM'}, \quad (1.2)$$

$$\sum_{2J}(-1)^{2(J_1+J_2-J)(2J-M)}\begin{pmatrix}2J_1 & 2J_2 & 2J\\ M_1 & M_2 & M\end{pmatrix} \times \begin{pmatrix}2J_1 & 2J_2 & 2J\\ M_1' & M_2' & M\end{pmatrix} = (-1)^{(2J_1-M_1)(2J_2-M_2)}\delta_{M_1M_1'}\delta_{M_2M_2'}, \quad (1.3)$$

and symmetries

$$\begin{pmatrix}2J_1 & 2J_2 & 2J\\ M_1 & M_2 & M\end{pmatrix} = (-1)^{f_1}\begin{pmatrix}2J_2 & 2J_1 & 2J\\ M_2 & M_1 & M\end{pmatrix} \quad (1.4)$$

$$= (-1)^{f_2}\begin{pmatrix}2J_1 & 2J_2 & 2J\\ -M_1 & -M_2 & -M\end{pmatrix} \quad (1.5)$$

$$= (-1)^{f_3}\begin{pmatrix}2J_1 & 2J & 2J_2\\ M_1 & -M & -M_2\end{pmatrix} \quad (1.6)$$

$$= (-1)^{f_4}\begin{pmatrix}2J & 2J_2 & 2J_1\\ -M & M_2 & -M_1\end{pmatrix}, \quad (1.7)$$

where

$$f_1 = n(J_1J_2J) + 2(J_1+J_2-J)\lambda(2J) + \lambda(2J_1)\lambda(2J_2) + \lambda(M_1)\lambda(M_2),$$

$$f_2 = n(J_1J_2J) + 2(J_1+J_2-J)(2J-M) + (2J_1-M_1)(2J_2-M_2),$$

$$f_3 = n(J_1\,0\,M_1/2) + [\lambda(2J_1)+1] \times [2(J_1+J_2-J)+\lambda(M_1)],$$

$$f_4 = n(J_2 M_2/2\,0) + [\lambda(2J_2)+1] \times [2(J_1+J_2-J)+\lambda(M_2)]$$

[see Ref. (1), for the meaning of $n(\ )$].

(2) The OSP(1,2) Racah coefficients $R_{J''J'} = R(J_1J_2JJ_3;J'J'')$ are defined as follows:

$$|2J,M\rangle_{J'} = \sum_{J''} R_{J''J'} |2J,M\rangle_{J''}, \tag{1.8}$$

$$(-1)^{2(J_2+J_3-J'')2(J_1+J''-J)}|2J,M\rangle_{J''} = \sum_{J'} (-1)^{2(J_1+J_2-J')2(J'+J_3-J)} R_{J''J'} |2J,M\rangle_{J'}. \tag{1.9}$$

They have the orthogonalities

$$\sum_{J''} (-1)^{2(J_2+J_3-J'')2(J_1+J''-J)} R_{J''J'} R_{J''\underline{J}'} = (-1)^{2(J_1+J_2-J')2(J'+J_3-J)} \delta_{J'\underline{J}'}, \tag{1.10}$$

$$\sum_{J'} (-1)^{2(J_1+J_2-J')2(J'+J_3-J)} R_{J''J'} R_{\underline{J}''J'} = (-1)^{2(J_2+J_3-J'')2(J_1+J''-J)} \delta_{J''\underline{J}''}, \tag{1.11}$$

and symmetries (for simplicity, we replace J_1, J_2, J, J_3, J', and J'', by a, b, c, d, e, and f, respectively)

$$R(abcd;ef) = (-1)^{g_1} R(badc;ef) \tag{1.12}$$

$$= (-1)^{g_2}(cdab;ef) \tag{1.13}$$

$$= (-1)^{g_3} R(acbd;fe) \tag{1.14}$$

$$= (-1)^{g_4} R(aefd;bc), \tag{1.15}$$

where

$$g_1 = \lambda(2a)2(a+d+e+f) + \lambda(2b)2(b+c+e+f) + 2(a+b+e)2(a+b+c+d),$$

$$g_2 = \lambda(2d)2(a+d+e+f) + \lambda(2b)2(b+c+e+f) + 2(b+d+f)2(a+b+c+d),$$

$$g_3 = 2(b+c+e+f)[\lambda(2b)+\lambda(2c)+1],$$

$$g_4 = n(edc) + n(bdf) + [\lambda(2a)+1]2(b+c+e+f).$$

(3) The OSP(1,2) irreducible tensors T^{2J}_M are defined as follows:

$$\langle q^2_m, T^{2J}_M\rangle = \epsilon(2J,M')\langle 2J,M'|q^2_m|2J,M\rangle T^{2J}_{M'}, \tag{1.16}$$

where q^2_m $(m=0,\pm 1,\pm 2)$ are the generators of the OSP(1,2) algebra, and $\epsilon(2J,M') = \langle 2J,M'|2J,M'\rangle$ is the norm of the left vector. The Wigner–Eckart theorem with respect to the matrix element of the irreducible tensor $T^{2J_2}_{M_2}$ is

$$\epsilon(2J,M)\langle 2J,M|T^{2J_2}_{M_2}|2J_1,M_1\rangle = (-1)^f (2J\|T^{2J_2}\|2J_1) \begin{pmatrix} 2J_1 & 2J_2 & 2J \\ M_1 & M_2 & M \end{pmatrix}, \tag{1.17}$$

where

$$f = 2(J_1+J_2-J)(2J-M) + (2J_1-M_1)(2J_2-M_2) + \lambda(M_1)\lambda(M_2),$$

$\epsilon(2J,M) = \langle 2J,M|2J,M\rangle$ is the norm of the vector $|2J,M\rangle$ generated from the action of $T^{2J_2}_{M_2}$ on the basis $|2J_1M_1\rangle$, and $(2J\|T^{2J_2}\|2J_1)$ is a reduced matrix element of the tensor $T^{2J_2}_{M_2}$.

II. OSP(1,2) WIGNER OPERATOR

A. Definition of the Wigner operator

As we know, the irrep J of the OSP(1,2) algebra is $4J+1$ dimensions. Corresponding to the irrep J, we can define $4J+1$ independent irreducible tensors, while every one of them has $4J+1$ independent components. We denote all these operators by the Gel'fand notation

$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle, \tag{2.1}$$

and call it the OSP(1,2) Wigner operator, where $2J = 0,1,\ldots;\ \Delta,M = 2J,\ldots,-2J$. For convenience, we denote simply the Wigner operator by the symbol $\langle 4J,0\rangle$ sometimes.

The Wigner operator of $J=\frac{1}{2}$ is the fundamental Wigner operator; it is composed of nine operators

$$\left\langle \begin{matrix} & 2 & \\ 2 & & 0 \\ & 2 & \end{matrix} \right\rangle, \left\langle \begin{matrix} & 1 & \\ 2 & & 0 \\ & 2 & \end{matrix} \right\rangle, \left\langle \begin{matrix} & 0 & \\ 2 & & 0 \\ & 2 & \end{matrix} \right\rangle,$$

$$\left\langle \begin{matrix} & 2 & \\ 2 & & 0 \\ & 1 & \end{matrix} \right\rangle, \left\langle \begin{matrix} & 1 & \\ 2 & & 0 \\ & 1 & \end{matrix} \right\rangle, \left\langle \begin{matrix} & 0 & \\ 2 & & 0 \\ & 1 & \end{matrix} \right\rangle,$$

$$\left\langle \begin{matrix} & 2 & \\ 2 & & 0 \\ & 0 & \end{matrix} \right\rangle, \left\langle \begin{matrix} & 1 & \\ 2 & & 0 \\ & 0 & \end{matrix} \right\rangle, \left\langle \begin{matrix} & 0 & \\ 2 & & 0 \\ & 0 & \end{matrix} \right\rangle.$$

The components of the OSP(1,2) Wigner operator may be classified as odd or even. We denote the degree of the Wigner operator by $\lambda(M)$. The degree of the Wigner operators of even $2J-M$ is $\lambda(2J)$.

The OSP(1,2) Wigner operator is defined as follows:

$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle = (-1)^f \begin{pmatrix} 2j & 2J & 2j+\Delta \\ m & M & m+M \end{pmatrix} |2j+\Delta,m+M\rangle, \tag{2.2}$$

where

$$f = (2J-\Delta)(2j+\Delta-m-M) + (2j-m)(2J-M) + \lambda(m)\lambda(M).$$

From the definition (2.2), we see that the Wigner operator $\langle 4J,0\rangle$ transforms the vectors in the space j into that of the space $j+\Delta/2$, and the transformation is impossible unless $|2J-2j| \leqslant 2j+\Delta \leqslant 2J+2j$, that is, j must satisfy $4j \geqslant 2J-\Delta$. If the condition is not satisfied, the Wigner operator $\langle 4J,0\rangle$ will turn the space j into 0. Generally, one names a vector set annihilated by an operator as the null space of the operator. The null spaces of the OSP(1,2) Wigner operator $\langle 4J,0\rangle$ are decided by

$$4j < 2J - \Delta. \tag{2.3}$$

In (2.2), let $j=0$, $m=0$, $\Delta = 2J$; one gets a simple and interesting result:

$$|2J,M\rangle = \left\langle \begin{matrix} & 4J & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |0,0\rangle . \tag{2.4}$$

B. The relations between the OSP(1,2) Wigner operators and SO(3) Wigner operators

We have shown[1] that there is a simple proportional relation between the OSP(1,2) Wigner coefficient and SO(3) Wigner coefficient. We may establish a similar relation for the Wigner operators of the two algebras.

In the usual form of representation, the bases of the SO(3) irrep i are denoted by $|i,m\rangle$, where $m = i, i-1,..., -i$; $i-m$ are always integers; and the SO(3) Wigner coefficients and Wigner operators are denoted by $\left(\begin{smallmatrix} i & I & i' \\ m & M & m' \end{smallmatrix}\right)$ and $\left\langle \begin{smallmatrix} & I+\Delta & \\ 2I & & 0 \\ & I+M & \end{smallmatrix}\right\rangle$, respectively, and the definition of the latter is[3]

$$\left\langle \begin{matrix} & I+\Delta & \\ 2I & & 0 \\ & I+M & \end{matrix} \right\rangle |i,m\rangle = C^{\,i\ \ I\ \ i+\Delta}_{\,m\ M\ m+M} |i+\Delta, m+M\rangle , \tag{2.5}$$

where $M,\Delta = I, I-1,..., -I$. However, in the new form[12] of the OSP(1,2) representation, the bases of the SO(3) irrep i are denoted by $|2i,m\rangle$, where $m = 2i, 2i-2,..., -2i$; $2i-m$ are always even. Correspondingly, the SO(3) Wigner coefficients and Wigner operators should be denoted by $C^{\,2i\ 2I\ 2i'}_{\,m\ M\ m'}$ and $\left\langle \begin{smallmatrix} & 2I+\delta & \\ 4I & & 0 \\ & 2I+M & \end{smallmatrix}\right\rangle$, respectively, and the definition (2.5) should be changed to

$$\left\langle \begin{matrix} & 2I+\delta & \\ 4I & & 0 \\ & 2I+M & \end{matrix} \right\rangle |2i,m\rangle = C^{\,2i\ 2I\ 2i+\delta}_{\,m\ M\ m+M} |2i+\delta, m+M\rangle , \tag{2.6}$$

where $\delta, M = 2I, 2I-2,..., -2I$. In Ref. 1, we have given the relation between the new form and the usual form of the SO(3) Wigner coefficient.

In (2.2), substituting the SO(3) Wigner coefficient for the OSP(1,2) Wigner coefficient, introducing the SO(3) Wigner operator, and considering the fact that $|2j,m\rangle$ are the bases of the SO(3) irrep j (if $2j-m$ is even) or $j-\frac{1}{2}$ (if $2j-m$ is odd), we can establish the connections between the OSP(1,2) Wigner operators and SO(3) Wigner operators. According to the odd and even properties of $2j-m$, $2J-M$, $2J-\Delta$, and $2j+\Delta-m-M$ (see Table I), we can divide these results into eight kinds as follows:

(1)
$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle = (-1)^{\lambda(m)\lambda(M)} \left(\frac{2j+J+\Delta/2+1}{2j+\Delta+1}\right)^{1/2} \times \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle_s |2j,m\rangle_s ,$$

TABLE I. The relations between the OSP(1,2) Wigner operators and SO(3) Wigner operators are displayed according to eight kinds of cases shown in this table. (+ , even; − , odd.)

	$2j-m$	$2J-M$	$2J-\Delta$	$2j+\Delta-m-M$
1	+	+	+	+
2	−	−	−	−
3	+	+	−	−
4	−	−	+	+
5	+	−	+	−
6	+	−	−	+
7	−	+	+	−
8	−	+	−	+

(2)
$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle = (-1)^{\lambda(m)\lambda(M)} \left(\frac{2j+J+\Delta/2+\frac{1}{2}}{2j+\Delta}\right)^{1/2} \times \left\langle \begin{matrix} & 2J+\Delta-1 & \\ 4J-2 & & 0 \\ & 2J+M-1 & \end{matrix} \right\rangle_s |2j-1,m\rangle_s ,$$

(3)
$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle = (-1)^{\lambda(2j)+\lambda(m)\lambda(M)} \left(\frac{J-\Delta/2+\frac{1}{2}}{2j+\Delta}\right)^{1/2} \times \left\langle \begin{matrix} & 2J+\Delta-1 & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle_s |2j,m\rangle_s ,$$

(4)
$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle = (-1)^{\lambda(2j)+\lambda(m)\lambda(M)} \left(\frac{J-\Delta/2}{2j+\Delta+1}\right)^{1/2} \times \left\langle \begin{matrix} & 2J+\Delta & \\ 4J-2 & & 0 \\ & 2J+M-1 & \end{matrix} \right\rangle_s |2j-1,m\rangle_s , \tag{2.7}$$

(5)
$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle = (-1)^{\lambda(2j)+\lambda(m)\lambda(M)} \left(\frac{J+\Delta/2}{2j+\Delta}\right)^{1/2} \times \left\langle \begin{matrix} & 2J+\Delta-2 & \\ 4J-2 & & 0 \\ & 2J+M-1 & \end{matrix} \right\rangle_s |2j,m\rangle_s ,$$

$$(6)\quad \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle$$
$$= (-1)^{\lambda(m)\lambda(M)} \left(\frac{2j-J+\Delta/2+\frac{1}{2}}{2j+\Delta+1}\right)^{1/2}$$
$$\times \left\langle \begin{matrix} & 2J+\Delta-1 & \\ 4J-2 & & 0 \\ & 2J+M-1 & \end{matrix} \right\rangle_s |2j,m\rangle_s,$$

$$(7)\quad \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle$$
$$= (-1)^{\lambda(m)\lambda(M)} \left(\frac{2j-J+\Delta/2}{2j+\Delta}\right)^{1/2}$$
$$\times \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle_s |2j-1,m\rangle_s,$$

$$(8)\quad \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle |2j,m\rangle$$
$$= (-1)^{\lambda(2j)+1+\lambda(m)\lambda(M)} \left(\frac{J+\Delta/2+\frac{1}{2}}{2j+\Delta+1}\right)^{1/2}$$
$$\times \left\langle \begin{matrix} & 2J+\Delta+1 & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle_s |2j-1,m\rangle_s,$$

where the subscript s stands for SO(3).

C. Adjoint operation of the Wigner operator

In order to express properly the orthogonality relations of the Wigner operators, it is necessary to introduce an adjoint operation (+). Our definition is as follows:

$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle^{\dagger} |2j,m\rangle$$
$$= (-1)^{\lambda(m-M)\lambda(M)+n(J,\Delta/2,0)+n(J,M/2,0)}$$
$$\times \begin{pmatrix} 2j-\Delta & 2J & 2j \\ m-M & M & m \end{pmatrix} |2j-\Delta,m-M\rangle. \qquad (2.8)$$

Applying the symmetry relations (1.4)–(1.7) of the OSP(1,2) Wigner coefficients, we easily prove that the adjoint operation defined in this way has the following properties:

$$\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle^{\dagger} = (-1)^{[\lambda(2J)+1][2J-\Delta]+\lambda(2J)\lambda(M)}$$
$$\times \left\langle \begin{matrix} & 2J-\Delta & \\ 4J & & 0 \\ & 2J-M & \end{matrix} \right\rangle, \qquad (2.9)$$

$$\left[\left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle^{\dagger}\right]^{\dagger} = \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle. \qquad (2.10)$$

D. Orthogonality relations of the Wigner operators

Applying the adjoint definition (2.8), we can express the orthogonality properties of the Wigner operators in very simple and beautiful forms:

$$\sum_M (-1)^{n(J,M/2,0)} \left\langle \begin{matrix} & 2J+\Delta' & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle$$
$$\times \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle^{\dagger}$$
$$= (-1)^{n(J,\Delta/2,0)} I \delta_{\Delta\Delta'}, \qquad (2.11)$$

$$\sum_\Delta (-1)^{n(J,\Delta/2,0)} \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M' & \end{matrix} \right\rangle^{\dagger}$$
$$\times \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle$$
$$= (-1)^{n(J,M/2,0)} I \delta_{MM'}, \qquad (2.12)$$

$$\sum_m (-1)^{2j-m} \epsilon(2j,m) \langle 2j,m| \left\langle \begin{matrix} & 2J'+\Delta' & \\ 4J' & & 0 \\ & 2J'+M' & \end{matrix} \right\rangle$$
$$\times \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle^{\dagger} |2j,m\rangle$$
$$= (-1)^{2J-M+n(J,\Delta/2,0)+n(J,M/2,0)} \delta_{JJ'} \delta_{\Delta\Delta'} \delta_{MM'}, \qquad (2.13)$$

where I is a unit operator. The relations (2.11)–(2.13) coincide with the orthogonality relations (1.2) and (1.3) of the OSP (1,2) Wigner coefficients; it is not difficult for the reader to check this.

E. An arbitrary irreducible tensor can be expressed in terms of the Wigner operators

Comparing relation (2.2) with (1.17), one can see that the Wigner operator defined by (2.2) has the characteristics of the unit irreducible tensor; the relation (2.2) is actually the Wigner–Eckart theorem about the unit irreducible tensor. Hence we can express the action of an arbitrary irreducible tensor T^{2J}_M on the space j in terms of $\langle 4J,0\rangle$ as

$$T^{2J}_M |2j,m\rangle = \left(\sum_\Delta (2j+\Delta \| T^{2J} \| 2j) \times \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle \right) |2j,m\rangle. \qquad (2.14)$$

The relation above shows that the action of an irreducible tensor on a given space is only determined by its reduced matrix elements.

The relation (2.14) may be rewritten in an operator form as

$$T^{2J}_M = \sum_\Delta R^{2J}(\Delta) \left\langle \begin{matrix} & 2J+\Delta & \\ 4J & & 0 \\ & 2J+M & \end{matrix} \right\rangle,$$

where $R^{2J}(\Delta)$ is an invariant operator whose result on the basis $|2j,m\rangle$ is irrelevant to m:

$$R^{2J}(\Delta)|2j,m\rangle = R'^{2J}(\Delta,j)|2j,m\rangle.$$

It is not difficult to see that, in fact, $R'^{2J}(\Delta,j)$ is a reduced matrix element of T^{2J}_M.

F. Coupling laws of the Wigner operators

In order to study the coupling laws of the OSP(1,2) Wigner operators, we define a new operator $R^{2a2b2c}_{\rho\,\sigma\,\tau}$, which is called Racah invariant. Its definition is

$$R^{2a2b2c}_{\rho\,\sigma\,\tau}|2j,m\rangle = R^{2a2b2c}_{\rho\,\sigma\,\tau}(2j)|2j,m\rangle, \quad \tau=\rho+\sigma, \tag{2.15}$$

where

$$R^{2a2b2c}_{\rho\,\sigma\,\tau}(2j) = (-1)^{2(a+b-c)(2c-\tau)+(2a-\rho)(2b-\sigma)} \times R(j-\tau/2,ajb;j-\sigma/2,c), \tag{2.16}$$

which is a new notation for the OSP(1,2) Racah coefficient and satisfies the same orthogonality relations as the Wigner coefficients in the form.

It is necessary to point out that $R^{2a2b2c}_{\rho\,\sigma\,\tau}(2j)$ has two important properties: (a) the movement property,

$$R^{2a2b2c}_{\rho\,\sigma\,\tau}(2j) = (-1)^k R^{2b\;\;2a\;\;2c}_{-\sigma-\rho-\tau}(2j-\tau),$$
$$k=\lambda(2a)(2a+2c-\sigma) + \lambda(2b)(2b+2c-\rho) + 2(a+b-c)(2a+2b-\tau), \tag{2.17}$$

which may be proved by using the symmetry relations of Racah coefficients; and (b) the asymptotic property

$$\lim_{j\to\infty} R^{2a2b2c}_{\rho\,\sigma\,\tau}(2j) = \begin{pmatrix} 2a & 2b & 2c \\ \rho & \sigma & \tau \end{pmatrix}, \tag{2.18}$$

which will be proved in Appendix A. This result asserts that the OSP(1,2) Wigner coefficient may be obtained as the asymptotic limit of the OSP(1,2) Racah coefficient.

From the orthogonality relations (1.10) and (1.11) of the Racah coefficients, we can prove that the Racah invariant operators have the following orthogonalities:

$$\sum_{\rho,\sigma}(-1)^{(2a-\rho)(2b-\sigma)} R^{2a2b2c}_{\rho\,\sigma\,\tau} R^{2a2b2c'}_{\rho\,c\,\tau} = (-1)^{2(a+b-c)(2c-\tau)} I\delta_{cc'}, \tag{2.19}$$

$$\sum_{c}(-1)^{2(a+b-c)(2c-\tau)} R^{2a2b2c}_{\rho\,\sigma\,\tau} R^{2a2b\,2c}_{\rho'\sigma'\tau} = (-1)^{(2a-\rho)(2b-\sigma)} I\delta_{\rho\rho'}\delta_{\sigma\sigma'}. \tag{2.20}$$

Now we work out an operator from two Wigner operators $\langle 4a,0\rangle$ and $\langle 4b,0\rangle$:

$$T^{2c}_{\gamma}(\sigma,\rho) = \sum_{\alpha,\beta}(-1)^{\lambda(\alpha)\lambda(\beta)} \begin{pmatrix} 2a & 2b & 2c \\ \alpha & \beta & \gamma \end{pmatrix} \times \left\langle \begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix} \right\rangle \left\langle \begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix} \right\rangle. \tag{2.21}$$

Interchanging a and b in the Wigner coefficient, and comparing the relation (2.21) with the definition of the product of two irreducible tensors,[2] we can see that $T^{2c}_{\gamma}(\sigma,\rho)$ is also an irreducible tensor. Since the Wigner–Eckart theorem about the matrix element of $T^{2c}_{\gamma}(\sigma,\rho)$ can be written as

$$\epsilon(2j+\tau,m+\gamma)\langle 2j+\tau,m+\gamma|T^{2c}_{\gamma}(\sigma,\rho)|2j,m\rangle = (2j+\tau\|\mathbf{T}^{2c}(\sigma,\rho)\|2j)\epsilon(2j+\tau,m+\gamma) \times \langle 2j+\tau,m+\gamma| \left\langle \begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix} \right\rangle |2j,m\rangle,$$

$T^{2c}_{\gamma}(\sigma,\rho)$ may be expressed in terms of $\langle 4c,0\rangle$,

$$T^{2c}_{\gamma}(\sigma,\rho) = R^{2a2b2c}_{\rho\,\sigma\,\tau} \left\langle \begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix} \right\rangle. \tag{2.22}$$

Here we have introduced the Racah invariant $R^{2a2b2c}_{\rho\,\sigma\,\tau}$. From (2.21) and (2.22), one has

$$R^{2a2b2c}_{\rho\,\sigma\,\tau} \left\langle \begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix} \right\rangle = \sum_{\alpha,\beta}(-1)^{\lambda(\alpha)\lambda(\beta)} \begin{pmatrix} 2a & 2b & 2c \\ \alpha & \beta & \gamma \end{pmatrix} \times \left\langle \begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix} \right\rangle \left\langle \begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix} \right\rangle. \tag{2.23}$$

This is the coupling law of the Wigner operators.

From (2.23) and the orthogonality relation (1.3) of the Wigner coefficients, we obtain

$$\left\langle \begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix} \right\rangle \left\langle \begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix} \right\rangle = \sum_{2c}(-1)^{t_1} R^{2a2b2c}_{\rho\,\sigma\,\tau} \times \begin{pmatrix} 2a & 2b & 2c \\ \alpha & \beta & \gamma \end{pmatrix} \left\langle \begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix} \right\rangle, \tag{2.24}$$

where

$$t_1 = 2(a+b-c)(2c-\gamma) + (2a-\alpha)(2b-\beta) + \lambda(\alpha)\lambda(\beta).$$

Applying the orthogonality relations of the Wigner operators and that of the Racah invariants, from (2.23), we get once more

$$\delta_{\tau\tau'} R^{2a2b2c}_{\rho\,\sigma\,\tau} = \sum_{\alpha,\beta,\gamma}(-1)^{t_2} \begin{pmatrix} 2a & 2b & 2c \\ \alpha & \beta & \gamma \end{pmatrix} \left\langle \begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix} \right\rangle \times \left\langle \begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix} \right\rangle \left\langle \begin{matrix} & 2c+\tau' & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix} \right\rangle^{\dagger}, \tag{2.25}$$

$$\delta_{cc'}\left\langle\begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix}\right\rangle$$
$$=\sum_{\substack{\alpha,\beta\\ \rho,\sigma}}(-1)^{t_3}R^{2a2b2c'}_{\rho\ \sigma\ \tau}\begin{pmatrix}2a & 2b & 2c\\ \alpha & \beta & \gamma\end{pmatrix}$$
$$\times\left\langle\begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix}\right\rangle\left\langle\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\rangle, \quad (2.26)$$

where

$$t_2=\lambda(\alpha)\lambda(\beta)+n(c,\gamma/2,0)+n(c,\tau/2,0),$$
$$t_3=2(a+b-c)(2c-\tau)$$
$$+(2a-\rho)(2b-\sigma)+\lambda(\alpha)\lambda(\beta).$$

A simple and clear method for verifying the correctness of the relations (2.23)–(2.26) is for the two sides of them to act on the basis $|2j,m\rangle$. For example, from (2.24), we have

$$\begin{pmatrix}2j & 2a & 2j+\rho\\ m & \alpha & m+\alpha\end{pmatrix}\begin{pmatrix}2j+\rho & 2b & 2j+\tau\\ m+\alpha & \beta & m+\gamma\end{pmatrix}$$
$$=\sum_c R\left(jaj+\frac{\tau}{2},b;j+\frac{\rho}{2},c\right)\begin{pmatrix}2a & 2b & 2c\\ \alpha & \beta & \gamma\end{pmatrix}$$
$$\times\begin{pmatrix}2j & 2c & 2j+\tau\\ m & \gamma & m+\gamma\end{pmatrix}. \quad (2.27)$$

This relation is in accord with the relation (4.5) in Ref. 1.

III. OSP(1,2) RACAH OPERATOR

A. Definition for the Racah operator

The Racah operator is associated with the Racah coefficient. The OSP(1,2) Racah operator is denoted by

$$\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}, \quad (3.1)$$

where $\rho,\sigma=2a,2a-1,\ldots,-2a$; its definition is

$$\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}|(2j_1,2j_2)2j,m\rangle$$
$$=(-1)^f R\left(j_1 a j j_2-\frac{\sigma}{2};j_1+\frac{\rho}{2},j_2\right)$$
$$\times|(2j_1+\rho,2j_2-\sigma)2j,m\rangle, \quad (3.2)$$

where

$$f=(2a-\rho)[2(j_1+j_2+j)+\sigma+\rho]+(2a-\sigma)$$
$$\times 2(j_1+j_2+j)+(2a-\sigma)\lambda(2j_2)$$
$$+\lambda(2j_2-\sigma)\lambda(2a)+\sigma.$$

The OSP(1,2) Racah operator may also be defined in terms of the OSP(1,2) Wigner operators, that is,

$$\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}$$
$$=\sum_\alpha(-1)^{n(a,\alpha/2,0)}$$
$$\times\left\langle\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\rangle_1\otimes\left\langle\begin{matrix} & 2a+\sigma & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\rangle_2^\dagger, \quad (3.3)$$

where $\langle 4a,0\rangle_i$ $(i=1,2)$ is the Wigner operator only acting on the space labeled by i. The relation (3.3) coincides with (3.2); this assertion may be proved by having the two sides of (3.3) act on the basis $|(2j_1 2j_2)2j,m\rangle$ of the coupling space. It should be noted that the action of the direct product of two Wigner operators on the direct product of two spaces follows the rule

$$(\langle 4a,0\rangle_1\otimes\langle 4a,0\rangle_2)|2j_1,m_1\rangle|2j_2,m_2\rangle$$
$$=(-1)^{\lambda(\alpha)\lambda(m_1)}\langle 4a,0\rangle_1|2j_1,m_1\rangle\langle 4a,0\rangle_2|2j_2,m\rangle.$$

The Racah operator of $a=\frac{1}{2}$ is the fundamental Racah operator; it is composed of nine operators

$$\left\{\begin{matrix} & 2 & \\ 2 & & 0 \\ & 2 & \end{matrix}\right\},\ \left\{\begin{matrix} & 2 & \\ 2 & & 0 \\ & 1 & \end{matrix}\right\},\ \left\{\begin{matrix} & 2 & \\ 2 & & 0 \\ & 0 & \end{matrix}\right\},$$
$$\left\{\begin{matrix} & 1 & \\ 2 & & 0 \\ & 2 & \end{matrix}\right\},\ \left\{\begin{matrix} & 1 & \\ 2 & & 0 \\ & 1 & \end{matrix}\right\},\ \left\{\begin{matrix} & 1 & \\ 2 & & 0 \\ & 0 & \end{matrix}\right\}, \quad (3.4)$$
$$\left\{\begin{matrix} & 0 & \\ 2 & & 0 \\ & 2 & \end{matrix}\right\},\ \left\{\begin{matrix} & 0 & \\ 2 & & 0 \\ & 1 & \end{matrix}\right\},\ \left\{\begin{matrix} & 0 & \\ 2 & & 0 \\ & 0 & \end{matrix}\right\}.$$

We could use a similar method for studying the relations between the OSP(1,2) Racah coefficients and SO(3) Racah coefficients in Ref. 1 so as to discuss the connections between the OSP(1,2) Racah operators and SO(3) Racah operators, but we ignore this procedure because of its complication.

B. Adjoint operation of the Racah operator

In Sec. II, we have defined the adjoint operation of the Wigner operator. Referring to the adjoint operation of the Racah operator, we use the definition

$$\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}^\dagger|(2j_1,2j_2)2j,m\rangle$$
$$=(-1)^{f'}R\left(j_1-\frac{\rho}{2},ajj_2;j_1 j_2+\frac{\sigma}{2}\right)$$
$$\times|(2j_1-\rho,2j_2+\sigma)2j,m\rangle \quad (3.5)$$

where

$$f'=n(a,\rho/2,0)+n(a,\sigma/2,0)+(2a-\sigma)\lambda(2j_2+\sigma)$$
$$+\lambda(2j_2)\lambda(2a)+\sigma.$$

The adjoint operator $\{\ \}^\dagger$ may be expressed in terms of the Wigner operators, too, that is

$$\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}^{\dagger}$$

$$= \sum_{\alpha} (-1)^{n(a,\alpha/2,0)}$$

$$\times \left\langle\begin{matrix} & 2a+\sigma & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\rangle_2 \otimes \left\langle\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\rangle_1^{\dagger} . \tag{3.6}$$

The reader may verify by himself the concordance of the relation (3.6) with (3.5).

Using the adjoint properties (2.9) and (2.10) of the Wigner operator, we easily prove that the adjoint operation of the Racah operator has the following properties:

$$\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}^{\dagger}$$

$$= (-1)^{\lambda(2a)(\rho+\sigma+1)+2a+\rho+\sigma} \left\{\begin{matrix} & 2a-\rho & \\ 2a & & 0 \\ & 2a-\sigma & \end{matrix}\right\}, \tag{3.7}$$

$$\left(\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}^{\dagger}\right)^{\dagger} = \left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}. \tag{3.8}$$

C. Orthogonality relations of the Racah operators

Using the adjoint operation in the above, we can give some simple and beautiful orthogonality relations for the Racah operators:

$$\sum_{\rho} (-1)^{n(a,\rho/2,0)} \left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}^{\dagger} \left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma' & \end{matrix}\right\}$$

$$= (-1)^{n(a,\sigma/2,0)} I \otimes I \delta_{\sigma\sigma'}, \tag{3.9}$$

$$\sum_{\sigma} (-1)^{n(a,\sigma/2,0)} \left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\} \left\{\begin{matrix} & 2a+\rho' & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}^{\dagger}$$

$$= (-1)^{n(a,\rho/2,0)} I \otimes I \delta_{\rho\rho'}, \tag{3.10}$$

$$\sum_{2j=|2j_1-2j_2|}^{2j_1+2j_2} (-1)^{2(j_1+j_2-j)} \epsilon |(2j_1,2j_2)2j,m\rangle\langle(2j_1,2j_2)2j,m|$$

$$\times \left\{\begin{matrix} & 2a'+\rho' & \\ 4a' & & 0 \\ & 2a'+\sigma' & \end{matrix}\right\} \left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\sigma & \end{matrix}\right\}^{\dagger}$$

$$\times |(2j_1,2j_2)2j,m\rangle$$

$$= (-1)^{n(a,\rho/2,0)+n(a,\sigma/2,0)+2a+\sigma} \delta_{\rho\rho'} \delta_{\sigma\sigma'} \delta_{aa'}, \tag{3.11}$$

where $I \otimes I$ is a direct product of two unit operators. Obviously, the relations (3.9)–(3.11) are very similar to the orthogonality relations (2.11)–(2.13) of the Wigner operators.

D. Coupling law of the Racah operators

We directly point out that the OSP(1,2) Racah operators have the following coupling law:

$$\left\{\begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix}\right\} \left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\}$$

$$= \sum_{c} (-1)^{h_1} \overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau} \overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma} \left\{\begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix}\right\}, \tag{3.12}$$

where

$$h_1 = [\lambda(2a) + 2a - \alpha][\lambda(2b) + 2b - \beta] + (2a-\alpha)(2b-\beta) + 2(a+b-c)(2c-\gamma),$$

and $\overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau}$ and $\overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma}$ are two Racah invariant operators defined as

$$\overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau} |(2j_1,2j_2)2j,m\rangle$$

$$= R^{2a2b2c}_{\rho\,\sigma\,\tau}(2j_1) |(2j_1,2j_2)2j,m\rangle, \tag{3.13a}$$

$$\overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma} |(2j_1,2j_2)2j,m\rangle$$

$$= R^{2a\ \ 2b\ \ 2c}_{-\alpha\,-\beta\,-\gamma}(2j_2) |(2j_1,2j_2)2j,m\rangle. \tag{3.13b}$$

They satisfy the same orthogonality relations as (2.18) and (2.19), that is,

$$\sum_{\rho,\sigma} (-1)^{(2a-\rho)(2b-\sigma)} \overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau} \overset{+}{R}{}^{2a2b2c'}_{\rho\,\sigma\,\tau}$$

$$= (-1)^{2(a+b-c)(2c-\tau)} I \delta_{cc'}, \tag{3.14}$$

$$\sum_{c} (-1)^{2(a+b-c)(2c-\tau)} \overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau} \overset{+}{R}{}^{2a2b\,2c}_{\rho'\,\sigma'\,\tau}$$

$$= (-1)^{(2a-\rho)(2b-\sigma)} I \delta_{\rho\rho'} \delta_{\sigma\sigma'} \tag{3.15}$$

and

$$\sum_{\alpha,\beta} (-1)^{(2a-\alpha)(2b-\beta)} \overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma} \overset{-}{R}{}^{2a2b2c'}_{\alpha\,\beta\,\gamma}$$

$$= (-1)^{2(a+b-c)(2c-\gamma)} I \delta_{cc'}, \tag{3.16}$$

$$\sum_{c} (-1)^{2(a+b-c)(2c-\gamma)} \overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma} \overset{-}{R}{}^{2a2b\,2c}_{\alpha'\,\beta'\,\gamma}$$

$$= (-1)^{(2a-\alpha)(2b-\beta)} I \delta_{\alpha\alpha'} \delta_{\beta\beta'}. \tag{3.17}$$

Using the two sides of (3.12) to act on the basis $|(2j_1,2j_2)2j,m\rangle$, we can prove that the coupling law of the OSP(1,2) Racah operators coincides with the following sum rule of the OSP(1,2) Racah coefficients:

$$\sum_{e} R(abcd;ef) R(db'cd';c'e) R(ba'ed';b'a)$$

$$= R(a'bc'd;b'f) R(fa'cd';c'a). \tag{3.18}$$

The proof of (3.18) is given in Appendix B.

From the orthogonalities of $\overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau}$ and $\overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma}$ and the coupling law (3.12), we easily obtain

$$\sum_{\rho,\sigma}(-1)^{h_2}\overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau}\left\{\begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix}\right\}\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\}$$

$$=\overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma}\left\{\begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix}\right\}, \tag{3.19}$$

$$\sum_{\alpha,\beta}(-1)^{h_3}\overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma}\left\{\begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix}\right\}\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\}$$

$$=\overset{+}{R}{}^{2a2b2c}_{\rho\,\sigma\,\tau}\left\{\begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix}\right\}, \tag{3.20}$$

$$\left\{\begin{matrix} & 2c+\tau & \\ 4c & & 0 \\ & 2c+\gamma & \end{matrix}\right\}\delta_{cc'}$$

$$=\sum_{\substack{\alpha,\beta\\ \rho,\sigma}}(-1)^{h_4}\overset{+}{R}{}^{2a2b2c'}_{\rho\,\sigma\,\tau}\overset{-}{R}{}^{2a2b2c}_{\alpha\,\beta\,\gamma}$$

$$\times\left\{\begin{matrix} & 2b+\sigma & \\ 4b & & 0 \\ & 2b+\beta & \end{matrix}\right\}\left\{\begin{matrix} & 2a+\rho & \\ 4a & & 0 \\ & 2a+\alpha & \end{matrix}\right\}, \tag{3.21}$$

where

$$\begin{aligned} h_2 &= [\lambda(2a)+2a-\alpha][\lambda(2b)+2b-\beta] \\ &\quad +(2a-\alpha)(2b-\beta)+(2a-\rho)(2b-\sigma) \\ &\quad +2(a+b-c)(\gamma+\tau), \\ h_3 &= [\lambda(2a)+2a-\alpha][\lambda(2b)+2b-\beta], \\ h_4 &= [\lambda(2a)+2a-\alpha][\lambda(2b)+2b-\beta] \\ &\quad +(2a-\rho)(2b-\sigma)+2(a+b-c)(2c-\tau). \end{aligned}$$

IV. CONCLUSION

We have established a comprehensive theory on the OSP(1,2) Wigner operators and Racah operators that is quite similar to the corresponding theory of SO(3). It can be believed that the former is a simple and direct generation of the latter to Lie superalgebras.

APPENDIX A: WIGNER COEFFICIENT AS THE ASYMPTOTIC LIMIT OF THE RACAH COEFFICIENT

For the SO(3) algebra, it has been proved that if

$$W^{I_2I_3I''}_{\alpha\beta\gamma}(I)=[(2I-2\beta+1)(2I''+1)]^{1/2}$$
$$\times W(I-\gamma,I_2II_3;I-\beta,I''), \tag{A1}$$

with $\gamma=\alpha+\beta$, then[3] one has

$$\lim_{I\to\infty}W^{I_2I_3I''}_{\alpha\beta\gamma}(I)=C^{I_2I_3I''}_{\alpha\beta\gamma}. \tag{A2}$$

For the OSP(1,2) algebra, let

$$R^{2J_22J_32J''}_{\rho\;\sigma\;\tau}(2J)$$
$$=(-1)^{(2J_2-\rho)(2J_3-\sigma)+2(J_2+J_3-J'')(2J''-\tau)}$$
$$\times R(J-\tau/2,J_2JJ_3;J-\sigma/2,J''), \tag{A3}$$

with $\tau=\rho+\sigma$, we now prove a similar relation, that is,

$$\lim_{J\to\infty}R^{2J_22J_32J''}_{\rho\;\sigma\;\tau}(2J)=\begin{pmatrix}2J_2 & 2J_3 & 2J''\\ \rho & \sigma & \tau\end{pmatrix}. \tag{A4}$$

First of all, we write the OSP(1,2) Wigner coefficient in the product form of two factors:

$$\begin{pmatrix}2J_1 & 2J_2 & 2J\\ M_1 & M_2 & M\end{pmatrix}=\left(\begin{matrix}J_1 & J_2\\ I_1 & I_2\end{matrix}\middle|\begin{matrix}J\\ I\end{matrix}\right)C^{2I_1\,2I_2\,2I}_{M_1M_2M}. \tag{A5}$$

The first factor is named as the scalar factor; the second one is the SO(3) Wigner coefficient. The values of I (I_i) equal J (J_i) if $2J-M$ ($2J_i-M_i$) is even, and $J-\frac{1}{2}$ ($J_i-\frac{1}{2}$) if $2J-M$ ($2J_i-M_i$) is odd. The values of the scalar factor are

$$\left(\begin{matrix}J_1 & J_2\\ J_1 & J_2\end{matrix}\middle|\begin{matrix}J\\ J\end{matrix}\right)=\left(\begin{matrix}J_1 & J_2\\ J_1-\frac{1}{2} & J_2-\frac{1}{2}\end{matrix}\middle|\begin{matrix}J+\frac{1}{2}\\ J\end{matrix}\right)$$
$$=\left(\frac{J_1+J_2+J+1}{2J+1}\right)^{1/2},$$

$$\left(\begin{matrix}J_1 & J_2\\ J_1-\frac{1}{2} & J_2-\frac{1}{2}\end{matrix}\middle|\begin{matrix}J\\ J\end{matrix}\right)=\left(\begin{matrix}J_1 & J_2\\ J_1 & J_2\end{matrix}\middle|\begin{matrix}J+\frac{1}{2}\\ J\end{matrix}\right)$$
$$=(-1)^{\lambda(2J_1)+1}\left(\frac{J_1+J_2-J}{2J+1}\right)^{1/2},$$

$$\left(\begin{matrix}J_1 & J_2\\ J_1-\frac{1}{2} & J_2\end{matrix}\middle|\begin{matrix}J\\ J-\frac{1}{2}\end{matrix}\right)=\left(\begin{matrix}J_1 & J_2\\ J_1 & J_2-\frac{1}{2}\end{matrix}\middle|\begin{matrix}J-\frac{1}{2}\\ J-\frac{1}{2}\end{matrix}\right)$$
$$=\left(\frac{J_1-J_2+J}{2J}\right)^{1/2},$$

$$\left(\begin{matrix}J_1 & J_2\\ J_1 & J_2-\frac{1}{2}\end{matrix}\middle|\begin{matrix}J\\ J-\frac{1}{2}\end{matrix}\right)=(-1)\left(\begin{matrix}J_1 & J_2\\ J_1-\frac{1}{2} & J_2\end{matrix}\middle|\begin{matrix}J-\frac{1}{2}\\ J-\frac{1}{2}\end{matrix}\right)$$
$$=(-1)^{\lambda(2J_1)}\left(\frac{J-J_1+J_2}{2J}\right)^{1/2}. \tag{A6}$$

Letting $J\to\infty$, and reserving J_2 limitedness (hence $J_1\to\infty$), we easily see that the scalar factor has the following asymptotic properties:

$$\lim_{\substack{J\to\infty\\ J_1\to\infty}}\left(\begin{matrix}J_1 & J_2\\ I_1 & I_2\end{matrix}\middle|\begin{matrix}J\\ J\end{matrix}\right)=\delta_{I_1J_1},$$
$$\lim_{\substack{J\to\infty\\ J_1\to\infty}}\left(\begin{matrix}J_1 & J_2\\ I_1 & I_2\end{matrix}\middle|\begin{matrix}J\\ J-\frac{1}{2}\end{matrix}\right)=\delta_{I_1J_1-\frac{1}{2}}. \tag{A7}$$

Now, we write down such a relation between the OSP(1,2) Racah coefficients and Wigner coefficients:

$$R_{J''J'}\begin{pmatrix}2J_1 & 2J'' & 2J\\ M_1 & M'' & M\end{pmatrix}$$
$$=\sum_{M_2M_3M'}(-1)^{2(J_2+J_3-J'')(2J''-M'')+(2J_2-M_2)(2J_3-M_3)}$$
$$\times\begin{pmatrix}2J_1 & 2J_2 & 2J'\\ M_1 & M_2 & M'\end{pmatrix}\begin{pmatrix}2J' & 2J_3 & 2J\\ M' & M_3 & M\end{pmatrix}$$
$$\times\begin{pmatrix}2J_2 & 2J_3 & 2J''\\ M_2 & M_3 & M''\end{pmatrix}. \tag{A8}$$

Making all the OSP(1,2) Wigner coefficients in (A8) the form (A5) and introducing the SO(3) Racah coefficients, we obtain

$$R_{J''J'} = \begin{pmatrix} J_1 & J'' \\ I_1 & I'' \end{pmatrix}\begin{vmatrix} J \\ I \end{pmatrix}^{-1} \sum_{I_2 I_3 I'} (-1)^{2(J_2+J_3-J'')2(J''-I'')+2(J_2-I_2)2(J_3-I_3)}$$
$$\times \begin{pmatrix} J_1 & J_2 \\ I_1 & I_2 \end{pmatrix}\begin{vmatrix} J' \\ I' \end{pmatrix}\begin{pmatrix} J' & J_3 \\ I' & I_3 \end{pmatrix}\begin{vmatrix} J \\ I \end{pmatrix}\begin{pmatrix} J_2 & J_3 \\ I_2 & I_3 \end{pmatrix}\begin{vmatrix} J'' \\ I'' \end{pmatrix}[(2I'+1)(2I''+1)]^{1/2} W(I_1 I_2 I I_3; I' I''). \tag{A9}$$

Further we replace all the Racah coefficients in the relation (A9) by the new notations (A1) and (A3); then we obtain

$$(-1)^{2(J_1+J_2-J')2(J'+J_3-J)+2(J_2+J_3-J'')2(J_1+J''-J)} R^{2J_2\;\;2J_3\;\;2J''}_{2(J'-J_1)\;2(J-J')\;2(J-J_1)}(2J)$$
$$= \begin{pmatrix} J_1 & J'' \\ I_1 & I'' \end{pmatrix}\begin{vmatrix} J \\ I \end{pmatrix}^{-1} \sum_{I_2 I_3 I'} (-1)^{2(J_2+J_3-J'')2(J''-I'')+2(J_2-I_2)2(J_3-I_3)}$$
$$\times \begin{pmatrix} J_1 & J_2 \\ I_1 & I_2 \end{pmatrix}\begin{vmatrix} J' \\ I' \end{pmatrix}\begin{pmatrix} J' & J_3 \\ I' & I_3 \end{pmatrix}\begin{vmatrix} J \\ I \end{pmatrix}\begin{pmatrix} J_2 & J_3 \\ I_2 & I_3 \end{pmatrix}\begin{vmatrix} J'' \\ I'' \end{pmatrix} W^{I_2\;\;I_3\;\;I''}_{I'-I_1\;I-I'\;I-I_1}(I). \tag{A10}$$

We now let $I = J$ in (A10), and make the limit transition of $J \to \infty$. We require that $\rho = 2(J' - J_1)$, $\sigma = 2(J - J')$, and $\tau = \rho + \sigma = 2(J - J_1)$ are limited. Hence we must let $J_1, J' \to \infty$, while letting $J \to \infty$.

When making the limit transition above ($I = J$, $J \to \infty$), according to the asymptotic behavior of the scalar factor, we must let $I_1 = J_1$ in the relation (A10), and notice that only the term of $I' = J'$ does not equal zero in the sum of (A10). This limit transition brings the relation (A4).

Relations (A2) and (A4) show that the Wigner coefficient of SO(3) and OSP(1,2) may all be obtained as the asymptotic limit of their Racah coefficients.

APPENDIX B: DERIVATION FOR THE SUM RULE (3.18)

In order to derive the sum rule (3.18) of the Racah coefficients, we first study the coupling of the three irreps J_1, J_2, and J_3 and consider the coupling coefficient of the following form:

$$S(J_1 J_2 J J_3; J_{12} J_{13}) = \epsilon(2J,M)_{J_{13}} \langle (2J_1,2J_3)2J_{13},2J_2,2J,M \mid (2J_1,2J_2)2J_{12},2J_3,2J,M \rangle, \tag{B1}$$

where $\epsilon(2J,M)_{J_{13}}$ is the norm of the final state vector generated from the coupling way of $2J_1,2J_3 \to 2J_{13}$; $2J_{13},2J_2 \to 2J$.

Similar to the Racah coefficients,[1] we can express S in terms of the Wigner coefficients as

$$S = \sum_{\substack{M_1 M_2 M_3 \\ M_{12} M_{13}}} (-1)^f \begin{pmatrix} 2J_1 & 2J_2 & 2J_{12} \\ M_1 & M_2 & M_{12} \end{pmatrix}\begin{pmatrix} 2J_{12} & 2J_3 & 2J \\ M_{12} & M_3 & M \end{pmatrix}\begin{pmatrix} 2J_1 & 2J_3 & 2J_{13} \\ M_1 & M_3 & M_{13} \end{pmatrix}\begin{pmatrix} 2J_{13} & 2J_2 & 2J \\ M_{13} & M_2 & M \end{pmatrix}, \tag{B2}$$

where

$$f = 2(J_2 + J_{13} - J)(2J - M) + (2J_2 - M_2)(2J_{13} - M_{13}) + 2(J_1 + J_3 - J_{13})(2J_{13} - M_{13})$$
$$+ (2J_1 - M_1)(2J_3 - M_3) + \lambda(M_2)\lambda(M_3).$$

Applying the symmetries of the Wigner coefficients, we can further prove the following relation between S and R:

$$S(J_1 J_2 J J_3; J_{12} J_{13}) = (-1)^{f'} R(J_2 J_1 J J_3; J_{12} J_{13}), \tag{B3}$$

where

$$f' = n(J_1 J_2 J_{12}) + \lambda(2J_1)\lambda(2J_2) + 2(J_1 + J_2 + J_{12})\lambda(2J_{12}) + n(J_{13} J_2 J) + \lambda(2J_{13})\lambda(2J_2) + 2(J_{13} + J_2 + J)\lambda(2J).$$

Now, we study the coupling of the four irreps J_1, J_2, J_3, and J_4, and consider the coupling coefficient of the following form:

$$\langle (2J_1,2J_2)2J_{12},2J_3,2J_{123},2J_4,2J,M \mid (2J_1,2J_4)2J_{14},(2J_2,2J_3)2J_{23},2J,M \rangle \epsilon(2J,M)_{J_{14}J_{23}}, \tag{B4}$$

where $\epsilon(2J,M)_{J_{14}J_{23}}$ is the norm of the right vector generated from the coupling way of $2J_1,2J_4 \to 2J_{14}$; $2J_2,2J_3 \to 2J_{23}$; $2J_{23},2J_{14} \to 2J$.

Inserting the completeness conditions of the state vectors with an indefinite metric into (B4), we have

$$\langle (2J_1,2J_2)2J_{12},2J_3,2J_{123},2J_4,2J,M \mid 2J_1,(2J_2,2J_3)2J_{23},2J_{123},2J_4,2J,M \rangle \epsilon(2J,M)_{J_{23}J_{123}}$$
$$\times \langle 2J_1,(2J_2,2J_3)2J_{23},2J_{123},2J_4,2J,M \mid (2J_1,2J_4)2J_{14},(2J_2,2J_3)2J_{23},2J,M \rangle \epsilon(2J,M)_{J_{14}J_{23}}$$
$$= \sum_{J_{124}} \langle (2J_1,2J_2)2J_{12},2J_3,2J_{123},2J_4,2J,M \mid (2J_1,2J_2)2J_{12},2J_4,2J_{124},2J_3,2J,M \rangle \epsilon(2J,M)_{J_{12}J_{124}}$$
$$\times \langle (2J_1,2J_2)2J_{12},2J_4,2J_{124},2J_3,2J,M \mid (2J_1,2J_4)2J_{14},2J_2,2J_{124},2J_3,2J,M \rangle \epsilon(2J,M)_{J_{14}J_{124}}$$
$$\times \langle (2J_1,2J_4)2J_{14},2J_2,2J_{124},2J_3,2J,M \mid (2J_1,2J_4)2J_{14},(2J_2,2J_3)2J_{23},2J,M \rangle \epsilon(2J,M)_{J_{14}J_{23}}. \tag{B5}$$

From the definitions for S and R, we can see that relation (B5) means

$$R(J_1J_2J_{123}J_3;J_{12}J_{23})S(J_1J_{23}J,J_4;J_{123}J_{14}) = \sum_{J_{124}} S(J_{12}J_3JJ_4;J_{123}J_{124})S(J_1J_2J_{124}J_4;J_{12}J_{14})R(J_{14}J_2JJ_3;J_{124}J_{23}). \tag{B6}$$

Inserting (B3) into (B6), and substituting the letters $a, b, c, d, e, f, a', b', c'$, and d' for $J_{14}, J_2, J, J_3, J_{124}, J_{23}, J_1, J_{12}, J_{123}$, and J_4, respectively, we obtain the sum rule of the Racah coefficients:

$$\sum_e R(abcd;ef)R(db'cd';c'e)R(ba'ed';b'a) = R(a'bc'd;b'f)R(fa'cd';c'a). \tag{B7}$$

[1]G.-J. Zeng, J. Phys. A: Math. Gen. **20**, 1961 (1987).
[2]G.-J. Zeng, J. Phys. A: Math. Gen. **20**, 5423 (1987).
[3]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics Theory and Applications* (Addison–Wesley, Reading, MA, 1981).

Analysis and solution of a nonlinear second-order differential equation through rescaling and through a dynamical point of view

P. G. L. Leach
Department of Computational and Applied Mathematics, The University of Witwatersrand, 2050 Wits Post Office, Johannesburg, South Africa
M. R. Feix and S. Bouquet[a)]
Physique Mathématique, Modélisation et Simulation, Centre National de la Recherche Scientifique, 3A av de la Recherche Scientifique, 45071 Orléans Cédex 2, France

(Received 3 May 1988; accepted for publication 3 August 1988)

The solutions of the equation $\ddot{y} + y\dot{y} + \beta y^3 = 0$, where β is a free parameter, are investigated. For $\beta = \frac{1}{9}$ the equation is linearizable through an eight-parameter symmetry group and is completely integrable. For $\beta \neq \frac{1}{9}$ only two symmetries subsist, but through a dynamical description the analytical asymptotic solutions and their behavior are given according to the value of β and according to the initial conditions.

I. INTRODUCTION

The differential equation

$$y'' + yy' + y^3/9 = 0 \tag{1}$$

arises in the study of the modified Emden equation[1–3]

$$\ddot{q} + \alpha(t)\dot{q} + \gamma(t)q^m = 0$$

in the case $m = 3$. It is also found in the study of univalued functions defined by second-order differential equations. Equation (1) is a member of the Riccati hierarchy and can be transformed into a linear third-order equation by means of the standard transformation $y(x) = 3u'(x)/u(x)$. It is also a member of the class of equations represented by

$$y'' + 3a(x)yy' + b(x)y' + a^2(x)y^3 + c(x)y^2 + d(x)y + e(x) = 0.$$

This is the general form of a second-order ordinary differential equation linear in the first derivative that can be transformed into a linear second-order equation by means of a point transformation.[4–6] As a consequence Eq. (1) possesses the algebra sl(3,R).

By virtue of its interest in both mathematical and physical contexts, we here make further investigations into Eq. (1). To provide interpretations based on an understanding of physics, we recast the problem as the classical mechanical problem of a particle moving in a one-dimensional potential. Further, we modify (1) so that the Newtonian equation of motion is taken to be

$$\ddot{q} + q\dot{q} + \beta q^3 = 0. \tag{2}$$

We are interested in the behavior of the solution of (2), in particular for varying values of β. One question to be addressed is the following. For $\beta = \frac{1}{9}$, (2) is linearizable, possesses eight symmetries, and is completely integrable. Consequently, we could expect that this remarkable mathematical property corresponds to an important physical one appearing (or disappearing) for this value which consequently would appear as a critical one. By setting the problem in the context of classical mechanics, we shall see that there are other values of β for which something can be deduced about (2) analytically and that the critical value of β is not $\frac{1}{9}$. This is something that was not detected in the symmetry analysis referred to above since, for all values of $\beta \neq \frac{1}{9}$, there exist only two symmetries, i.e., the Lie point symmetry analysis distinguishes between $\beta = \frac{1}{9}$ and $\beta \neq \frac{1}{9}$ only. We note that Ince[7] includes (2) in his detailed discussion of second-order nonlinear equations. However, our treatment is differently based.

In the case $\beta = \frac{1}{9}$, application of the Riccati transformation $q = 3\dot{u}/u$ to (2) gives the third-order equation

$$\dddot{u} = 0$$

with solution

$$u(t) = A_0 + A_1 t + A_2 t^2,$$

whence

$$q(t) = \frac{3(A_1 + 2A_2 t)}{A_0 + A_1 t + A_2 t^2}, \quad q(0) = \frac{3A_1}{A_0}, \quad \dot{q}(0) = \frac{3(2A_0A_2 - A_1^2)}{A_0^2}. \tag{3}$$

From (3) we see that only A_1/A_0 and A_2/A_0 matter and consequently we take $A_0 = 1$.

The asymptotic behavior of $q(t)$ depends not only on the value of A_2 but also on the existence and sign of the roots of $1 + A_1 t + A_2 t^2 = 0$. If there is no positive root for this equation, the asymptotic behavior ($t \to \infty$) of $q(t)$ is

$$q(t) = 6/t, \quad A_2 \neq 0, \qquad q(t) = 3/t, \quad A_2 = 0. \tag{3'}$$

On the other hand, if a positive root exists, then the solution exhibits an explosive character [i.e., $q(t)$ goes to infinity in a finite time]. The problem is to obtain the boundary curves for the initial conditions.

If $A_2 < 0$, the equation $1 + A_1 t + A_2 t^2$ has real roots of opposite sign and consequently one is positive leading to an explosive solution. If $A_2 > 0$, two cases occur. For $A_1 > 0$, either there is no root or the two roots are negative and consequently no explosive solution can take place. For $A_1 < 0$, if

[a)] Present address: C.E.A. Centre d'Etude de Vaujours, DFI/FN, BP No. 7, 77181 Courtry, France.

 0022-2488/88/122563-07$02.50

the roots exist, they are both positive and the explosive solution takes place.

Consequently, the boundary curves are

$$A_2=0, \quad \text{for } A_1>0, \tag{3''}$$
$$A_1^2-4A_2=0, \quad \text{for } A_1<0. \tag{3'''}$$

Taking into account (3), (3″) and (3‴) can be written, respectively, as

$$A_2=0 \Rightarrow \dot{q}_0=-3A_1^2.$$

This relation together with

$$q_0=3A_1 \Rightarrow \dot{q}_0=-\tfrac{1}{3}q_0^2.$$

Taking into account the two last relations of (3),

$$A_1^2-4A_2=0 \Rightarrow \dot{q}_0=-\tfrac{1}{6}q_0^2.$$

Figure 1 gives the sign of A_1, A_2, and $\Delta=A_1^2-4A_2$ and shows that the boundary curve for initial conditions leading to an explosive solution are given by the two half-parabolas. In fact, we are going to show the generalization of this result for $0<\beta<\frac{1}{8}$.

II. SELF-SIMILAR ANALYSIS

For β taking on general values, we may use rescaling[2] to determine the asymptotic behavior of $q(t)$. The transformation

$$(t,q)\rightarrow(\bar{t},\bar{q}):\ t=a^A\bar{t},\ q=a^B\bar{q})$$

is self-similar if

$$B-2A=2B-A=3B.$$

This system has the consistent solution $B=-A$ with A arbitrary. We note that this tranformation corresponds to the second of the Lie point symmetries associated with (2),[4] viz.,

$$G_2=t\frac{\partial}{\partial t}-q\frac{\partial}{\partial q}$$

as can be seen by setting $a=(1+\epsilon)^{-1}$. As (2) is autonomous, the other symmetry is obviously

$$G_1=\frac{\partial}{\partial t}.$$

We will see, later on, that the combination of these two symmetries will give both the asymptotic behavior and the classification of the initial conditions. The elementary invariants of the similarity transformation are[8]

$$\xi=qt, \quad \eta=\dot{q}t^2. \tag{4}$$

Noting that

$$\frac{d\xi}{dt}=\frac{1}{t}(\dot{q}t^2+qt)=\frac{1}{t}(\eta+\xi),$$

we take as new variables the invariants ξ (for position) and $\omega=\eta+\xi$ (for velocity), so that Eq. (2) becomes

$$\frac{d\xi}{d\theta}=\omega, \tag{5a}$$

$$\frac{d\omega}{d\theta}=(3-\xi)\omega+(\xi^2-2\xi-\beta\xi^3), \tag{5b}$$

where $\theta=\log t$ is the new time.

Equations (5) are interpreted as a system of first-order equations of motion in the new phase space-time (ξ,ω,θ) of a particle moving under the influence of a velocity-dependent drag force $(3-\xi)\omega$ and a force derivable from the potential

$$V(\xi)=\tfrac{1}{4}\beta\xi^4+\xi^2-\tfrac{1}{3}\xi^3. \tag{6}$$

We note that the velocity-dependent drag force is damping for $\xi>3$ and accelerating for $\xi<3$.

III. BEHAVIOR OF THE POTENTIAL WITH VARYING β

For the convenience of later reference, we categorize the behavior of the potential as β varies downwards. This is presented in Fig. 2. We first note that $V(\xi)$ is always zero at the origin $\xi=0$.

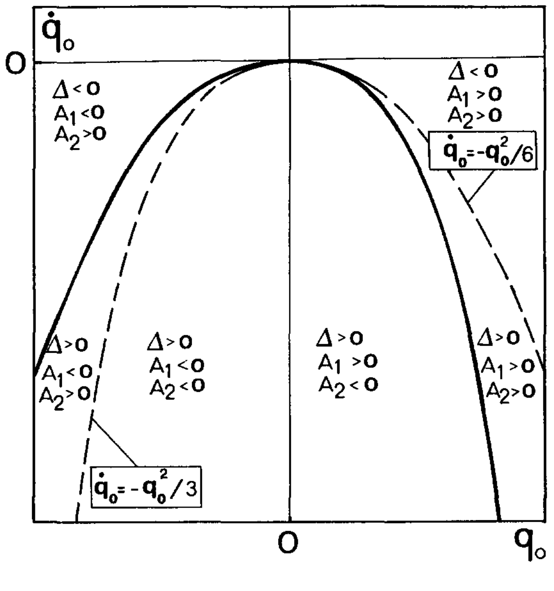


FIG. 1. The boundary curve between explosive and nonexplosive solutions is the half-parabola $\dot{q}_0=-q_0^2/6$ for $q_0<0$ and the half-parabola $\dot{q}_0=-q_0^2/3$ for $q_0>0$ in the case $\beta=\frac{1}{9}$.

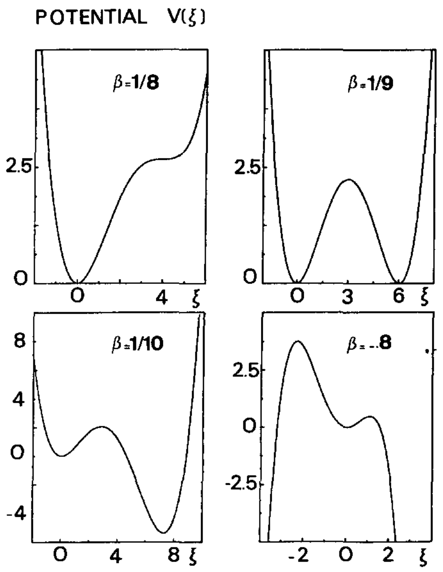


FIG. 2. Plots of the potential $V(\xi)=\frac{1}{4}\beta\xi^4-\frac{1}{3}\xi^3+\xi^2$ for different values of β.

$\boxed{\beta > \frac{1}{8}}$: $V(\xi)$ has a single minimum at $\xi = 0$. It is strictly monotonic increasing for $|\xi| \to \infty$.

$\boxed{\beta = \frac{1}{8}}$: $V(\xi)$ has a single minimum at $\xi = 0$, a stationary point of inflection ($dV/d\xi = d^2V/d\xi^2 = 0$) at $\xi = 4$, and now it is monotonic increasing for $|\xi| \to \infty$.

$\boxed{\frac{1}{8} > \beta > \frac{1}{9}}$: The stationary point of inflection of the potential splits into a maximum at ξ_M in [3,4] and a minimum at ξ_m in [4,6].

$\boxed{\beta = \frac{1}{9}}$: $V(\xi)$ has minima at $\xi = 0$ and $\xi_m = 6$ and a maximum at $\xi_M = 3$. Note that the potential reads $V(\xi) = (\xi - 6)^2\xi^2/36$ in this peculiar case and that it is symmetric about the maximum $\xi_M = 3$. Moreover, the extremum ξ_M is the limit abscissa between a damping and an accelerating velocity-dependent drag force.

$\boxed{0 < \beta < \frac{1}{9}}$: The maximum ξ_M is in [2,3] and the minimum ξ_m is in $[6, +\infty[$. As $\beta \to 0$, ξ_M moves towards the limit value 2 and the minimum ξ_m behaves as $1/\beta$. The potential is strictly monotonic increasing for $|\xi| \to \infty$.

$\boxed{\beta = 0}$: The maximum is at $\xi_M = 2$. The potential is strictly montonic increasing in [0,2] and strictly monotonic decreasing in $]-\infty, 0]$ and in $[2, +\infty[$.

$\boxed{\beta < 0}$: $V(\xi)$ has a minimum at $\xi = 0$ and two maxima on either side of the origin. The maximum on the left of the origin moves in from $\xi = -\infty$ for β increasingly negative and away from $\xi = 2$ to the right. From this rather detailed analysis, it appears that the critical values of β are $\beta = 0$ and $\beta = \frac{1}{8}$.

IV. SOME SPECIAL SOLUTIONS

Before we discuss the qualitative behavior of the motion for varying β, we consider some special solutions. Elimination of θ from Eqs. (5a) and (5b) gives

$$\omega \frac{d\omega}{d\xi} = (3 - \xi)\omega + \xi^2 - 2\xi - \beta\xi^3. \tag{7}$$

An ansatz of the form

$$\omega = a_0 + a_1\xi + a_2\xi^2 \tag{8}$$

for the solution of (7) yields the following possible solutions:

$$\text{(i)}\ \omega = 2\xi - \tfrac{1}{3}\xi^2, \quad \beta = \tfrac{1}{9}, \tag{9}$$

$$\text{(ii)}\ \omega = -6 + 3\xi - \tfrac{1}{3}\xi^2, \quad \beta = \tfrac{1}{9}, \tag{10}$$

$$\text{(iii)}\ \omega = \xi + a_2\xi^2, \quad \beta = -a_2(2a_2 + 1). \tag{11}$$

[It is a trivial matter to show that a finite polynomial solution to (7) can only have the form of (8).]

We may then solve (5a) for each of these solutions. We also list here the corresponding solution of the original equation (2).

Case (i),

$$\xi = 6e^{2\theta}/(K + e^{2\theta}), \quad q = 6t/(K + t^2); \tag{12}$$

case (ii),

$$\xi = \frac{3K + 6e^\theta}{K + e^\theta}, \quad q = \frac{3K + 6t}{Kt + t^2} = \frac{6(\frac{1}{2}K + t)}{-K^2/4 + (\frac{1}{2}K + t)^2}; \tag{13}$$

case (iii),

$$\xi = -e^\theta/(K + a_2e^\theta), \quad q = -1/(K + a_2t); \tag{14}$$

where in all cases K is a constant of integration. We see that solutions (i) and (ii) are almost identical. As expected, they are particular forms of the general solution (3) for the $\beta = \frac{1}{9}$ case.

We consider case (iii) and the condition $\beta = -a_2(2a_2 + 1)$ in more detail. The expression for the potential (6) is now

$$V(\xi) = -(a_2/4)(2a_2 + 1)\xi^4 + \xi^2 - \tfrac{1}{3}\xi^3. \tag{15}$$

If we are interested in real valued solutions to (2), the interpretation in terms of a_2 is valid for $\beta \leqslant \frac{1}{8}$ since for $\beta > \frac{1}{8}$, a_2 as given by (11) becomes complex. Hence the special solution (iii) is only to be considered for $\beta \leqslant \frac{1}{8}$. We recall that it is this value of β that separates two distinct regions of behavior of the potential. From Fig. 2 and Sec. III we see that the potential $V(\xi)$ has three extrema, $\xi_0 = 0$, $\xi_1 = \xi_m$, and $\xi_2 = \xi_M$, for $0 < \beta \leqslant \frac{1}{8}$. These extrema are obtained from Eq. (6) and they are given by

$$(\beta\xi_i^2 - \xi_i + 2)\xi_i = 0 \quad (i = 0,1,2). \tag{16}$$

In fact, Eq. (11) is transformed into Eq. (16) if we take $-1/a_2$ as the new variable. Consequently, introducing the solution ξ_m and ξ_M of Eq. (16) for a given β, we can write the special solution (iii) [Eq. (11)] as

$$\omega_M(\xi) = \xi(1 - \xi/\xi_M), \tag{17a}$$

$$\omega_m(\xi) = \xi(1 - \xi/\xi_m). \tag{17b}$$

Equations (17a) and (17b) show clearly the important role that will be played by the special solution (iii). The curves $\omega_M(\xi)$ and $\omega_m(\xi)$, interpreted as initial conditions in the phase space (ξ,ω), are frontier curves for different types of time evolution. Indeed, the solution $\omega_M(\xi)$ describes a particle arriving with a zero velocity on the top of the potential hill (with possible subsequent bifurcations) while, for the solution $\omega_m(\xi)$, the particle falls in the potential bottom and has a zero velocity at $\xi = \xi_m$ (consequently playing the role of an attractor solution). These two curves are exactly the ones we found in the case $\beta = \frac{1}{9}$ (cf. the discussions of the end of Sec. I where we found that the boundary curve for bifurcating initial conditions is indeed the self-similar solution going through the point $\xi = \xi_M = 3$).

It is interesting to understand why such simple special solutions can be obtained and to see the roles of the two symmetries G_1 and G_2 (see Sec. II) which for our equation exists for all values of β.

In fact, the introduction of the new phase space (ξ,ω) makes the system autonomous (i.e., invariant under the symmetry $\partial/\partial\theta$). To compute the boundaries, we must consequently solve Eq. (7). In our case, the solutions of this equation can be obtained by using the symmetry G_1 [i.e., the invariance of the equation in $(q,\dot{q})$ space]. Let us consider at the initial time $t = 1$ the conditions $q = \xi_m$, $\dot{q} = -\xi_m$, which correspond in the (ξ,ω) space to $\xi = \xi_m$, $\omega = 0$. In this (ξ,ω) space, nothing happens and the particle is motionless. Of course, in the $(q,\dot{q})$ space we have an evolution with

$$q = \xi_m/t, \quad \dot{q} = -\xi_m/t^2.$$

Let us consider the system at $t = T$. Since the system is invariant under time translation, we can reintroduce the values ξ_m/T and $-\xi_m/T^2$ as new initial position and velocity and bring back the clock to $t = 1$. The new ξ and ω are consequently

$$\xi = \xi_m/T, \quad \omega = -\xi_m/T^2 + \xi_m/T.$$

Eliminating T, we indeed obtain

$$\omega = \xi(1 - \xi/\xi_m) \tag{18}$$

and we have obtained the parabola going through the equilibrium point $(\xi_m,0)$. In the same way, we obtain the parabola going through the other equilibrium point $(\xi_M,0)$. These boundary curves appear consequently as prolongation of the two equilibrium points obtained by going into the $(q,\dot{q})$ space (where evolution takes place) and coming back to the (ξ,ω) space.

V. PAINLEVÉ ANALYSIS OF EQ. (2)

We perform the Painlevé analysis of (2) as follows. First, we determine the dominant behavior by substituting

$$q(t) \propto \alpha_0(t - t_1)^n \tag{19}$$

into (2). We find that $n = -1$ and all terms are dominant. This reflects the fact that (2) is invariant under the scale change $t \to at$, $q \to a^{-1}q$ as we saw already in the self-similar analysis of Sec. II. The value of α_0 is determined from the solution of

$$2\alpha_0 - \alpha_0^2 + \beta\alpha_0^3 = 0, \tag{20}$$

i.e.,

$$\alpha_0 = (1 \pm \sqrt{1 - 8\beta}/2\beta \Leftrightarrow \beta = (\alpha_0 - 2)/\alpha_0^2 \tag{21}$$

since $\alpha_0 \neq 0$. We have two particular solutions corresponding to the two roots. They are $\alpha_0^+(t - t_1)^{-1}$ and $\alpha_0^-(t - t_1)^{-1}$. These solutions fail to be real for $\beta > \frac{1}{8}$. In the second step of the Painlevé analysis we determine the resonances (Kowaleski exponents). Writing

$$q = \alpha_0\tau^{-1} + \alpha_1\tau^{-1+r}, \quad \tau = t - t_1, \tag{22}$$

and substituting into (2) we find that

$$r^2 + (\alpha_0 - 3)r + (\alpha_0 - 4) = 0, \tag{23}$$

so that r takes the values $r_1 = -1$, which always occurs in such analyses, and $r_2 = 4 - \alpha_0$. Since the resonances are required to be integral and the leading behavior is determined by $(t - t_1)^{-1}$, we have that $4 - \alpha_0$ must be a non-negative integer, i.e., $\alpha_0 \in \{3,2,1,0,-1,...\}$. Consequently this specifies the value of β. For α_0^+ we have the sequence of permissible (α_0^+,β) to be $\{-n, -(n+2)/n^2, n \in \mathbb{Z}^+\}$ and, for α_0^-, $\{3,\frac{1}{9}; 2,0; 1,-1\}$, where in the case of (2,0), l'Hôpital's rule must be used.

By way of example, the solution for $\alpha_0^- = 3$ and $\beta = \frac{1}{9}$ is

$$q(t) = 3(t - t_1)^{-1} + a_1\left[1 - \frac{a_1}{3}(t - t_1) + \frac{a_1^2}{3^2}(t - t_1)^2 - \frac{a_1^3}{3^3}(t - t_1)^3 + \frac{a_1^4}{3^4}(t - t_1)^4 - \frac{a_1^5}{3^5}(t - t_1)^5 \cdots\right].$$

Performing the summation, we obtain

$$q(t) = 3(t - t_1)^{-1} + 3a_1/[3 + a_1(t - t_1)], \tag{24}$$

where a_1 is arbitrary. We recover a particular case of the general solution given by (3).

For $\alpha_0^- = 1$ and $\beta = -1$, it is

$$q(t) = (t - t_1)^{-1} + a_3(t - t_1)^2\left[1 + \frac{a_3}{21}(t - t_1)^3 + \frac{a_3^2}{3.21}(t - t_1)^6 + \frac{11a_3^3}{39.21.21}(t - t_1)^9 + \frac{4a_3^4}{39.21.21}(t - t_1)^{12}\cdots\right], \tag{25}$$

where a_3 is arbitrary.

VI. QUALITATIVE BEHAVIOR

Now we present qualitative behavior obtained through numerical simulations of the system (5). The discussion is made in the phase space (ξ,ω) according to the values of initial conditions ξ_i and ω_i (ξ_i: initial position; ω_i: initial velocity), and according to the range in which the parameter β lies.

A. The case $0 < \beta \leqslant \frac{1}{8}$

We first study the range bounded by the critical values $\beta = 0$ and $\beta = \frac{1}{8}$. In this interval, we have the two special solutions $\omega_M(\xi) = \xi(1 - \xi/\xi_M)$ [Eq. (17a)] and $\omega_m(\xi) = \xi(1 - \xi/\xi_m)$ [Eq. (17b)], where $\xi_M = (1 - \sqrt{1 - 8\beta})/2\beta$ and $\xi_m = (1 + \sqrt{1 - 8\beta})/2\beta$.

As shown in Fig. 3, these solutions are two parabolic trajectories in the phase space. These curves pass through the point (0;0) where they have a common tangent line, $(d\omega/d\xi) = 1$.

The "large" parabola corresponds to $\omega = \omega_m(\xi)$ and the intersection with the ξ axis is at $\xi = \xi_m$. The small parabola represents $\omega = \omega_M(\xi)$ and $\xi = \xi_M$ is the intersection with the ξ axis.

In order to structure the discussion, we divide the phase space into four strips defined by

$$S_1 = \{\xi/\xi \geqslant \xi_m\},$$
$$S_2 = \{\xi/\xi_M \leqslant \xi \leqslant \xi_m\},$$
$$S_3 = \{\xi/0 \leqslant \xi \leqslant \xi_M\},$$
$$S_4 = \{\xi/\xi \leqslant 0\}.$$

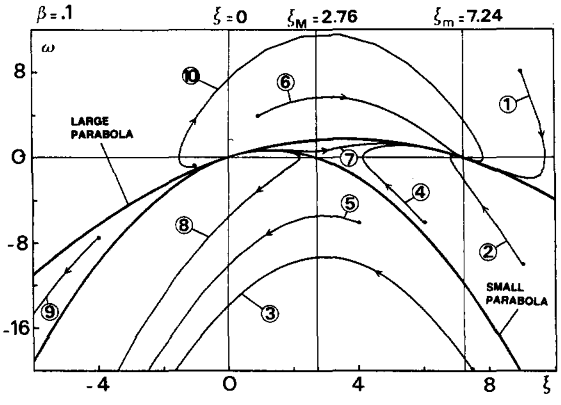


FIG. 3. Evolution in the phase space (ξ,ω) for $\beta = \frac{1}{10}$ $(0 < \beta \leqslant \frac{1}{8})$. The "large" and "small" parabola represent, respectively, $\omega = \omega_m(\xi)$ and $\omega = \omega_M(\xi)$. The curves ①, ②, ④, ⑥, ⑦, ⑩ lead to the *attractor point* $\xi = \xi_m$ and $\omega = 0$. The curves ③, ⑤, ⑧, ⑨ describe explosive solutions.

We now describe the behavior for different initial conditions (ξ_i, ω_i). We denote the asymptotic solution for the time-evolution of the particle by $(\xi_\infty, \omega_\infty)$.

1. $\xi_i \in S_1$

We have the following results:

(i) $\omega_i > \omega_M(\xi)$, then $(\xi_\infty, \omega_\infty) = (\xi_m, 0)$, (26)

(ii) $\omega_i \leqslant \omega_M(\xi)$, then $\omega_\infty(\xi) \sim -\xi^2/\xi_M$, for $\xi \to -\infty$. (27)

In case (i), the particle always asymptotically falls into the bottom of the potential with a zero velocity. Two features must be pointed out. First, if the particle moves initially towards the positive ξ, it comes back after a finite time and during the way back, the work done by the friction force $(3-\xi)\omega$ exactly balances the potential energy in such a way that the particle reaches ξ_m with a zero velocity (curve ① in Fig. 3). As a matter of fact, also for negative initial velocity [but greater than $\omega_M(\xi_i)$], the particle cannot pass over the top $(\xi = \xi_M)$ of the potential and it is finally trapped at $\xi = \xi_m$ (curve ②). For the second case [case (ii)], the initial velocity is highly negative. The particle climbs up the top of the potential and for $\xi < 0$, the ω-dependent force exceeds the potential force. As a consequence, an explosive solution arises and the particle moves outwards $(\xi \to -\infty)$ with an increasing negative velocity (curve ③). The behavior is given by Eq. (27) which is obtained by seeking a solution of the form $A\xi^n$ in Eq. (7). We may now solve q as a function of t. Combining on one hand (4) and (26), we obtain the evolution corresponding to case (i),

$$q(t) \to \frac{1+\sqrt{1-8\beta}}{2\beta} \cdot \frac{1}{t} = \frac{\xi_m}{t}, \quad t \to \infty. \tag{28}$$

This is an extension of solution (12) to $\beta \neq \frac{1}{9}$ for large t and $K > 0$. On the other hand, combining (4) and (27) we obtain [case (ii)]

$$q(t) \sim \frac{1}{t_0 \log(t/t_0)} \sim \frac{1}{t - t_0}, \quad t \to t_0. \tag{29}$$

The particle moves to infinity in a finite time and we recover the results of the Painlevé analysis as given in Sec. V. This solution is again an extension of (12) to $\beta \neq \frac{1}{9}$, where the constant K is negative.

2. $\xi_i \in S_2$

The behavior is similar to the one obtained for the case $\xi_i \in S_1$. The curve $\omega = \omega_M(\xi)$ separates the initial conditions in two classes. The first class $[\omega > \omega_M(\xi)]$ leads to evolution (28) (curve ④), corresponding to the attractor, while the second class $[\omega \leqslant \omega_M(\xi)]$ leads to solution (29) (curve ⑤) corresponding to the explosive solution). A blowup of Fig. 3 in the region $[0, \xi_m]$ is shown in Fig. 4.

3. $\xi_i \in S_3$

The asymptotic behavior is again the same as the case $\xi_i \in S_1$. Nevertheless (see Fig. 4), the transient stage is quite different. In fact, for $\omega_i > \omega_M(\xi_i)$, the particle climbs up the top of the potential before falling to $\xi = \xi_m$ with a zero velocity (curves ⑥ and ⑦). For $\omega_i \leqslant \omega_M(\xi_i)$, the particle can-

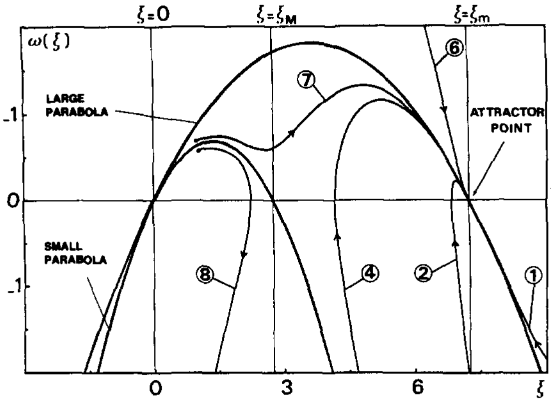


FIG. 4. Blowup of the region $[0, \xi_m]$ shown in Fig. 3.

not jump over the top and, if ω_i is positive, the particle falls back and for large negative ξ, $\omega_\infty(\xi)$ is given by (27) (curve ⑧).

4. $\xi_i \in S_4$

We have the following results:

(i) $\omega_i > \omega_m(\xi_i)$, then $(\xi_\infty, \omega_\infty) = (\xi_m, 0)$; (30)

(ii) $\omega_i \leqslant \omega_m(\xi_i)$, then $\omega_\infty(\xi) \sim -\xi^2/\xi_M$, for $\xi \to -\infty$. (31)

For high negative velocity $[\omega_i < \omega_m(\xi_i)]$ the particle climbs up the potential (the velocity dependent force exceeds the potential force) and experiences an explosive instability (curve ⑨). By contrast for $\omega_i > \omega_m(\xi_i)$ the particle is always trapped at $\xi = \xi_m$ with a zero velocity (curve ⑩).

Let us point out that this result is again a generalization of the result obtained for $\beta = \frac{1}{9}$. In that case, the explosive instability appears at time τ for which $1 + A_1\tau + A_2\tau^2 = 0$. Taking (3) into account we find that, for $t \to \tau$, we have

$$q(t) = \frac{3(A_1 + A_2\tau)}{A_2(\tau - \tau')} \frac{1}{t - \tau}. \tag{32}$$

In the above formula, τ' is the second root of $1 + A_1 t + A_2 t^2 = 0$. Since $\xi = qt$ and $\omega = qt + \dot{q}t^2$, we have, for $t \to \tau$,

$$\xi \approx \frac{3\tau(A_1 + 2A_2\tau)}{A_2(\tau - \tau')} \frac{1}{t - \tau}, \qquad \omega \approx \frac{3\tau^2(A_1 + 2A_2\tau)}{A_2(\tau - \tau')} \frac{(-1)}{(t-\tau)^2}. \tag{33a}$$

For the computation of ω, we can neglect qt when $t \to \tau$. We have consequently, for $t \to \tau$,

$$\frac{\omega}{\xi^2} = \left(-\frac{1}{3}\right) \frac{A_2(\tau - \tau')}{A_1 + 2A_2\tau}. \tag{33b}$$

But since τ and τ' are the two roots of $A_2 t^2 + A_1 t + 1 = 0$, the second member of (33b) is just equal to $(-\frac{1}{3})$, i.e., to $-1/\xi_M$ in agreement with (31).

As a conclusion of this section, we have two kinds of asymptotic behavior. The particle falls to $\xi_\infty = \xi_m$ with $\omega_\infty = 0$ for $\xi_i > 0$ and $\omega_i > \omega_M(\xi_i)$ and for $\xi_i < 0$ and $\omega_i > \omega_m(\xi_i)$.

If $\omega_i \leqslant \omega_M(\xi_i)$ for $\xi_i \geqslant 0$ and if $\omega_i \leqslant \omega_m(\xi_i)$ for $\xi_i \leqslant 0$, the particle moves towards the region of negative ξ with an increasing negative velocity and $\omega(\xi) \sim -\xi^2/\xi_M$ for

$\xi \to -\infty$ (explosive solution). This behavior is described in terms of q and t by Eqs. (28), (29), and (31). Note that (28), (29), and (31) also display the asymptotic behavior for the case $\beta = \frac{1}{9}$. *Consequently, asymptotic solutions obtained analytically for the case $\beta = \frac{1}{9}$ may be generalized to the whole range $0 < \beta \leqslant \frac{1}{8}$.*

B. The case $\beta > \frac{1}{8}$

For $\beta > \frac{1}{8}$, the situation is completely modified. The potential has only a minimum at $\xi = 0$. The two pivot points $\xi = \xi_m$ and $\xi = \xi_M$ and the corresponding "small" and "large" parabolas do not exist any more. Since we have shown that for $0 < \beta < \frac{1}{8}$, these elements give both bifurcation boundaries and the attracting solutions, we can wonder if, for $\beta > \frac{1}{8}$, there is any interest in working in the new space. Consequently, the study in terms of q, $\dot{q}$, and t appears more relevant. In the following, we consider therefore the basic equation (2) and we will present simulations performed in the phase space $(q,\dot{q})$.

Equation (2) holds symmetry properties: it is invariant under the transformation $t \to -t$ and $q \to -q$. The trajectories in the phase space are therefore expected to be symmetric with respect to the $\dot{q}$ axis. Starting initially from $-q_0 < 0$ with a zero velocity ($\dot{q} = 0$), the particle moves towards the region $q > 0$ and reaches the point q_0 with zero velocity. For the same reason, the evolution from this point q_0 with this initial zero velocity brings back the particle at $-q_0$. In other words, the trajectory in the $(q,\dot{q})$ space is a closed curve. The particle oscillates in the range $[-q_0, +q_0]$.

Figure 5 exhibits the evolution for $\beta = 0.15$, $\beta = 0.3$, and $\beta = 0.6$. The initial position is $-q_0 = -1$.

We note that the larger the value of β, the larger is the amplitude of $\dot{q}$. This behavior comes out from the fact that the potential $\beta q^4/4$ stiffens for increasing β. Consequently, the particle is more accelerated and the amplitude of the velocity increases.

The lower part ($\dot{q} < 0$) of the trajectory presents a little bump. This is a slowing down of the particle, when going from $q > 0$ to $q < 0$, due to the velocity-dependent drag force.

As one brings β nearer the critical value $\frac{1}{8}$, the particle passes through $q = 0$ at a slower pace and in the limit $\beta = \frac{1}{8}$, the particle arrives at $q = 0$ with no velocity while the time to reach that point goes to infinity. It is this approach to the

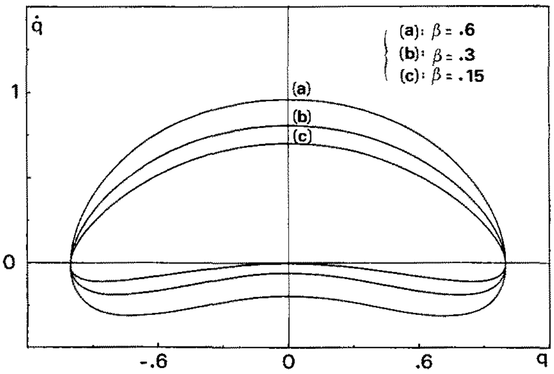


FIG. 5. Solution in the phase space $(q,\dot{q})$ for different values of β ($\beta > \frac{1}{8}$).

point $q = \dot{q} = 0$ that is subsequently described and solved in the new space. For $\beta > \frac{1}{8}$ we, of course, do not need this new space.

C. The case $\beta < 0$

Since we now recover the self-similar solutions, we come back to the (ξ,ω) space. Figure 6 exhibits the two parabolic trajectories in the phase space (ξ,ω). These trajectories correspond to the special solutions (17a) and (17b) for $\beta < 0$. But in contradistinction to Sec. VI A, the curve $\omega_m(\xi)$ is now concave. Note that ξ_m is always negative while ξ_M always lies in the range $[0,2]$.

The new feature brought in this subsection is that no attractor point can exist in the potential $V(\xi)$.

This can be easily seen in terms of the variables q and t. Indeed, coming back to Eq. (2), the force $-\beta q^3$ derives from the potential $V(q) = (\beta/4)q^4$, which is always negative and monotonic decreasing for $|q| \to +\infty$. Consequently, the point $(q, \dot{q}) = (0,0)$ is unstable and the particle always falls to the right or to the left.

The question is now the following: Starting from $q < 0$ (resp. $q > 0$) with a positive velocity $\dot{q}$ (resp. $\dot{q} < 0$) can a particle reach the top ($q = 0$) of the potential and fall towards the zone $q > 0$ (resp. $q < 0$)?

The answer is yes. It is obtained from the study in the (ξ,ω) phase space given in Fig. 6. Keeping in mind that the curves $\omega_m(\xi)$ and $\omega_M(\xi)$ are trajectories of the particle and noting that two distinct trajectories cannot cross, the phase space (ξ,ω) separates in two regions. The frontier between these two regions is given by the curve $\omega = \omega_m(\xi)$ for $\xi < 0$ and by $\omega = \omega_M(\xi)$ for $\xi > 0$. *We see, therefore, that this property, already obtained in the range $0 < \beta < \frac{1}{8}$, is preserved for $\beta < 0$.*

The first region contains every initial point (ξ_i,ω_i) whose evolution brings asymptotically the particle towards $\xi \to +\infty$ with positive velocity. For any initial condition (ξ_i,ω_i) taken in the second region, the particle always moves towards $\xi \to -\infty$ with negative velocity. The initial points labeled ① to ④ (respectively, ⑤ to ⑧) give the evolution towards $\xi \to +\infty$ (resp. $-\infty$) with an increasing positive (respectively, negative) velocity $[\omega(\xi) \sim \xi^2]$.

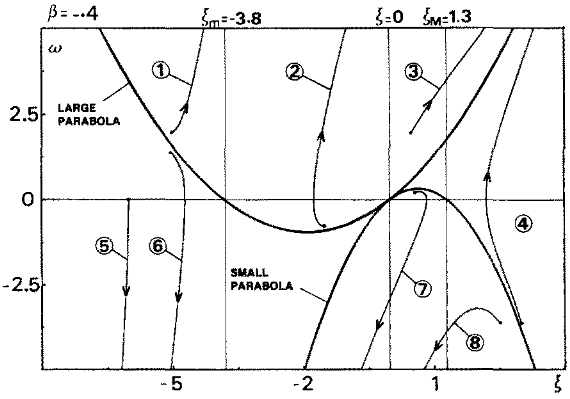


FIG. 6. Evolution in the phase space (ξ,ω) for $\beta = -0.4$ ($\beta < 0$). The curves ① to ④ (resp. ⑤ to ⑧) describe a particle moving towards $\xi > 0$ (resp. $\xi < 0$).

VII. CONCLUSION

The purpose of this paper was, first, to exhibit the importance and role of the different symmetries obtained by the usual Lie analysis, and, second, to show that rescaling and casting the problem in a new phase space allows one to use qualitative arguments about the sign of the drag force, the form of the new potential, and the role of the self-similar solutions.

The first intriguing result is the critical position from a purely "mathematical" point of view of the case $\beta = \frac{1}{9}$ where everything can be analytically expressed while from a "physical" point of view it plays no role at all. This somewhat arbitrary distinction between mathematical and physical points of view just means that the bifurcation values (for the parameter) or the bifurcation boundaries (for the initial conditions) will be labeled physical results while the fact that an analytical expression can be given is a mathematical one.

The second result was the "pivot" role played by the self-similar solutions $q = \xi_m/t$ and ξ_M/t. Moreover, since the equation is also invariant under time translation, these solutions can be extended (curves $\omega = \xi - \xi^2/\xi_m$ and $\omega = \xi - \xi^2/\xi_M$) providing all the interesting results about the boundaries for the initial conditions leading to bifurcations and the nature of the asymptotic solutions. When these self-similar solutions disappear, the nature of the general solution is totally modified.

The third result is the interesting concept of rescaling where, introducing the new dynamical variables, we can find easily the equilibrium points and their nature not only in their neighborhood but, sometimes, far away in a strongly nonlinear region. Of course, in some situations ($\beta > \frac{1}{8}$), the new phase space does not present any interest and the problem is simpler in the original $(q,\dot{q},t)$ space.

Now we point out that these methods and concepts can be generalized to the case of the second-order differential equation with self-similar solutions [either time invariant as (3) or not]. Of course, the results will depend on the structure of the different terms of the equation, but the methodology will essentially remain the same. We will present later work on this generalization.

ACKNOWLEDGMENT

One of us (P. G. L. Leach) thanks the CNRS at Orléans for its hospitality in July 1987 during which time part of this work was undertaken.

[1]J. de C. Moreira, "Comments on 'A direct approach to finding exact invariants for one-dimensional time-dependent classical Hamiltonians,' " preprint IF/UFRJ/83/25 of the Institute of Physics, Federal University of Rio de Janeiro, 1983.
[2]M. R. Feix and H. R. Lewis, J. Math. Phys. **26**, 68 (1985).
[3]P. G. L. Leach, J. Math. Phys. **26**, 2510 (1985).
[4]F. M. Mahomed and P. G. L. Leach, Quaestiones Math. **8**, 241 (1985).
[5]F. M. Mahomed and P. G. L. Leach, "Symmetries of nonlinear differential equations and linearization II," preprint CNLS-CAMS, University of the Witwatersrand, 1987.
[6]L. G. S. Duarte, S. F. S. Duarte, and J. de C. Moreira, "One dimensional equations with the maximum number of symmetry generators," preprint 4RFJ-IF-FIT-87/03 of the Department of Theoretical Physics, Federal University of Rio de Janeiro, 1983.
[7]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. 14.
[8]M. R. Feix, S. Bouquet, and H. R. Lewis, Physica D **28**, 80 (1987).

Gauge and Bäcklund transformations for the variable coefficient higher-order modified Korteweg–de Vries equation

Yu-kun Zheng and W. L. Chan
Department of Mathematics, Science Centre, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

(Received 18 May 1988; accepted for publication 10 August 1988)

A family of higher-order modified Korteweg–de Vries equations with variable coefficients (t-ho-mKdV) is introduced. A one-to-one correspondence between a real solution of these equations and a complex solution of the variable coefficient higher-order Korteweg–de Vries (t-ho-KdV) equations is established through a complex Miura transformation. An auto-Bäcklund transformation for these t-ho-mKdV equations is derived from that of the t-ho-KdV equations. The associated gauge transformations of the corresponding AKNS systems are presented. They enable one to construct a hierarchy of solutions of the t-ho-mKdV equations from a known hierarchy without solving the differential equations for the wave functions except the first one. A new family of higher-order evolution equations with an auto-Bäcklund transformation is also derived in connection with the gauge transformation of the t-ho-mKdV equations.

I. INTRODUCTION

It is well known that the Miura transformation[1] (MT)

$$u = q_x + q^2 \tag{1.1}$$

connects the solution u of the Korteweg–de Vries (KdV) equation

$$u_t + u_{xxx} + 6uu_x = 0, \tag{1.2}$$

and the solution q of the modified KdV (mKdV) equation

$$q_t = 6q^2 q_x + q_{xxx} = 0. \tag{1.3}$$

Chern and Peng[2] generalized these two equations to two higher-order families of equations and also proved the existence of the same MT (1.1) connecting them. However, (1.1) only furnishes a method for obtaining a solution u of the KdV equation (1.2) from a known solution q of the mKdV equation (1.3) but not the other way round. Hence it is of much interest to find an auto-Bäcklund transformation (BT) for (1.2) or (1.3) so that one can construct more new solutions from a known one. Wadati and Sogo[3] pointed out that there also exists a complex form of MT,

$$u = iq_x + q^2 \quad (q \text{ real}), \tag{1.4}$$

between the solutions of Eqs. (1.2) and (1.3). To check this we need only to substitute (1.4) into (1.2). Thus

$$u_t + u_{xxx} + 6uu_x = \left(i\frac{\partial}{\partial x} + 2q\right)(q_t + 6q^2 q_x + q_{xxx}) = 0. \tag{1.5}$$

Equations (1.4) and (1.5) indicate that if a complex solution u of the KdV equation (1.2) possesses the form of (1.4), then the square root of the real part (or the integral of the imaginary part) of u is a solution of the mKdV equation (1.3) and vice versa. On the other hand, the authors of this paper had derived an auto-BT for the t variable coefficient higher-order KdV (t-ho-KdV) family and a gauge transformation (GT) for the corresponding AKNS system.[4,5] In this paper, we will generalize the results obtained there to the case of the t-ho-KdV equation with complex solutions and derive a family of the t variable coefficient higher-order mKdV (t-ho-mKdV) equation from the t-ho-KdV equation. Then we show that the same MT (1.4) also provides a connection between the families of the t-ho-KdV equation and the t-ho-mKdV equation. Finally, we will use the MT (1.4) and the auto-BT for the generalized t-ho-KdV equation to derive an auto-BT for the t-ho-mKdV equation and use the GT for the AKNS system corresponding to the t-ho-KdV equation to derive a GT for the AKNS system corresponding to the t-ho-mKdV equation.

II. THE t-ho-mKdV EQUATION

The starting point of our problem is the following AKNS system:

$$d\Psi = \Omega\Psi, \tag{2.1}$$

where Ψ is a column vector function of x and t,

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \tag{2.2}$$

and

$$\Omega = P\,dx + Q\,dt, \tag{2.3}$$

$$P = \begin{pmatrix} \eta & q \\ -q & -\eta \end{pmatrix}, \tag{2.4}$$

η a real parameter, independent of x and t, (2.5)

q a function of x and t, (2.6)

$$Q = \begin{pmatrix} A & B \\ C & -A \end{pmatrix}, \tag{2.7}$$

A a functional of q, (2.8)

$$B = \frac{A_x}{2q} + \frac{1}{\eta} qA + \frac{1}{4\eta}\left(\frac{A_x}{q}\right)_x, \tag{2.9}$$

$$C = \frac{A_x}{2q} - \frac{1}{\eta} qA - \frac{1}{4\eta}\left(\frac{A_x}{q}\right)_x. \tag{2.10}$$

Theorem 1: A necessary and sufficient condition for the

 0022-2488/88/122570-06$02.50

integrability of the AKNS system (2.1) is that A and q satisfy the following equation:

$$q_t + \eta \frac{A_x}{q} - \frac{1}{\eta}(qA)_x - \frac{1}{4\eta}\left(\frac{A_x}{q}\right)_{xx} = 0. \qquad (2.11)$$

To prove this theorem it is sufficient to verify that the exterior differential equality

$$d\Omega - \Omega \wedge \Omega = 0 \qquad (2.12)$$

holds under the condition (2.11).

Now, consider the following complex gauge transformation:

$$G_1\colon\ \Psi \to \Phi = G_1 \Psi, \qquad (2.13)$$

where

$$G_1 = \begin{pmatrix} iq - 2\eta & -q \\ 1 & i \end{pmatrix}. \qquad (2.14)$$

We have

$$\Phi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \begin{pmatrix} -2\eta\psi_1 - q\psi_2 + iq\psi_1 \\ \psi_1 + i\psi_2 \end{pmatrix}. \qquad (2.15)$$

Differentiating in (2.13) and using (2.1) and (2.11) we get a complex AKNS system with Φ as its wave function:

$$d\Phi = \Theta\Phi, \qquad (2.16)$$

where

$$\Theta = M\,dx + N\,dt, \qquad (2.17)$$

$$M = (G_1)_x G^{-1} + G_1 P G_1^{-1} = \begin{pmatrix} \eta & iq + q^2 \\ -1 & -\eta \end{pmatrix}, \qquad (2.18)$$

$$N = (G_1)_t G_1^{-1} + G_1 Q G_1^{-1} = \frac{i}{2\eta}\begin{pmatrix} \alpha & \sigma \\ \tau & -\alpha \end{pmatrix}, \qquad (2.19)$$

$$\alpha = \eta \frac{A_x}{q} + \frac{1}{2}\left(\frac{A_x}{q}\right)_x - i(2\eta A + A_x), \qquad (2.20)$$

$$\sigma = q_x A - i\left(i\frac{\partial}{\partial x} + 2q\right)\left[qA + \eta\frac{A_x}{q} + \frac{1}{2}\left(\frac{A_x}{q}\right)_x\right], \qquad (2.21)$$

$$\tau = \frac{i}{q}\left(i\frac{\partial}{\partial x} + 2q\right)A. \qquad (2.22)$$

Denote

$$R = iD + 2q, \quad D = \frac{\partial}{\partial x}, \qquad (2.23)$$

$$\hat{C} = -(1/2\eta q)RA. \qquad (2.24)$$

In view of (1.4), (2.23), (2.24), and (2.20)–(2.22), we get

$$\alpha = 2\eta i(\tfrac{1}{2}\hat{C}_x + \eta\hat{C}), \qquad (2.25)$$

$$\sigma = 2\eta i(\tfrac{1}{2}\hat{C}_{xx} + \eta\hat{C}_x + u\hat{C}), \qquad (2.26)$$

$$\tau = -2\eta i\hat{C}. \qquad (2.27)$$

The AKNS system (2.16) now reads

$$d\Phi = \begin{pmatrix} \eta & u \\ -1 & -\eta \end{pmatrix}\Phi\,dx + \begin{pmatrix} -\tfrac{1}{2}\hat{C}_x - \eta\hat{C} & -\tfrac{1}{2}\hat{C}_{xx} - \eta\hat{C}_x - u\hat{C} \\ \hat{C} & \tfrac{1}{2}\hat{C}_x + \eta\hat{C} \end{pmatrix}\Phi\,dt. \qquad (2.28)$$

For this AKNS system we have the following.

Theorem 2: A necessary and sufficient condition for the integrability of the AKNS system (2.28) is that $\hat{C}$ and u satisfy the following equation:

$$u_t + \tfrac{1}{2}\hat{C}_{xxx} - 2(\eta^2 - u)\hat{C}_x + u_x\hat{C} = 0. \qquad (2.29)$$

To prove this theorem it is sufficient to verify that the exterior differential equality

$$d\Theta - \Theta \wedge \Theta = 0 \qquad (2.30)$$

holds under the condition (2.29).

By substituting (1.4) and (2.24) into (2.29), we get the following identity:

$$u_t + \tfrac{1}{2}\hat{C}_{xxx} - 2(\eta^2 - u)\hat{C}_x + u_x\hat{C} = R\left\{q_t + \eta\frac{A_x}{q} - \frac{1}{\eta}(qA)_x - \frac{1}{4\eta}\left(\frac{A_x}{q}\right)_{xx}\right. . \qquad (2.31)$$

This identity establishes a relation between the two equations (2.11) and (2.29). Since (2.31) is a complex identity and the expression on the left-hand side of (2.11) is a multiple of the real part of the right-hand side of (2.31). Hence we arrive at the following.

Theorem 3: Under the condition that $\hat{C}$ and q are connected by (2.24), a necessary and sufficient condition for q and A to satisfy Eq. (2.11) is that u and $\hat{C}$ satisfy Eq. (2.29).

Equation (2.29) contains a family of KdV-type equations. We have discussed it in some detail in our previous paper.[4] We now use (2.29) to derive a family of mKdV equations.

Choosing $\hat{C}$ in (2.29) to be a polynomial of η,

$$\hat{C} = \hat{C}_n = 4\sum_{j=0}^{n}\sum_{m=0}^{j} E^m k_{j-m}(t)\eta^{2(n-j)}, \qquad (2.32)$$

where

$$E = \tfrac{1}{4}D^2 + u - \tfrac{1}{2}D^{-1}u_x, \qquad (2.33)$$

$$D = \frac{\partial}{\partial x}, \quad D^{-1} = \int dx, \qquad (2.34)$$

and $k_j(t)$ ($j = 0,1,2,\ldots,n$) are some arbitrary real functions of t. Denote

$$F = \tfrac{1}{4}D^2 + u + \tfrac{1}{2}u_x D^{-1}, \qquad (2.35)$$

$$T = \tfrac{1}{4}D^3 + uD + \tfrac{1}{2}u_x. \qquad (2.36)$$

Using the commutative relation

$$TE = FT, \qquad (2.37)$$

after the substitution of (2.32) into (2.29), the η in (2.29) will be eliminated and (2.29) leads to

$$u_t + 4\sum_{j=0}^{n} k_{n-j}(t)F^j u_x = 0. \qquad (2.38)$$

For $n = 1$, $k_0 = 2$, and $k_1 = 0$, (2.38) becomes the well-known KdV equation (1.2). Therefore (2.38) is called the t variable coefficient higher-order KdV (t-ho-KdV) equation.

Denote

$$L = \tfrac{1}{4}D^2 + q^2 + q_x D^{-1}q. \qquad (2.39)$$

It is easy to verify that the commutative relation

$$FR = RL \tag{2.40}$$

holds. Substituting (1.4) into (2.38) and using (2.40), we get the identity

$$u_t + 4 \sum_{j=0}^{n} k_{n-j}(t) F^j u_x = R \left[q_t + 4 \sum_{j=0}^{n} k_{n-j}(t) L^j q_x \right]. \tag{2.41}$$

Now consider the following equation:

$$q_t + 4 \sum_{j=0}^{n} k_{n-j}(t) L^j q_x = 0. \tag{2.42}$$

This is a family of evolution equations; for $n = 1$, $k_0 = 1$, and $k_1 = 0$, (2.42) gives the well-known mKdV equation (1.3). Therefore we call (2.42) the t variable coefficient higher-order mKdV (t-ho-mKdV) equation. Thus, by (2.41), we have the following.

Theorem 4: If a real function q and a complex function u with independent variables x and t are connected by (1.4), then a necessary and sufficient condition for the function q to be a solution of the t-ho-mKdV equation (2.42) is that the function u be a solution of the t-ho-KdV equation (2.38).

Furthermore, by Theorem 1, Theorem 3, (2.32), and Theorem 4, we get the following.

Theorem 5: A necessary and sufficient condition for integrability of an AKNS system (2.1) is that the function q satisfies the t-ho-mKdV equation (2.42).

III. SOME RESULTS ABOUT THE t-ho-KdV EQUATION

In the last section, we revealed a relation between the t-ho-KdV equation and the t-ho-mKdV equation, that is, the t-ho-mKdV equation can be derived from a complex solution t-ho-KdV equation. Our next step is to derive an auto-BT for the t-ho-mKdV equation from the auto-BT of the t-ho-KdV equation. Therefore we will first cite here the main results about the t-ho-KdV equation that we obtained previously.[4] Although these results originally were derived for the case where the solution of the t-ho-KdV equation is a real function, they are not subjected to this restriction in the derivation.

Assume that u is a known solution of the t-ho-KdV equation (2.38), and $\Phi = (\varphi_1, \varphi_2)^T$ is a corresponding solution of the AKNS system (2.28). Let

$$v = \varphi_{2x}/\varphi_2 = -\varphi_1/\varphi_2 - \eta; \tag{3.1}$$

then Eq. (2.38) has a new solution

$$u' = u + 2v_x. \tag{3.2}$$

This is an auto-BT for the t-ho-KdV equation (2.38). To this new solution u' of (2.38), by Theorem 2, there is a corresponding new AKNS system

$$d\Phi' = \begin{pmatrix} \eta & u' \\ -1 & -\eta \end{pmatrix} \Phi' \, dx + \begin{pmatrix} -\frac{1}{2}\hat{C}'_x - \eta \hat{C}' & -\frac{1}{2}\hat{C}'_{xx} - \eta \hat{C}'_x - u\hat{C}' \\ \hat{C}' & \frac{1}{2}\hat{C}'_x + \eta \hat{C}' \end{pmatrix} \Phi' \, dt. \tag{3.3}$$

There also exists a gauge transformation

$$G_2\colon \ \Phi \to \Phi' = G_2 \Phi, \tag{3.4}$$

which transforms (2.28) into (3.3), where G_z is a 2×2 matrix as follows:

$$G_2 = \frac{1}{\beta^2} \begin{pmatrix} a + (\eta - v)c & (\eta + v)a - b + (\eta^2 - v^2)c - (\eta - v)d \\ -c & -(\eta + v)c + d \end{pmatrix}. \tag{3.5}$$

The notation in (3.5) has the following meanings in the present case. Variables η and v have been defined in (2.5) and (3.1). Denote

$$\varphi_2^0 = \varphi_2(x_0, t), \tag{3.6}$$

$$\beta = \varphi_2^0 \exp\left(\int_{x_0}^{x} v \, dx \right), \tag{3.7}$$

$$B = \int_{x_0}^{x} \beta^2 \, dx, \tag{3.8}$$

$$B = \int_{x_0}^{x} \beta^{-2} \, dx; \tag{3.9}$$

then

$$a = \beta^4 (a_0 - b_0 B'), \tag{3.10}$$

$$b = b_0 \beta^2, \tag{3.11}$$

$$c = \beta^2 (a_0 B - b_0 B B' + c_0 - d_0 B'), \tag{3.12}$$

$$d = b_0 B + d_0. \tag{3.13}$$

Originally a, b, c, and d were expressed by some indefinite integrals, but now, instead, we have expressed these quantities in (3.6)–(3.13) by some definite integrals for the convenience of determining the a_0, b_0, c_0, and d_0 in (3.10)–(3.13). Consequently these quantities can now be determined as functions of t as follows:

$$a_0 = a_1 + b_1 A, \tag{3.14}$$

$$b_0 = b_1, \tag{3.15}$$

$$c_0 = -a_1 A' - b_1 A A' + c_1 + d_1 A, \tag{3.16}$$

$$d_0 = -b_1 A' + d_1, \tag{3.17}$$

where

$$A = \int_{t_0}^{t} (\varphi_2^0)^{-2} \hat{C}_n [u(x_0, t), t] \, dt, \tag{3.18}$$

$$A' = \int_{t_0}^{t} (\varphi_2^0)^2 \hat{C}_n [u'(x_0, t), t] \, dt, \tag{3.19}$$

and a_1, b_1, c_1, and d_1 are some arbitrary constants satisfying the following condition:

$$a_1 d_1 - b_1 c_1 = 1. \tag{3.20}$$

IV. A BÄCKLUND TRANSFORMATION FOR THE *t*-ho-mKdV EQUATION

We now use the above results to derive a BT for the *t*-ho-mKdV equation (2.42).

Suppose that q is a known solution of the *t*-ho-mKdV equation (2.42); then, by Theorem 4, the function u, determined by (1.4), is a complex solution of the *t*-ho-KdV equation (2.38), and, further, u', defined by (3.2), is a new solution of (2.38). We want to show that the BT (3.2) for the *t*-ho-KdV equation (2.38) is a complex function of the form of a complex MT (1.4). Denote

$$iq'_x + q^* = u' = u + 2v_x, \tag{4.1}$$

where q' and q^* are two real functions. It is necessary to show that the following equality holds:

$$q^* = (q')^2. \tag{4.2}$$

Substituting the complex function φ_2 in (2.15) into (3.1) and then (1.4) and (3.1) into (4.1), we have

$$iq_x + q^* = iq_x + q^2 + 2 \times [\ln(\psi_1^2 + \psi_2^2)^{1/2} + i\tan^{-1}(\psi_2/\psi_1)]_{xx}, \tag{4.3}$$

or, equating the imaginary part and the real part of the two sides of equality (4.3), respectively,

$$q' = q + 2\left(\tan^{-1}\frac{\psi_2}{\psi_1}\right)_x = q + 2\,\frac{\psi_1\psi_{2x} - \psi_2\psi_{1x}}{\psi_1^2 + \psi_2^2}, \tag{4.4}$$

$$q^* = q^2 + 2[\ln(\psi_1^2 + \psi_2^2)^{1/2}]_{xx} = q^2 + 2\left(\frac{\psi_1\psi_{1x} + \psi_2\psi_{2x}}{\psi_1^2 + \psi_2^2}\right)_x. \tag{4.5}$$

By (2.1)–(2.7) we get

$$\psi_{1x} = \eta\psi_1 + q\psi_2, \tag{4.6}$$

$$\psi_{2x} = -q\psi_1 - \eta\psi_2, \tag{4.7}$$

$$\psi_{1t} = A\psi_1 + B\psi_2, \tag{4.8}$$

$$\psi_{2x} = C\psi_1 - A\psi_2. \tag{4.9}$$

Using (4.6) and (4.7), (4.4) and (4.5) can be simplified. We have

$$q' = -q - \frac{4\eta\psi_1\psi_2}{\psi_1^2 + \psi_2^2}, \tag{4.10}$$

$$q^* = q^2 + 2\eta\left(\frac{\psi_1^2 - \psi_2^2}{\psi_1^2 + \psi_2^2}\right)_x = \left(q + \frac{4\eta\psi_1\psi_2}{\psi_1^2 + \psi_2^2}\right)^2. \tag{4.11}$$

Equations (4.10) and (4.11) indicate that equality (4.2) holds. Therefore (3.2) can be rewritten in the form

$$u' = iq'_x + q'^2, \tag{4.12}$$

where q' is the function defined in (4.4). Thus, by the definition of u' in (3.2), the equality (4.12), and Theorem 4, we arrive at the following.

Theorem 6: Assume that q is a solution of the *t*-ho-mKdV equation (2.42) and that ψ_1 and ψ_2 are the corresponding solutions of the AKNS system (2.1); then the function q', defined in (4.4), is a new solution of the *t*-ho-mKdV equation (2.42), that is, (4.4) is a BT for the *t*-ho-mKdV equation (2.42).

For $n = 1$, $k_0 = 1$, and $k_1 = 0$ in (2.42), or the mKdV equation (1.3), this result was obtained by Wadati *et al.*[6] (see, also, Rogers and Shadwick[7]).

V. GAUGE TRANSFORMATION AND THE BÄCKLUND TRANSFORMATION

It is obvious that the application of the BT (4.4) for finding a new solution of the *t*-ho-mKdV equation requires the solutions ψ_1 and ψ_2 of the AKNS system (2.1). In this section, we will introduce an easy method to obtain a new solution of (2.1) from a known solution, that is, the gauge transformation method for the AKNS system (2.1).

To the solution q' in (4.4) of the *t*-ho-mKdV equation (2.42), by Theorem 5, there is a corresponding integrable AKNS system

$$d\Psi' = \Omega'\Psi', \tag{5.1}$$

where Ψ' is a column vector function of x and t,

$$\Psi' = \begin{pmatrix} \psi'_1 \\ \psi'_2 \end{pmatrix}, \tag{5.2}$$

and

$$\Omega' = P'\,dx + Q'\,dt, \tag{5.3}$$

$$P' = \begin{pmatrix} \eta & q' \\ -q' & -\eta \end{pmatrix}, \tag{5.4}$$

$$\eta \text{ kept the same as (2.5)}, \tag{5.5}$$

$$q' \text{ a function defined in (4.4)}, \tag{5.6}$$

$$Q' = \begin{pmatrix} A' & B' \\ C' & -A' \end{pmatrix}, \tag{5.7}$$

$$A' \text{ a functional of } q', \tag{5.8}$$

$$B' = \frac{A'_x}{2q'} + \frac{1}{\eta}q'A + \frac{1}{4\eta}\left(\frac{A'_x}{q'}\right)_x, \tag{5.9}$$

$$C' = \frac{A'_x}{2q'} - \frac{1}{\eta}q'A - \frac{1}{4\eta}\left(\frac{A'_x}{q'}\right)_x. \tag{5.10}$$

Referring to (2.13)–(2.28), there exists a complex gauge transformation

$$G_3\colon\ \Psi' \to \Phi' = G_3\Psi', \tag{5.11}$$

with

$$G_3 = \begin{pmatrix} iq' - 2\eta & -q' \\ 1 & i \end{pmatrix} \tag{5.12}$$

and

$$\Phi' = \begin{pmatrix} \varphi'_1 \\ \varphi'_2 \end{pmatrix} = \begin{pmatrix} -2\eta\psi'_1 - q\psi'_2 + iq\psi'_1 \\ \psi'_1 + i\psi'_2 \end{pmatrix}, \tag{5.13}$$

such that Φ' satisfies the complex AKNS system

$$d\Phi' = \begin{pmatrix} \eta & u' \\ -1 & -\eta \end{pmatrix}\Phi'\,dx + \begin{pmatrix} -\frac{1}{2}\hat{C}'_x - \eta\hat{C}' & -\frac{1}{2}\hat{C}'_{xx} - \eta\hat{C}'_x - u\hat{C}' \\ \hat{C}' & \frac{1}{2}\hat{C}'_x + \eta\hat{C}' \end{pmatrix}\Phi'\,dt, \tag{5.14}$$

where u' is a solution of the *t*-ho-KdV equation (2.38) defined in (3.2) and $\hat{C}'$ is connected to q' and A' as follows:

$$\hat{C}' = \frac{1}{2\eta q'} R'A' \quad \left(R' = i\frac{\partial}{\partial x} + 2q'\right). \tag{5.15}$$

Thus (5.14) is just the AKNS system (3.3), and we know from Sec. III that there exists a gauge transformation (3.4) that transforms (2.28) into (3.3) [namely (5.14)]. Let

$$G = G_3^{-1}G_2G_1, \tag{5.16}$$

where G_1, G_2, and G_3 are defined in (2.14), (3.5), and (5.12), respectively. Thus, by (2.13), (3.4), and (5.11), (5.16) is a gauge transformation

$$G:\ \Psi \to \Psi' = G\Psi \tag{5.17}$$

that transforms the AKNS systems (2.1) into (5.14).

Denote

$$w = \tan^{-1}(\psi_2/\psi_1), \quad w' = \tan^{-1}(\psi_2'/\psi_1'); \tag{5.18}$$

then the BT (4.4) can be expressed in terms of w:

$$q = q + 2w_x. \tag{5.19}$$

Now we can use the gauge transformation (5.17) or, equivalently, (3.4) to derive a relationship between w and w'. Using (5.13), (3.4), (3.5), (2.15), and (3.1), we have

$$\begin{aligned}\psi_1' + i\psi_2' &= \varphi_2' \\ &= (1/\beta^2)\{-c\varphi_1 - [(\eta+v)c - d]\varphi_2\} \\ &= (\varphi_2/\beta^2)\{-c(\varphi_1/\varphi_2) - (\eta+v)c + d\} \\ &= \varphi_2^{-1}\{c(\eta+v) - (\eta+v)c + d\} \\ &= |\varphi_2|^{-2}(\psi_1 - i\psi_2)d.\end{aligned} \tag{5.20}$$

Let μ and ν be the real and imaginary part of the complex function d defined in (3.13), respectively:

$$d = \mu + i\nu. \tag{5.21}$$

Substituting (5.21) into (5.20) we get

$$\psi_1' + \psi_2' = |\varphi_2|^{-2}[(\psi_1\mu + \psi_2\nu) + i(\psi_1\nu - \psi_2\mu)]. \tag{5.22}$$

Denote

$$w_0 = \tan^{-1}(\nu/\mu); \tag{5.23}$$

then by (5.18), (5.22), and (5.23), we obtain

$$\begin{aligned}\tan w' &= (\psi_1\nu - \psi_2\mu)/(\psi\mu + \psi\nu) \\ &= (\tan w_0 - \tan w)/(1 + \tan w_0 \tan w) \\ &= \tan(w_0 - w).\end{aligned} \tag{5.24}$$

So w possesses the transformation formula

$$w' = w_0 - w. \tag{5.25}$$

To find w_0, defined in (5.23), we must know how to obtain the real functions μ and ν in (5.21). From (3.13), (3.8), (3.17), and (3.19), we see that it is enough to clarify what the real part and imaginary part of the functional $\hat{C}_n$ are in (3.19). This can be done by the following procedure. Let S be an operator defined by

$$S = D^{-1}qD\,[\tfrac{1}{4}Dq^{-1}D + q], \tag{5.26}$$

it is easy to verify that the commutative relation

$$qEq^{-1}R = RS \tag{5.27}$$

holds, where R and E are defined in (2.23) and (2.33), respectively. Then we consider the operator E^m acting on the constant 1; repeatedly using (5.27), we have

$$\begin{aligned}E^m &= E^m\cdot 1 \\ &= E^{m-1}E\cdot 1 = \tfrac{1}{2}E^{m-1}q^{-1}(qEq^{-1}R)\cdot 1 \\ &= \tfrac{1}{2}E^{m-1}q^{-1}RS = \tfrac{1}{2}E^{m-2}q^{-1}(qEq^{-1}R)S \\ &= \tfrac{1}{2}E^{m-2}q^{-1}RS^2 = \cdots = \tfrac{1}{2}q^{-1}RS^m \\ &= S^m + i\tfrac{1}{2}q^{-1}DS^m.\end{aligned} \tag{5.28}$$

Substituting (5.28) into (2.32), we get the expression of $\hat{C}_n$ separated into real and imaginary parts as follows:

$$\hat{C}_n = 4\sum_{j=0}^{n}\sum_{m=0}^{j} S^m k_{j-m}(t)\eta^{2(n-j)} + i2q^{-1}D\left[\sum_{j=0}^{n}\sum_{m=0}^{j} S^m k_{j-m}(t)\eta^{2(n-j)}\right]. \tag{5.29}$$

Comparing (5.29) with (2.24), we get the explicit expression of the functional A in (2.7)–(2.10) in terms of a polynomial in η:

$$A = A_n = -4\sum_{j=0}^{n}\sum_{m=0}^{j} S^m k_{j-m}(t)\eta^{2(n-j)+1}. \tag{5.30}$$

Now we can express $\hat{C}_n$ in a simpler form:

$$\hat{C}_n = -(1/\eta)A_n - i(1/2\eta q)DA_n. \tag{5.31}$$

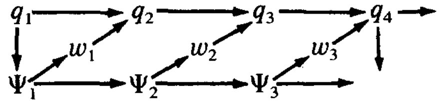
By the BT, (5.19), (5.25), and the gauge transformation (3.4) we can now obtain a hierarchy of solutions of the t-ho-mKdV equation (2.42),

$$q_1, q_2, q_3, \ldots, q_k, \ldots, \tag{5.32}$$

from a known solution q_1 of that equation and a hierarchy of solutions of the corresponding AKNS system (2.1)–(2.10),

$$\Psi_1, \Psi_2, \Psi_3, \ldots, \Psi_k, \ldots, \tag{5.33}$$

without solving any differential equation except for Ψ_1 in the following manner:



(5.34)

The equality (5.25) is, in the fact, a BT of another evolution equation. We now derive this equation. Taking the derivative with respect to x in the first equality of (5.18) and using (4.6) and (4.7), we get

$$w_x = -2\eta w - q(1 + w^2). \tag{5.35}$$

Solving for q from (5.35) gives

$$q = -(w_x + 2\eta w)/(1 + w^2). \tag{5.36}$$

Again taking the derivative with respect to t in the first equality of (5.18) and using (4.8) and (4.9), we get

$$w_t = -2Aw - Bw^2 + C. \tag{5.37}$$

Substitute (2.9) and (2.10) into (5.37),

$$w_t = -2Aw + (A_x/2q)(1 - w^2) - \left[\frac{1}{\eta}qA + \frac{1}{4\eta}\left(\frac{A_x}{q}\right)_x\right](1 + w^2); \tag{5.38}$$

further, substitute (5.30) into (5.38),

$$w_t = 4\sum_{j=0}^{n}\sum_{m=0}^{j} HS^m k_{j-m}(t)\eta^{2(n-j)+1}, \tag{5.39}$$

where H is an operator defined as

$$H = [2w - (1/2q)(1-w^2)D + (1+w^2)((1/\eta)q + (1/4\eta)Dq^{-1}D)]. \tag{5.40}$$

The function q contained in H and S can be expressed in terms of w by (5.36). Therefore (5.39) is a class of t variable higher-order nonlinear evolution equations. This is just the equation that we want to find, and possesses the BT (5.25). For $n = 1$, Eq. (5.39) reads

$$w_t + k_0[w_{xxx} + [2w^2w_x/(1+w^2)^2](12\eta^2 - 4\eta w w_x + w_x^2) - 6ww_xw_{xx}/(1+w^2)] + 4k_1w_x = 0. \tag{5.41}$$

[1]M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[2]S. S. Chern and C.-K. Peng, Manuscripta Math. **28**, 207 (1979).
[3]M. Wadati and K. Sogo, J. Phys. Soc. Jpn. **52**, 394 (1983).
[4]Y.-K. Zheng and W. L. Chan, J. Math. Phys. **29**, 308 (1988).
[5]W. L. Chan and Y.-K. Zheng, Lett. Math. Phys. **14**, 293 (1987).
[6]M. Wadati, H. Sanuki, and K. Kono, Prog. Theor. Phys. **53**, 418 (1975).
[7]C. Rogers and W. F. Shadwick, *Bäcklund Transformations and their Applications* (Academic, New York, 1982).

A generalization of manifolds as space-time models

J. Gruszczak
Department of Physics, The Pedagogical University, ul. Podchorążych, 2, 30-084 Cracow, Poland
M. Heller
Vatican Astronomical Observatory, V-00120 Vatican City State
P. Multarzyński[a]
Joint Institute for Nuclear Research, Dubna, USSR

(Received 20 May 1987; accepted for publication 27 July 1988)

A smooth manifold can be defined as the pair (M, C), where C is a family of functions on a set M, satisfying suitable axioms. The manifold concept can be easily generalized by dropping the axiom ensuring the manifold to be locally diffeomorphic to $\mathbb{R}^n$. The resulting concept, the so-called d-space, turns out to be a geometrically workable structure. The existence of the pseudo-Riemannian structure (both Riemannian and Lorentz) on d-spaces is discussed. It is proposed to model the physical space-time by a d-space rather than by a manifold. In some quantum gravity situations space-time may still be a d-space but already not a manifold.

I. INTRODUCTION

The set $C(X)$ of continuous functions on a "reasonable" topological space X forms a ring, and the set C_x of functions that vanish at a point $x \in X$ forms a maximal ideal. Moreover, each maximal ideal in $C(X)$ is of the form C_x for some $x \in X$. Since the space of maximal ideals in $C(X)$ is isomorphic to X, one can reconstruct the geometric and topological structure of X from the knowledge of the algebraic structure of the ring $C(X)$. In the spirit of this program, Penrose and Rindler,[1] following Chevalley[2] and Nomizu,[3] have developed the theory of differential manifolds[4] (d-manifolds, for the sake of brevity) by defining a manifold to be an abstract set of points, the structure of which is determined by a nonempty set $\mathscr{T}$ of scalar fields on M satisfying suitable axioms. However, it turns out that if one drops the axiom enforcing the manifold to be locally diffeomorphic to the Euclidean space of some dimension, one obtains the more general (but still geometrically manageable) concept of the so-called *differential space* or *d-space*, for short.

Quite a number of different generalizations of the manifold concept were proposed by mathematicians. For instance, Aronszajn[5] and Marshall[6] developed the theory of the so-called *sub-Cartesian spaces* that essentially are manifolds with "singularities" such as piecewise manifolds and quasianalytic sets of $\mathbb{R}^n$. A certain modification of this approach was suggested by Spallek[7] (see also Ref. 8). Mostow[9] introduced his concept of the differential space within the context of Milnor's classifying spaces. Chen[10] considered a differential space structure in the loop space which led him to a version of the de Rham theorem in this space. A theory of differential spaces, together with the de Rham theorem, was also elaborated by Smith.[11] In Ref. 12 we compare these generalizations of the smooth manifold concept and establish some dependences between them.

As far as we know, no attempt has ever been made to employ such generalizations to model the physical space-time, and this is precisely the goal of the present paper. For reasons established in Ref. 12, we shall do this with the help of Sikorski's approach (in this approach generality seems to be best balanced with workability). In order to be able to fulfill our goal, we must develop and adapt mathematical formalism to physical purposes (especially, the existence of the Lorentz structure should be fully discussed). The beautiful monograph by Sikorski[13] (which, unfortunately, exists only in the Polish version) presents differential geometry in terms of d-spaces. Here we shall give only necessary definitions and theorems, providing the English reader with a more comprehensive review in Ref. 14.

Of course, every d-manifold is a d-space, but not vice versa. Those d-spaces that are not d-manifolds will be called *d-spaces proper*. The macroscopic space-time of contemporary relativistic physics doubtlessly should be modeled by a four-dimensional d-manifold, but it should be expected that when going deeper and deeper to smaller and smaller scales or closer and closer to the cosmological singularity, one reaches the level at which the space-time is already not a d-manifold but it still continues to be a d-space.

In other words, sufficiently near to the cosmological singularity or at sufficiently small scales general relativity is commonly believed to break down. We might speculate that it is the principle of equivalence that, in such circumstances, ceases to be valid. This principle, in turn, is coded into the geometric structure that postulates that locally every space-time should be diffeomorphic to that of special relativity (i.e., that the gravitational field locally can be transformed away). A space-time manifold without this structure is exactly a d-space proper.

The differential space method turns out to be a very efficient tool in dealing with the classical singularity problem. Singularities (at least some kinds of them) need not be considered as belonging to "singular boundaries" of space-time, but can be regarded as "internal domains" of a corre-

[a] On leave of absence from the Institute of Nuclear Physics, ul. Radzikowskiego 152, 31-342 Cracow, Poland.

sponding d-space. Moreover, a d-manifold with boundary turns out to be a d-space, and one may find it useful to study space-time boundaries (such as the causal boundary or Schmidt's b-boundary) in terms of the d-space theory. We shall present some striking results concerning these fields of research in a forthcoming paper.

We begin our study with a short description of how the topology of a set is determined by a family of functions defined on it (Sec. II). Then the definition of d-space (Sec. III) is followed by a comparison of the d-space concept with that of a d-manifold (Sec. IV). The tangent space of a d-space is defined, and the dimensionality problem of d-spaces is discussed in Sec. V (dimension is not a part of the d-space definition). Our aim is to introduce pseudo-Riemannian structures (both Riemannian and Lorentz) on d-spaces (Sec. VI), and to prove theorems on their existence (Sec. VII). In Sec. VIII, we reflect briefly upon the physical significance of our proposal to model the physical space-time by a d-space rather than by a d-manifold.

II. TOPOLOGIES IN TERMS OF A FAMILY OF FUNCTIONS

Let M be any set, $\mathscr{S}$ a family of topological spaces, and C a nonempty family of functions defined on M with values in a topological space of the family $\mathscr{S}$, i.e., $C=\{f\colon M\to S_f \in\mathscr{S}\}$. The weakest topology in which all functions $f\in C$ are continuous will be called the topology *induced* by C, and denoted by T_C.

Let $\mathscr{P}$ be a family of partially ordered sets. The partial order relation in a set $Q\in\mathscr{P}$ will be denoted by $\leqslant_Q$. We define the relation $<_Q$ in the following way: if $x,y\in Q$, $x<_Q y$ iff $x\leqslant_Q y$ and $x\neq y$. The topology, with the subbase consisting of all sets of the form $\{p\in M\colon f(p)<x\}$ and $\{p\in M\colon f(p)>x\}$, $f\in C\subset\{f\colon M\to\mathscr{P}_f\in\mathscr{P}\}$, $x\in\mathscr{P}$, is said to be the topology T_C *induced* in M by a nonempty family C.

The set $\mathbb{R}$ of reals is both a topological space (with the natural topology) and an ordered set (with the natural order relation). Let $C\subset\{f\colon M\to\mathbb{R}\}$. The topology in M can be introduced according to the above-mentioned methods, provided we treat $\mathbb{R}$ as a topological space or an ordered set, respectively. It is evident that in this case both topologies coincide, and we have the topology T_C uniquely defined in M.

III. d-SPACES

Let (M, T) be any topological space, and C any nonempty family of functions on M with values in any set.

Definition 3.1: A function f, defined on $A\subset M$, is said to be a *local C-function*, if for every $p\in A$ there is a neighborhood B in the topological subspace (A, T_A), where T_A is the topology induced in A by T, and a function $g\in C$ such that $f|_B=g|_B$. The set of local C-functions on $A\subset M$ will be denoted by C_A. We obviously have $C\subset C_M$.

Definition 3.2: The set C is said to be *closed with respect to localization* if $C=C_M$.

Definition 3.3: Let C be a family of real functions on M, and $\mathscr{E}$ the set of all C^∞-functions (smooth functions) on $\mathbb{R}^n$. Then C is said to be *closed with respect to superposition with smooth functions on the Euclidean space* $\mathbb{R}^n$ if for any natural number n and any function $\omega\in\mathscr{E}$ one has $f_1,\dots,f_n\in C \Rightarrow \omega\circ(f_1,\dots,f_n)\in C$.

Definition 3.4: A family C of real functions on M is said to be a *differential structure* (*d-structure*) on M if it is closed with respect to localization and closed with respect to superposition with smooth functions on the Euclidean space $\mathbb{R}^n$, for any $n\in N$. The pair (M, C) is called the *differential space* (*d-space*); the set M being the *support* of the d-structure C. We shall always assume that M is a topological space with the topology T_C induced by C.

Let us notice that the dimensionality of (M, C) is not a part of the above definition.

It can be easily shown that if C is a d-structure, C is an algebra (in the foreign literature called also a linear ring) over $\mathbb{R}$ containing all constant functions (see Ref. 13, p. 77). A d-space (M, C) is a Hausdorff topological space iff for any $p, q\in M$, there is a function $f\in C$ such that $f(p)\neq f(q)$. Any d-space (M, C) can always be made Hausdorff. Indeed, define the equivalence relation $p\# q$ iff $f(p)=f(q)$, for all $f\in C$, and consider the set $M/\#$. In the following, we shall always assume that all d-spaces considered are Hausdorff.

Proposition 3.5: Let M be any set. For any set C_0 of real functions on M, there is the smallest d-structure C such that $C_0\subset C$, and the topology T_C coincides with the topology T_{C_0}. Then C_0 is said to *generate* the d-structure C. ■

Proof: Let C be the family of all functions $f\colon M\to\mathbb{R}$ such that $f\in C$ iff, for every $p\in M$, there is a neighborhood U of p, in the topology T_{C_0}, such that $f|_U=\omega\circ(f_1,\dots,f_n)_U$, where $f_1,\dots,f_n\in C_0$ and $\omega\in\mathscr{E}$. The family C is closed with respect to localization and closed with respect to superposition with smooth functions on $\mathbb{R}^n$. One can easily check that if C_1 is a d-structure and $C_0\subset C_1\subset C$, then $C_1=C$. Any function $f\in C$ is continuous in the topology T_{C_0} and $C_0\subset C$ implies $T_C=T_{C_0}$. □

Instead of quoting examples of d-spaces let us prove the following.

Proposition 3.6: Every subset A of $\mathbb{R}^n$ can be made into a d-space. ■

Proof: Let $\pi_i\colon \mathbb{R}^n\ni(x_1,\dots,x_n)\to x_i\in\mathbb{R}$, and $p_i=\pi_i|_A$, $i=1,\dots,n$. The topology T_{p_i} coincides with the topology on A induced from $\mathbb{R}^n$. The family $\{p_i\}$ generates the d-structure $\mathscr{E}_A$ on A. □

Let (M, C) be a d-space, and $N\subset M$. One can easily see that (N, C_N), where C_N is the set of C-functions on N, is also a d-space. It will be called the *d-subspace* of the d-space (M, C).

IV. d-SPACES AND d-MANIFOLDS

Definition 4.1: Let (M, C) and (N, D) be d-spaces. A one-to-one mapping $f\colon M\to N$ is said to be the *diffeomorphism* of (M, C) onto (N, D), iff $h\circ f\in C$ and $g\circ f^{-1}\in D$, for each $h\in D$, $g\in C$.

Definition 4.2: A d-space (M, C) is said to be an *n-dimensional differential manifold* (*d-manifold*, for short) if, for every $p\in M$, there is a neighborhood $V\in T_C$ of p and a neighborhood $U\in T_{\mathscr{E}}$, U being a subset of $\mathbb{R}^n$, and a diffeomorphism of $(U,\mathscr{E}_U)$ onto (V,C_V); the pairs $(U,\mathscr{E}_U)$ and

(V, C_V) are d-subspaces of $(\mathbb{R}^n, \mathscr{E})$ and (M, C), respectively.

It can be shown that Definition 4.2 is equivalent to the time honored definition of C^∞-manifolds in terms of atlases on M. Indeed, if $(M, \mathscr{A})$ is a d-manifold, where M is an n-dimensional topological (Hausdorff) space, and $\mathscr{A}$ an atlas on M, i.e., the set of all maps g: $U \to M$, U being an open subset of $\mathbb{R}^n$, satisfying the well-known conditions, then a function f: $M \to \mathbb{R}$ is said to be a *smooth function on M* if, for every function $g \in \mathscr{A}$, the function $f \circ g$ is C^∞. If C denotes all smooth functions on M, (M, C) is an n-dimensional d-manifold in the sense of Definition 4.2. And vice versa, if (M, C) is a d-manifold in the sense of Definition 4.2, one can easily see that functions of the form $h \circ g \in \mathscr{E}$, where g: $U \to M$, $U \subset \mathbb{R}^n$, and $h \in C$, form maps belonging to the atlas $\mathscr{A}$. (For details see Ref. 13, pp. 100–104.)

Moreover, the definition of d-manifolds in terms of the family C turns out to be more natural than the traditional one. Let $\mathscr{A}_1$ and $\mathscr{A}_2$ be two atlases on M and C_1 and C_2 two families of all-smooth functions determined by the atlases $\mathscr{A}_1$ and $\mathscr{A}_2$, respectively. We have the following proposition.

Proposition 4.3: $C_1 = C_2$ iff, for any $f \in \mathscr{A}_1$ and $g \in \mathscr{A}_2$, $g^{-1} \circ f$ is a diffeomorphism. □

This proposition establishes an equivalence relation between d-manifolds: $(M, \mathscr{A}_1) \sim (M, \mathscr{A}_2)$ iff $f \in \mathscr{A}_1$ and $g \in \mathscr{A}_2$ implies that $g^{-1} \circ f$ is a diffeomorphism. Two d-manifolds belonging to the same equivalence class are indistinguishable from the point of view of differential geometry, and, strictly speaking, one should work not with individual d-manifolds but with their equivalence classes. From this point of view, atlases turn out to be contingent and nonessential entities. On the other hand, families C of functions on M identify a given equivalence class uniquely. Moreover, by defining d-space as a pair (M, C), M can be assumed to be any set and there is no need to ascribe to it, from the beginning, the structure of a topological space.

It should be noticed that one can define a C^r-manifold by assuming that a d-space (M, C) is locally diffeomorphic (in the sense of Definition 4.1) to $(\mathbb{R}^n, \mathscr{E}^{(r)})$, where $\mathscr{E}^{(r)}$ is a d-structure on $\mathbb{R}^n$ containing C^r-functions.

V. TANGENT SPACE OF A d-SPACE. DIFFERENTIAL AND TOPOLOGICAL DIMENSIONS

In the following we shall assume that (M, C) is a d-space (not necessarily a d-manifold), and the functions belonging to C will be called *smooth functions* on M.

Definition 5.1: Any linear mapping v: $C \to \mathbb{R}$, satisfying the Leibniz condition

$$v(fg) = v(f)g(p) + f(p)v(g),$$

for $f, g \in C$, is said to be a *tangent vector* to a d-space (M, C) at $p \in M$, and the set of all tangent vectors to (M, C) at p is called the *tangent space* to (M, C) at p, and will be denoted by M_p.

One should notice that M_p is a nonempty set since the zero mapping 0: $C \to \mathbb{R}$ defined by $f \mapsto 0$, $f \in C$, belongs to M_p (for every $p \in M$). It is also easily seen that M_p is a vector space.

Definition 5.2: A cross section of the tangent bundle $\bigcup_{p \in M} M_p$, i.e., a mapping V: $M \to \bigcup_{p \in M} M_p$ such that $\pi \circ V = \mathrm{id}_M$, where π is the bundle projection, is said to be a *tangent vector field* on the d-space (M, C). The tangent vector field on (M, C) is said to be *smooth* if, for every $f \in C$, the mapping $V(\cdot)(f)$: $M \to \mathbb{R}$, defined by $M \ni p \mapsto V(p)(f)$, belongs to C. For definitions of tensors and smooth tensor fields on d-spaces the reader should refer to the literature.[13,14] It is worthwhile to notice that any vector bundle can be defined in terms of the theory of d-spaces (see Ref. 14).

The theory of tangent spaces of d-spaces can be developed analogously to that of d-manifolds. However, one should pay attention to some important peculiarities of the d-space concept, one of the most striking being the notion of dimension.

Definition 5.3: (1) A number $n \in N$ is called the (global) *differential dimension* (*d-dimension*) of (M, C) if (i) $\dim(M_p) = n$, for every $p \in M$; (ii) for every $p \in M$ and every vector $v \in M_p$, there is a smooth tangent vector field V on (M, C) such that $V(p) = v$. (2) The dimension of M_p, $\dim(M_p)$, is called the *local dimension* (*l-dimension*) of the d-space (M, C) *at the point p*.

Example 5.4: Let $(\mathbb{R}^2, \mathscr{E})$ be a d-space, the so-called Euclidean d-space, and let us consider the subset of $\mathbb{R}^2$: $A = \{(x, y) \in \mathbb{R}^2: xy = 0\}$. The pair $(A, \mathscr{E}_A)$ is a d-space. It is evident that $\dim(A_p) = 1$, for all $p \neq (0, 0)$. Let $f \in \mathscr{E}_A$; $p = (x, y) \in A$ implies $f(x, y) \in \mathbb{R}$, and there are two linear mappings V_1, V_2: $\mathscr{E}_A \to \mathbb{R}$ defined by $f \mapsto (\partial f / \partial x)(0, 0)$ and $f \mapsto (\partial f / \partial y)(0, 0)$, correspondingly. The mappings satisfy the Leibniz condition and are linearly independent. Therefore, they span the two-dimensional tangent space of the d-manifold $(A, \mathscr{E}_A)$ at the point $p = (0, 0)$. This d-space, owing to the existence of the point $p = (0, 0)$, fails to be a d-manifold. Clearly, the d-space $(A, \mathscr{E}_A)$ has no (global) d-dimension.

For d-manifolds, l-dimension is the same everywhere and is equal to its topological dimension (e.g., in the sense of Menger or Urysohn). Both these statements are not true with respect to d-spaces proper. The following example illustrates that d-dimension is a property of d-structure C rather than of its support M.

Example 5.5: Let $(\mathbb{R}, C_1)$ and $(\mathbb{R}, C_2)$ be two d-spaces with different d-structures defined in the following way: $C_1 = \{f: \mathbb{R} \to \mathbb{R}, f \in C^0$ and has one-sided derivatives$\}$, $C_2 = \mathscr{E} = C^\infty(\mathbb{R})$. Both d-spaces have the same topology and therefore the same topological dimension, whereas the d-dimension of $(\mathbb{R}, C_1)$ is equal to 2, and that of $(\mathbb{R}, C_2)$ is equal to 1.

All these peculiarities follow from the fact that d-dimension is invariant with respect to diffeomorphisms but not with respect to homeomorphisms.

VI. PSEUDO-RIEMANNIAN d-SPACES

The set $\mathfrak{M}(M)$ of all smooth tangent vector fields on (M, C), from the algebraic point of view, is a module over the algebra C of all smooth real functions on M. The set $V_1, \ldots, V_n \subset \mathfrak{M}(M)$ is said to be the *vector basis of the module* $\mathfrak{M}(M)$ if (i) for every $p \in M$, $V_1(p), \ldots, V_n(p)$ is a basis in M_p, and (ii) $V_1, \ldots, V_n$ is a C-basis of the module $\mathfrak{M}(M)$, i.e., if, for every $V \in \mathfrak{M}(M)$, there is exactly one finite sequence $f^i \in C$, $i = 1,\ldots, n$, such that

$$V = \sum_i f^i V_i.$$

Definition 6.1: Let $G(p)$ denote a scalar product in the tangent space M_p. From now on, we additionally assume that the module $\mathfrak{M}(M)$ has a finite vector basis (and consequently M has a d-dimension); such modules will be called *differential modules*. We define the *scalar product in* $\mathfrak{M}(M)$ by $G(V, W)(p) = G(p)(V(p), W(p))$, $V, W \in \mathfrak{M}(M)$, $p \in M$. The scalar product G is said to be *smooth* if, for any $V, W \in \mathfrak{M}(M)$, $G(V, W) \in C$, i.e., if $G \in \mathscr{L}_C(\mathfrak{M}(M), \mathfrak{M}(M); C)$, where the last symbol denotes the set of all bilinear module mappings with values in C.

In other words, G is a two-covariant C-tensor, symmetric and nondegenerate.

Theorem 6.2: If G is a smooth scalar product in the differential module $\mathfrak{M}(M)$, every point $p \in M$ has a neighborhood U on which there is a G-orthonormal vector basis $V_1,\ldots, V_n$ of the module $\mathfrak{M}(M)$, i.e., $G(V_i, V_j) = \delta^i_j \epsilon_i$, where $\epsilon_i = G(V_i, V_i) = \pm 1$, $i, j = 1,\ldots,n = \dim(\mathfrak{M}(M))$. ■

Proof: By construction through the standard Gramm–Schmidt G-orthogonalization (see, Ref. 13, p. 311). □

Whenever convenient, the G-orthogonalization may be ordered so that negative signs (if any) come first. The number I of minus signs is basis independent and the same everywhere in $\mathfrak{M}(M)$; it is called the *index* of the module $\mathfrak{M}(M)$.

Definition 6.3: The pair $(\mathfrak{M}(M), G)$, where $\mathfrak{M}(M)$ is a differential module and G a smooth scalar product on it, is called a *pseudo-Riemannian module* on the d-space (M, C). Also, G is referred to as the *metric* on (M, C). If additionally $I = 0$ or $I = \dim(\mathfrak{M}(M))$, $(\mathfrak{M}(M), G)$ is called a *Riemannian module* on (M, C); if $I = 1$ or $I = n - 1$, it is called a *Lorentz module* on (M, C).

Theorem 6.4: Let $(\mathfrak{M}(M), G)$ be a pseudo-Riemannian module. There is one and only one covariant derivative D in $\mathfrak{M}(M)$ such that $DG = 0$. (For the elementary proof, see Ref. 13, pp. 224–227.) □

The covariant derivative of the above theorem is called the *natural derivative* in $\mathfrak{M}(M)$.

Definition 6.5: The triple (M, C, G) is said to be a *pseudo-Riemannian d-space* if (M, C) is a d-space, and $\mathfrak{M}(M)$ a pseudo-Riemannian differential module on (M, C). If $(\mathfrak{M}(M), G)$ is a Riemann or Lorentz differential module on (M, C), (M, C, G) is said to be *Riemannian* or *Lorentz d-space*, respectively.

If in Definition 6.5 "d-space" is replaced by "d-manifold" one obtains the usual *pseudo-Riemannian*, *Riemannian* or *Lorentz d-manifold*, respectively.

Two pseudo-Riemannian d-spaces (M_1, C_1, G_1) and (M_2, C_2, G_2) are said to be *isometric* if there is a diffeomorphism f: $(M_1, C_1) \to (M_2, C_2)$, which preserves the scalar product. It can be shown that any isometry transforms the natural covariant derivative into the natural covariant derivative. One can also define curvature tensors on pseudo-Riemannian d-spaces,[15] and write down on them Einstein's equations. However, for the time being we postpone these interesting topics to focus on the existence of the Lorentz structure on d-spaces.

VII. THE EXISTENCE OF THE RIEMANN AND LORENTZ STRUCTURES ON d-SPACES

Let (M, C) be a d-space, and T and W two tensor fields on M having the same number of upper and lower indices. For any $f \in C$, we define the f-equivalence relation, $T \sim W$, iff $T|_{\mathscr{F}_f} = W|_{\mathscr{F}_f}$, where $\mathscr{F} := \{p \in M\colon f(p) \neq 0\}$. For a given $f \in C$, let $\mathfrak{M}^{(k)}_{(l)}(f)$ denote the set of all f-equivalence classes, where (k) and (l) signify the number of upper and lower indices of a given tensor field, correspondingly.

Lemma 7.1: $C(f) := \mathfrak{M}^{(0)}_{(0)}(f)$ with the natural definitions of sums and products is a commutative ring with identity. Here $\mathfrak{M}(f) := \mathfrak{M}^{(0)}_{(0)}(f)$ is a module over $C(f)$ (see Ref. 1, p. 99). □

Theorem 7.2: In a differential d-module $\mathfrak{M}(M)$ of all smooth vector fields on (M, C), there exists a Riemann metric G [so that $(\mathfrak{M}(M), G)$ is a Riemann module on the d-space (M, C)] if the following conditions are satisfied: (i) there exists a finite set of non-negative functions $f_1,\ldots, f_d \in C$ such that

$$\sum_{i=1}^{d} f_i = 1;$$

(ii) each module $\mathfrak{M}^{(1)}_{(0)}(f_i)$ has a finite vector basis. ■

Proof: Essentially, conditions (i) and (ii) define a partition of unity on M (see Ref. 1, p. 99). With the help of the partition, one assembles together, in a standard way,[16] all Riemann metrics defined locally. □

From condition (ii) of the above theorem, it follows that a d-space (M, C) carrying a Riemann structure must have a d-dimension. However, one should notice that in order to carry a Riemann structure a d-space need not be a d-manifold.

Definition 7.3: Let M_p be a tangent space of a d-space (M, C) at a point $p \in M$. A one-dimensional vector subspace Q_p of M_p is called a *direction* (or a *line element*) in M_p. The function Q: $p \mapsto Q_p$ is called a *direction field* on (M, C).

It is worth noticing that, in view of Definition 5.1, a direction field is, in fact, defined above in terms of a mapping from C to $\mathbb{R}$.

Theorem 7.4: In a differential d-module $\mathfrak{M}(M)$, there exists a Lorentz metric G [so that $(\mathfrak{M}(M), G)$ is a Lorentz module on the d-space (M, C)] if, in addition to conditions (i) and (ii) of Theorem 7.2, a continuous direction field Q: $p \mapsto Q_p$, $p \in M$, exists on (M, C). ■

Proof: The proof is a repetition of the standard demonstration of the existence of a Lorentz structure on a d-manifold (e.g., Ref. 16, p. 293). □

VIII. COMMENTS

As we have seen, the definition of a d-manifold in terms of the algebra of functions is at least as workable as the traditional one in terms of atlases, and the former could equally well be employed to model the physical space-time as the latter has done it for a long time with great success. In fact, the former approach has been chosen by Penrose and Rindler.[1] The algebra of functions, defining the manifold structure, is interpreted by them as a "system of scalar fields." A justification for such an option could be that the measurement results are always scalars [although what is called by physicists "observable" is a function on the phase space (cotangent bundle) rather than on space-time itself]. In the new approach "even coordinate systems may be thought of simply as sets of scalar fields" (Ref. 1, p. 180). One should notice that, in view of Proposition 3.5 above, if C_0 is any set of scalar fields, there is a d-structure C generated by C_0 (i.e., the smallest C such that $C_0 \subset C$). Correspondingly, one could truly speak of a d-space determined by measurement results (by C_0).

Changing from the manifold definition in terms of atlases to that in terms of the algebra of functions might be a question of elegance or of better or worse operational intuition, but one can hardly expect new physical insights. An interesting advantage of such a change is the possibility of generalizing the space-time model and to consider it to be a d-space rather than a d-manifold. It is true that in ordinary macroscopic situations local resemblance of space-time to a patch of a smooth Euclidean space is a natural feature. However, it is also highly restrictive. As we have seen a relaxation of these restrictions still leaves quite a lot of an effectively workable structure. This may turn out to be desirable when one penetrates more exotic situations (quantum gravity level, space-time with boundary—as was discussed in the Introduction).

Therefore, we propose to consider the triple $(M, C, (\mathfrak{M}(M), G))$, where (M, C) is a d-space and $(\mathfrak{M}(M), G)$ is a Lorentz differential module on (M, C), as a model for physical space-time. If the d-space in question is a d-manifold, the traditional model of space-time is automatically obtained.

One could feel a little disappointed that the metric G in $\mathfrak{M}(M)$ presupposes a d-dimension of (M, C). Some successes of multidimensional cosmologies (of Kaluza–Klein, supergravity, superstring, or combined types) in the last years, have evoked a hope that the dimension of space-time could possibly be treated as a dynamical variable, the number of dimensions being undetermined in the beginning, and then gradually emerging as enforced by the dynamics of the universe. The present work shows that it is the Lorentz metric that requires a fixed number of space-time dimensions. In spite of the above-mentioned disappointment, this is a valuable result: in order to manipulate dimension, one must go beneath the metric level. If it is the quantization of gravity that is expected to determine the dimensionality of space-time, it cannot be a quantization of space-time metric; it has to be a quantization of at least the space-time d-structure. The orthodox theory of general relativity seems to be too restrictive to face such exotic expectations.

The postulate that a d-space modeling space-time should be a d-manifold is, in fact, the implementation of Einstein's idea that locally space-time is indistinguishable from a small region of the Euclidean space. If one models space-time by a d-space proper, this version of the equivalence principle is automatically discarded. How does physics look without the principle of equivalence? Most probably, this question will be answered when the correct form of a quantum cosmology is elaborated for the very early universe.

We are far from thinking that the d-space model of space-time, argued for in the present work, will solve all problems of contemporary physics and cosmology. We argue that it is worth being explored as a step towards this ambitious goal. But even this provisional model is not yet complete. Parallel propagation structure, curvature structure, and perhaps Einstein's equations on d-spaces await their elaboration. We hope to deal with these problems in our subsequent works.

ACKNOWLEDGMENT

The work of two of the authors (J. G. and P. M.) was supported by the Polish Interdisciplinary Research Project CPBP 01 03.

[1]R. Penrose and W. Rindler, *Spinors and Space-Time*, Vol. 1 of *Two-Spinor Calculus and Relativistic Fields* (Cambridge U.P., Cambridge, 1984).
[2]C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, NJ, 1946).
[3]K. Nomizu, *Lie Groups and Differential Geometry* (Mathematical Society of Japan, Tokyo, 1956).
[4]Against the common usage we propose, after Sikorski,[13] to replace the term *differentiable* manifold (and correspondingly *differentiable* space) by the term *differential* manifold (space), the difference being analogous to that between metrizable and metric space.
[5]N. Aronszajn, Not. Am. Math. Soc. **14**, 111 (1967).
[6]Ch. D. Marshall, J. Differ. Geom. **10**, 575 (1975).
[7]K. Spallek, Rend. Math. **4**, Ser. VI, 237 (1971); **5**, Ser. VI, 375 (1971).
[8]M. Breuer and C. Marshall, Math. Ann. **237**, 105 (1978).
[9]M. A. Mostow, J. Differ. Geom. **14**, 255 (1979).
[10]K. T. Chen, Bull. Am. Math. Soc. **83**, 831 (1977).
[11]J. W. Smith, Tohoku Math. J. **18**, 115 (1966).
[12]M. Heller, P. Multarzyński, W. Sasin, and Z. Żekanowski, "On some generalizations of the manifold concept and their applications to physics," to be published.
[13]R. Sikorski, *Wstep do geometrii rozniczkowej* (*An Introduction to Differential Geometry*) (PWN, Warsaw, 1972).
[14]M. Heller, P. Multarzyński, and W. Sasin, to be published in Acta Cosmologica.
[15]A. Kowalczyk, in *Differential Geometry*, Banach Center Publications, Vol. 12 (Polish Scientific, Warsaw, 1984), p. 163.
[16]See, for example, Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (revised edition) (North-Holland, Amsterdam, 1982), p. 292.

Theory of fluctuations and small oscillations for quantum lattice systems

D. Goderis,[a] A. Verbeure, and P. Vets[a]
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

(Received 17 May 1988; accepted for publication 3 August 1988)

The theory of fluctuations for a quantum lattice system is rigorously defined. The property of stability of equilibrium states against macroscopic fluctuations is formulated and small oscillations around equilibrium are studied.

I. INTRODUCTION

It is well known that for macroscopic systems there are essentially two types of observables that are primordially relevant in physics. The first type being the so-called extensive observables, which are described for infinite systems by their space means, e.g., energy densities, entropy densities, mean magnetization, etc. In general, take any local observable $A(x)$ localized around the space point $x\in\mathbb{R}^\nu$ ($\nu = 1,2,...$); then the mathematical problem is to prove the existence of the limit

$$\lim_{V\to\mathbb{R}^\nu} \frac{1}{V}\int_V dx\, A(x) = \overline{A},$$

mostly taken in an equilibrium state of the system, and for a specific choice of sequence of volumes V.

In mathematical physics, the set of observables $\overline{A}$ obtained in this way are called observables at infinity,[1] because these observables are independent of the strictly local structure of $A(x)$ for small x. From the point of view of probability theory they are obtained as the limit points in the law of large numbers, and the convergence to the limit is exponential with a rate function determined by the principle of large deviations.[2] Always in an equilibrium state the set of observables at infinity form a commutative algebra pointwise invariant for the time evolution.

The second type of observables that are important in the physics of systems with a large number of degrees of freedom are the so-called fluctuation observables or macroscopic fluctuations. Denote by ω some state of the system and again by $A(x)$ a localized observable, then the mathematical problem here is to prove under suitable conditions the existence of the limit

$$\lim_{V\to\mathbb{R}^\nu} \frac{1}{\sqrt{V}}\int_V dx(A(x) - \omega(A(x))) = B_\omega(A).$$

Recently[3] we studied the mathematical aspects of this limit and proved quantum mechanical central limit theorems. We obtained a complete mathematical description of the central limits $B_\omega(A)$ for all local A. The set of fluctuations $B_\omega(A)$ forms again an algebra but this time a noncommutative one satisfying specific canonical commutation relations. Moreover the natural time evolution of the system induces a nontrivial time evolution on the fluctuations. At this point one should refer to the existing literature, where one computes fluctuations for particular examples of mean-field models.[4–6] Here we are interested in a model-independent theory of fluctuations and we are particularly investigating the quantum aspects. In this paper we implement and apply the results of Ref. 3 in the algebraic approach of statistical mechanics, i.e., a mathematical theory of macroscopic fluctuations is presented for systems with, what one calls, normal fluctuations. As mean-field systems are the prototypes of such systems we limit ourselves here to such systems, remembering, however, that for most of the results the only condition is that the system shows normal fluctuations.

In Sec. II we provide the necessary material about mean-field systems to give in Sec. III the theory of fluctuations. Our main result here is that we show that macroscopic fluctuations are used to consider perturbations of the equilibrium state. In fact, we prove that the equilibrium state is stable against its macroscopic fluctuations. Technically we prove that the relative entropy is the relevant notion to extremalize in order to get the equilibrium state. We derive the explicit formula for the relative entropy and prove that it is a quadratic expression in the perturbation. Remark that all this remains within the frame of equilibrium statistical mechanics.

Finally in Sec. IV we make clear how the initial time evolution of the system induces a time evolution on the set of macroscopic fluctuations and derive some of its properties.

II. MEAN FIELD SYSTEMS

The prototype of a quantum system with normal fluctuations are the so-called mean field models. Mathematically they have the following structure. Denote by M the algebra of $m\times m$ complex matrices and by $\mathscr{B}$ the C^*-algebra generated by the sequence $\otimes_{i=1}^{n} M_i$, where M_i is a copy of M. For each $X\in M$, denote by X_i the imbedding of X in $\mathscr{B}$,

$$X_i = 1\otimes\cdots\otimes X\otimes 1\otimes\cdots,$$

where X is in the ith site.

Let ρ be a state of M, then we denote by ω_ρ the product state of $\mathscr{B}$ defined by

$$\omega_\rho(X\otimes Y\otimes\cdots\otimes Z) = \rho(X)\rho(Y)\cdots\rho(Z), \tag{2.1}$$

for $X,Y,...,Z\in M$. Remark also that the state ρ of M is determined by a density matrix that we denote by the same ρ, i.e., $\rho(X) = \operatorname{tr}\rho X$, $X\in M$.

A mean-field model is given by the following local Hamiltonians:

$$H_N = \sum_{i=1}^{N} A_i + \frac{1}{2N}\sum_{\substack{i\neq j\\ i,j=1}}^{N} B_{ij}, \tag{2.2}$$

where A_i are copies of $A^* = A\in M$ and B_{ij} are copies of

[a] Onderzoeker IIKW, Belgium.

$B^* = B \in M \otimes M$ such that B_{ij} is invariant under the symmetry $B_{ij} = B_{ji}$.

The equilibrium states of these systems are given by states ω_ρ such that the density matrix ρ satisfies the gap equation[7]

$$\rho = e^{-\beta H_\rho}/\mathrm{tr}\, e^{-BH_\rho}, \tag{2.3}$$

where

$$H_\rho = A + B_\rho, \quad B_\rho = \mathrm{tr}_2(1 \otimes \rho)B, \tag{2.4}$$

and tr_2 is the partial trace over the second space.

Denote by $\{E_{ij} | i,j = 1,...,m\}$ the set of matrix units of M in the basis diagonalizing the effective Hamiltonian H_ρ, i.e., the E_{ij} satisfy

$$\sum_i E_{ii} = 1, \quad E_{ij}^* = E_{ji},$$

$$E_{ij}E_{kl} = \delta_{jk}E_{il},$$

$$H_\rho E_{ij} = \epsilon_i E_{ij}, \quad E_{ij}H_\rho = \epsilon_j E_{ij},$$

the $\{\epsilon_i | i = 1,...,m\}$ are the eigenvalues of H_ρ. For technical simplicity, but without loss of generality we assume that the eigenvalues are nondegenerate and that they are ordered as follows:

$$\epsilon_1 > \epsilon_2 > \cdots > \epsilon_n.$$

Of course, an arbitrary element $X \in M$ takes now the form

$$X = \sum_{i,j} x_{ij} E_{ij}.$$

The effective time evolution on M in the equilibrium state is now given by

$$\alpha_t(X) = e^{itH_\rho} X e^{-itH_\rho}, \quad X \in M. \tag{2.5}$$

Denote by M_{sa} the real vector space of the self-adjoint elements of M and by M_{sa}^0 the constants of the motion, i.e.,

$$M_{\mathrm{sa}}^0 = \{X \in M_{\mathrm{sa}} | \alpha_t(X) = X, \text{ for all } t \in \mathbb{R}\}.$$

Then M_{sa} has a unique decomposition

$$M_{\mathrm{sa}} = M_{\mathrm{sa}}^0 \oplus M_{\mathrm{sa}}^1.$$

A basis for M_{sa}^1 is given by

$$e_{kl} = E_{kl} + E_{lk}, \quad f_{kl} = i(E_{kl} - E_{lk}), \text{ with } k < l.$$

Any element $X \in M_{\mathrm{sa}}^1$ has a unique decomposition

$$X = \sum_{k<l} x_{kl} e_{kl} + \tilde{x}_{kl} f_{kl}, \quad x_{kl}, \tilde{x}_{kl} \in \mathbb{R}.$$

Define the linear operator J on M_{sa}^1 by

$$Je_{ij} = f_{ij}, \quad J^2 = -1,$$

and the real bilinear form s_ρ on M_{sa}^1 by

$$s_\rho(X,Y) = -i\rho([X,JY]), \quad X,Y \in M_{\mathrm{sa}}^1. \tag{2.6}$$

Proposition 2.1: The bilinear form s_ρ satisfies

(i) $s_\rho(X,Y) = s_\rho(Y,X)$, $X,Y \in M_{\mathrm{sa}}^1$;

(ii) $s_\rho(JX,Y) = -s_\rho(X,JY)$;

(iii) $s_\rho(X,X) > 0$, for all $X \neq 0$;

(iv) $Xs_\rho(e_{ij},e_{i'j'}) = s_\rho(f_{ij},f_{i'j'}) = \delta_{ii'}\delta_{jj'}s_\rho(e_{ij},e_{ij})$, $s_\rho(e_{ij},f_{kl}) = 0$.

Proof: Remark that $\rho(E_{ij}) = 0$ for $i \neq j$ (time invariance of the state ρ), then for

$$X = \sum_{i<j} x_{ij}e_{ij} + \tilde{x}_{ij} f_{ij}$$

and

$$Y = \sum_{i<j} y_{ij}e_{ij} + \tilde{y}_{ij} f_{ij},$$

one computes easily

$$s(X,Y) = -2 \sum_{i<j} (x_{ij}y_{ij} + \tilde{x}_{ij}\tilde{y}_{ij})\rho(E_{ii} - E_{jj}); \tag{2.7}$$

(i), (ii), and (iv) follow readily from this formula.

To prove (iii) remark that from (2.3) it follows that

$$\rho(E_{ii}) = e^{-\beta\epsilon_i}/\mathrm{tr}\, e^{-\beta H_\rho}$$

and hence from (2.7),

$$s(X,X) = 2 \sum_{i<j} (x_{ij}^2 + \tilde{x}_{ij}^2) \frac{(e^{-\beta\epsilon_j} - e^{-\beta\epsilon_i})}{\mathrm{tr}\, e^{-\beta H_\rho}} > 0, \quad \text{if } X \neq 0.$$

■

III. FLUCTUATIONS AND STABILITY

As usual, the local fluctuations are defined by

$$\tilde{X}^n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \rho(X)), \tag{3.1}$$

for all $X \in M_{\mathrm{sa}}$.

In physics the interesting objects are the macroscopic fluctuations, which are the limits for n tending to infinity of the operators $\tilde{X}^n$ in the equilibrium state of the system given by ω_ρ (2.1) with ρ satisfying (2.3).

In Ref. 3 we proved the following central limit theorem:

$$\lim_{n\to\infty} \omega_\rho(\exp i\tilde{X}^n) = \exp - \tfrac{1}{2}(\rho(X^2) - \rho(X)^2) \tag{3.2}$$

such that we can give a meaning to

$$\lim_{n\to\infty} \tilde{X}^n = B_\rho(X). \tag{3.3}$$

It is shown that $B_\rho(X)$ is a boson field for all $X \in M_{\mathrm{sa}}$, i.e., a linear (unbounded) self-adjoint operator on a well defined Hilbert space $\mathscr{H}_\rho$ satisfying the boson commutation relations

$$[B_\rho(X), B_\rho(Y)] = \rho([X,Y]).$$

In this way we are able to define the creation and annihilation operators of fluctuations by

$$a_\rho^\pm(X) = (1/\sqrt{2})(B_\rho(X) \mp iB_\rho(JX)) \tag{3.4}$$

for all $X \in M_{\mathrm{sa}}^1$, satisfying

$$[a_\rho^-(X), a_\rho^+(Y)] = \rho([X,Y]) - i\rho([X,JY]),$$

$$[a_\rho^-(X), a_\rho^-(Y)] = 0.$$

We considered also locally perturbed equilibrium states by perturbing the Hamiltonian (2.2) by a local fluctuation, i.e., we considered the Hamiltonians for large N:

$$H_N + \tilde{Y}^n = H_N + \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \rho(Y)), \tag{3.5}$$

where H_N is given by (2.2). In fact, we consider this type of perturbation of the original Hamiltonian together with its equilibrium state. Clearly we get a perturbed state for a Hamiltonian that is perturbed by a fluctuation. We will show that the equilibrium state is stable against this type of perturbations. Remark that the limit $N\to\infty$ corresponds to the thermodynamic limit. We are interested in the limits $N\to\infty$, but also in the central limit $n\to\infty$. The order in which the limits should be taken is always first the thermodynamic limit ($N\to\infty$) and then $n\to\infty$.

The thermodynamic equilibrium state ($N\to\infty$) of the perturbed Hamiltonian (3.5) is again a product state[7] of $\mathscr{B}$ denoted by $\omega_\rho^{\tilde{Y}^n}$ and defined by

$$\omega_\rho^{\tilde{Y}^n}(\otimes_j X_j)=\prod_{j=1}^{n}\rho^{Y/\sqrt{n}}(X_j)\prod_{j>n}\rho(X_j),\quad X_j\in M, \tag{3.6}$$

where the density matrix $\rho^{Y/\sqrt{n}}$ satisfies again the gap equation

$$\rho^{Y/\sqrt{n}}=\frac{\exp-\beta(H_\rho+Y/\sqrt{n})}{\operatorname{tr}\exp-\beta(H_\rho+Y/\sqrt{n})}. \tag{3.7}$$

In Ref. 3 we proved also the existence and finiteness of the following limits:

$$\lim_{n\to\infty}\omega_\rho^{\tilde{Y}^n}(\exp i\tilde{X}^n)=\exp[-\tfrac{1}{2}(\rho(X^2)-\rho(X)^2) - i\beta(X-\rho(X),Y-\rho(Y))_\sim], \tag{3.8}$$

for all $X,Y\in M_{sa}$, where $(Z_1,Z_2)_\sim$ is the Duhamel two-point function[8]; for $Z_1,Z_2\in M$

$$(Z_1,Z_2)_\sim=\frac{1}{\beta}\int_0^\beta ds\,\rho(Z_1^*\alpha_{is}(Z_2)). \tag{3.9}$$

One has also the existence and finiteness of the following limits:

$$\lim_{n\to\infty}\omega_\rho^{\tilde{Y}^n}((\tilde{X}^n)^k),\quad k=1,2,3,... \tag{3.10}$$

whose values can be obtained from formula (3.8). In particular for $k=1$,

$$\lim_{n\to\infty}\omega_\rho^{\tilde{Y}^n}(\tilde{X}^n)=-\beta(X-\rho(X),Y-\rho(Y))_\sim.$$

In the following we study the response of this type of perturbation on the thermodynamic functions. As usual we define the local thermodynamic functions on the set of states of $\mathscr{B}$ as follows: let ω be any state of $\mathscr{B}$ and ω_N the restriction of ω to $\otimes_{i=1}^N M_i$, then ω_N is of the form $\omega_N(\cdot)=\operatorname{tr}\sigma_N\cdot$, with σ_N a density matrix on $\mathbb{C}^{N\cdot m}$. Then the energy functional is

$$E_N(\omega)=\operatorname{tr}\sigma_N H_N=\omega(H_N), \tag{3.11}$$

the entropy functional is

$$S_N(\omega)=-\operatorname{tr}\sigma_N\log\sigma_N, \tag{3.12}$$

the free-energy functional is

$$F_N(\omega)=E_N(\omega)-(1/\beta)S_N(\omega),$$

and the relative entropy of ω_1 with respect to ω_2 is

$$S_N(\omega_1|\omega_2)=-\operatorname{tr}\sigma_{1,N}(\log\sigma_{1,N}-\log\sigma_{2,N}). \tag{3.13}$$

About the thermodynamic limits we have the following proposition.

Proposition 3.1:

(i) $\lim_{N\to\infty}(E_N(\omega_\rho^{\tilde{Y}^n})-E_N(\omega_\rho))=\sqrt{n}\,\omega_\rho^{\tilde{Y}^n}(\tilde{H}_\rho^n)$;

(ii) $\lim_{N\to\infty}(S_N(\omega_\rho^{\tilde{Y}^n})-S_N(\omega_\rho)) = S_n(\omega_\rho^{\tilde{Y}^n}|\omega_\rho)+\sqrt{n}\beta\omega_\rho^{\tilde{Y}^n}(\tilde{H}_\rho^n)$;

(iii) $\lim_{N\to\infty}(F_N(\omega_\rho^{\tilde{Y}^n})-F_N(\omega_\rho))=-(1/\beta)S_n(\omega_\rho^{\tilde{Y}^n}|\omega_\rho)$.

Proof: Using the product property of the states and the definition of the effective Hamiltonian (2.4), one computes consecutively,

$$\begin{aligned}
&\lim_{N\to\infty}(\omega_\rho^{\tilde{Y}^n}(H_N)-\omega_\rho(H_N))\\
&=\lim_{N\to\infty}\Bigg(\sum_{i=1}^n(\omega_\rho^{\tilde{Y}^n}-\omega_\rho)(A_i)
+\frac{1}{N}\sum_{i<j=2}^{n}(\omega_\rho^{\tilde{Y}^n}-\omega_\rho)(B_{ij})
+\frac{1}{N}\sum_{i=1}^{n}\sum_{j=n+1}^{N}(\omega_\rho^{\tilde{Y}^n}-\omega_\rho)(B_{ij})\Bigg)\\
&=n(\rho^{Y/\sqrt{n}}-\rho)(H_\rho)=\sqrt{n}\,\omega_\rho^{\tilde{Y}^n}(\tilde{H}_\rho^n)
\end{aligned}$$

proving (i). Furthermore for $N\geqslant n$

$$\begin{aligned}
&S_N(\omega_\rho^{\tilde{Y}^n})-S_N(\omega_\rho)\\
&=S_n(\omega_\rho^{\tilde{Y}^n})-S_n(\omega_\rho)\\
&=n(-\operatorname{tr}\rho^{Y/\sqrt{n}}(\log\rho^{Y/\sqrt{n}}-\log\rho)\\
&\quad-\operatorname{tr}(\rho^{Y/\sqrt{n}}-\rho)\log\rho)\\
&=S_n(\omega_\rho^{\tilde{Y}^n}|\omega_\rho)+\beta\sqrt{n}\,\omega_\rho^{\tilde{Y}^n}(\tilde{H}_\rho^n),
\end{aligned}$$

where we used (2.3) in the last step, proving (ii). Now (iii) follows from (i) and (ii). ∎

Remark that as far as the thermodynamic limits are concerned the densities remain unchanged under the perturbation, i.e.,

$$\lim_{N\to\infty}\left(\frac{E_N(\omega_\rho^{\tilde{Y}^n})}{N}-\frac{E_N(\omega_\rho)}{N}\right)=0,$$

and analogously for the entropy and free-energy densities.

Remark also that the limits $N\to\infty$ of E_N, S_N, and F_N are all infinite for the state ω_ρ as well as for the perturbed states $\omega_\rho^{\tilde{Y}}$, for all $Y\in M_{sa}$.

In Proposition 3.1 we compute the differences of the energy, entropy, and free-energy functionals and observe that these are the thermodynamic parameters distinguishing the equilibrium state ω_ρ and the perturbed states $\omega_\rho^{\tilde{Y}^n}$. Remark also that the free-energy difference is nothing but the relative entropy.

Naturally we are now interested in the limit $n\to\infty$, i.e., we are interested in the response on the thermodynamics due to a macroscopic fluctuation as perturbation of the equilibrium state.

Clearly it follows from (3.10) and Proposition 3.1 that

$$\lim_{N\to\infty}(E_N(\omega_\rho^{\tilde{Y}^n})-E_N(\omega_\rho))\approx\sqrt{n},$$

for large n. The behavior for $n\to\infty$ of the entropy and free-energy functionals are studied in the following theorem.

Theorem 3.2: For all $Y\in M_{sa}$,

$$\lim_{n\to\infty}\lim_{N\to\infty}(F_N(\omega_\rho^{\tilde{Y}^n})-F_N(\omega_\rho))$$
$$=\lim_n -(1/\beta)S_n(\omega_\rho^{\tilde{Y}^n}|\omega_\rho)$$
$$=\tfrac{1}{2}\beta(Y-\rho(Y),Y-\rho(Y))_\sim\geqslant 0.$$

The equality sign is attained if and only if

$$Y=\rho(Y),\quad \text{i.e., } \omega_\rho^{\tilde{Y}^n}=\omega_\rho.$$

Proof: Using the product property of the states and (2.3) and (3.7), one gets

$$S_n(\omega_\rho^{\tilde{Y}^n}|\omega_\rho)$$
$$=-\operatorname{tr}\otimes_{i=1}^n\rho^{Y/\sqrt{n}}(\log\otimes_{i=1}^n\rho^{Y/\sqrt{n}}-\log\otimes_{i=1}^n\rho)$$
$$=-n\operatorname{tr}\rho^{Y/\sqrt{n}}(\log\rho^{Y/\sqrt{n}}-\log\rho)$$
$$=\beta\sqrt{n}\operatorname{tr}\rho^{Y/\sqrt{n}}Y+n\log[Z(Y/\sqrt{n})/Z(0)],$$

where

$$Z(Y/\sqrt{n})=\operatorname{tr}e^{-\beta(H_\rho+Y/\sqrt{n})}.$$

Remark that the map

$$\mu\in\mathbb{R}\to Z(\mu X)$$

is analytic for all $X\in M_{sa}$. Hence

$$\frac{Z(Y/\sqrt{n})}{Z(0)}=1-\frac{\beta}{\sqrt{n}}\rho(Y)+\frac{\beta^2}{2n}\int_0^1 ds\,\rho(Y\alpha_{i\beta s}(Y))+O\Big(\frac{1}{n^{3/2}}\Big),$$

where $O(1/n^{3/2})$ is bounded by $c/n^{3/2}$ with $c\in\mathbb{R}^+$. Also

$$\beta\sqrt{n}\operatorname{tr}\rho^{Y/\sqrt{n}}Y=\beta\omega_\rho^{\tilde{Y}^n}(\tilde{Y}^n)+\sqrt{n}\beta\rho(Y)$$
$$=\beta\omega_\rho^{\tilde{Y}^n}(\tilde{Y}^n)+n\log\Big(1+\frac{\beta}{\sqrt{n}}\rho(Y)+\frac{\beta^2}{2n}\rho(Y)^2+O\Big(\frac{1}{n^{3/2}}\Big)\Big).$$

Therefore we get

$$S_n(\omega_\rho^{\tilde{Y}^n}|\omega_\rho)$$
$$=\beta\omega_\rho^{\tilde{Y}^n}(\tilde{Y}^n)+n\log\Big[\Big(1+\frac{\beta}{\sqrt{n}}\rho(Y)+\frac{\beta^2}{2n}\rho(Y)^2+O\Big(\frac{1}{n^{3/2}}\Big)\Big)\Big(1-\frac{\beta}{\sqrt{n}}\rho(Y)+\frac{\beta^2}{2n}\int_0^1 ds\,\rho(Y\alpha_{i\beta s}(Y))+O\Big(\frac{1}{n^{3/2}}\Big)\Big)\Big]$$
$$=\beta\omega_\rho^{\tilde{Y}^n}(\tilde{Y}^n)+\log\Big(1+\frac{\beta^2}{2n}(Y-\rho(Y),Y-\rho(Y))_\sim+O\Big(\frac{1}{n^{3/2}}\Big)\Big)^n,$$

where we used formula (3.9).

Now we are able to take the limit n tending to infinity. For the first term we use the result of formula (3.10) with $k=1$, for the second term we use

$$\lim_{n\to\infty}(1+x/n)^n=e^x$$

and get the finite result

$$\lim_n S_n(\omega^{\tilde{Y}^n}|\omega_\rho)=-(\beta^2/2)(Y-\rho(Y),Y-\rho(Y))_\sim.$$

The expression for the free-energy difference functional follows then from Proposition 3.1. It is well known[8] that the Duhamel two-point function is positive definite, i.e.,

$$(X,X)_\sim\geqslant 0,\quad \text{for all } X\in M,$$

and $(X,X)_\sim=0$ implies $X=0$. This proves the last statement of the theorem. ■

From this theorem it follows that, like the energy difference functional, also the entropy difference functional behaves like $\sqrt{n}$, but that the free-energy difference functional or the relative entropy remains finite for n tending to infinity. This shows that the latter one is the relevant thermodynamic parameter to measure the effect of the macroscopic fluctuations. This relative entropy functional is a quadratic expression in the perturbation and attains its extremum at the equilibrium state, which is a strict extremum. This type of stability of equilibrium states is situated between the notion of dynamic stability,[9,10] i.e., perturbing with a strictly local observable, and the notion of global thermodynamic stability,[11] i.e., perturbing with an extensive observable, proportional to n. In this context our perturbation is proportional to $\sqrt{n}$.

IV. SMALL OSCILLATIONS AROUND EQUILIBRIUM

Finally in this section we make some remarks about the time evolution of the fluctuations of the weakly perturbed states of the type $\omega_\rho^{\tilde{Y}^n}$ considered above.

In Sec. II, we introduced the time evolution α_t on M (2.5), induced from the Hamiltonian (2.2) and explicitly given by

$$\alpha_t(X)=e^{itH_\rho}Xe^{-itH_\rho},\quad X\in M.$$

We defined also the algebra of macroscopic fluctuations in formula (3.3). Clearly the dynamics α_t induces a dynamics $\tilde{\alpha}_t$ on the fluctuations by the formula

$$\tilde{\alpha}_tB_\rho(X)=B_\rho(\alpha_tX),\quad X\in M_{sa}. \tag{4.1}$$

It is immediate from this definition that if $X\in M$ is a constant of the motion α_t, then the corresponding fluctuation $B_\rho(X)$ is also a constant of the motion $\tilde{\alpha}_t$. Therefore we restrict ourselves from now on to the set M_{sa}^1, excluding the constants of the motion. Now we diagonalize the dynamics $\tilde{\alpha}_t$. Denote

$$\hat{e}_{kl}=\frac{e_{kl}}{s_\rho(e_{kl},e_{kl})^{1/2}},\quad \hat{f}_{kl}=\frac{f_{kl}}{s_\rho(e_{kl},e_{kl})^{1/2}},$$

i.e., $\{\hat{e}_{kl},\hat{f}_{kl}|k<l\}$ is an orthonormal basis with respect to the scalar product s_ρ for M_{sa}^1 introduced in Sec. II, then the following theorem can be shown.

Theorem 4.1: The creation and annihilation operators $a_\rho^\pm(\hat{e}_{kl})$ are eigenvectors of the evolution $\tilde{\alpha}_t$ and

$$\tilde{\alpha}_ta_\rho^\pm(\hat{e}_{kl})=e^{\pm i(\epsilon_k-\epsilon_l)t}a_\rho^\pm(\hat{e}_{kl}).$$

Moreover, $\tilde{\alpha}_t$ is implemented by the Hamiltonian

$$\hat{H}_\rho = \sum_{k<l} (\epsilon_k - \epsilon_l) a_\rho^+ (\hat{e}_{kl}) a_\rho^- (\hat{e}_{kl}),$$

i.e., $\tilde{\alpha}_t(\cdot) = e^{it\hat{H}_\rho} \cdot e^{it\hat{H}_\rho}$.

Proof: The proof is based on a straightforward computation taking into account the following easily checkable properties:

$$a_\rho^\pm (JX) = \pm i a_\rho^\pm (X), \quad X \in M_{\rm sa}^1,$$
$$\alpha_t J(X) = J\alpha_t(X).$$

One has immediately that

$$\tilde{\alpha}_t a_\rho^\pm (X) = a_\rho^\pm (\alpha_t X).$$

The result follows from

$$\alpha_t \hat{e}_{ij} = \hat{e}_{ij} \cos(\epsilon_i - \epsilon_j)t + J\hat{e}_{ij} \sin(\epsilon_i - \epsilon_j)t$$

and the definitions (3.4). The last statement follows immediately from the commutation relation:

$$[a_\rho^- (\hat{e}_{ij}), a_\rho^+ (\hat{e}_{kl})] = \delta_{ik}\delta_{jl}$$
$$= 0. \qquad \blacksquare$$

Remark that the spectrum of the time evolution $\tilde{\alpha}_t$ of the macroscopic fluctuations is given by $\{n\, \text{spectr}[H_\rho, \cdot\,] | n \in \mathbb{Z}\}$ and coincides here with the spectrum of the original physical Hamiltonian (2.2) in the equilibrium state. For more complicated systems one has to analyze again the basic definition relation (4.1) of $\tilde{\alpha}_t$ in terms of the original α_t. Clearly $\tilde{\alpha}_t$ turns out to be a second-quantized form of α_t. Therefore this theorem might provide an interesting method for the study of the spectrum in equilibrium via the study of the fluctuations. We will work out this point for a variety of models at a later occasion.

Now we look for the connection of the spectrum of $\tilde{\alpha}_t$ of the fluctuations with the explicit form of the thermodynamic potential or relative entropy.

Consider again the decomposition, for each $Y \in M_{\rm sa}^1$,

$$Y = \sum_{k<l} y_{kl} \hat{e}_{kl} + \tilde{y}_{kl} \hat{f}_{kl},$$

then we have the following theorem.

Theorem 4.2:

$$\lim_{n\to\infty} -(1/\beta) S_n (\omega_\rho^{\widetilde{\alpha_t Y^n}} | \omega_\rho) = \lim_{n\to\infty} -(1/\beta) S_n (\omega_\rho^{\tilde{Y}^n} | \omega_\rho)$$
$$= (\beta/2)(Y,Y)_\sim$$
$$= \frac{1}{2} \sum_{k<l} (y_{kl}^2 + \tilde{y}_{kl}^2) \frac{1}{\epsilon_k - \epsilon_l}.$$

Proof: The time invariance, i.e., the first equality, follows from Theorem 3.2 and

$$(\alpha_t X, \alpha_t X)_\sim = (X,X)_\sim, \quad X \in M.$$

The second equality follows from the time invariance of the state ρ implying

$$\rho(Y) = 0, \quad \text{if } Y \in M_{\rm sa}^1.$$

The last equality is obtained by explicit computation, using

$$\frac{\beta}{2}(Y,Y)_\sim = \frac{\beta}{2} \sum_{k<l} (y_{kl}^2 + \tilde{y}_{kl}^2)(\hat{e}_{kl}, \hat{e}_{kl})_\sim$$

and

$$(\hat{e}_{ij}, \hat{e}_{ij})_\sim = 1/\beta(\epsilon_i - \epsilon_j). \qquad \blacksquare$$

Remark that although the thermodynamic potential is time invariant, the expectation values of the fluctuations in the perturbed states are not, e.g., one has for $X, Y \in M_{\rm sa}^1$,

$$\lim_{n\to\infty} \omega_\rho^{\tilde{Y}^n}((\widetilde{\alpha_t X})^n) = \lim_{n\to\infty} \omega_\rho^{\widetilde{\alpha_{-t} Y^n}}(\tilde{X}^n)$$
$$= -\beta(X, \alpha_{-t} Y)_\sim$$

and also

$$i\frac{d}{dt}(X, \alpha_t Y)_\sim = -\rho([X, \alpha_t Y]).$$

Finally, observe that the thermodynamic potential is proportional to the inverse of the spectrum of the original Hamiltonian.

[1]O. E. Lanford and D. Ruelle, J. Math. Phys. **13**, 194 (1969).
[2]S. R. S. Varadhan, Commun. Pure Appl. Math. **19**, 261 (1966).
[3]D. Goderis, A. Verbeure, and P. Vets, "Non-commutative central limits," preprint KUL-TF-88/4, 1988.
[4]K. Hepp and E. H. Lieb, Ann. Phys. (NY) **76**, 360 (1973).
[5]K. Hepp and E. H. Lieb, Helv. Phys. Acta **46**, 573 (1974).
[6]W. F. Wreszinski, Helv. Phys. Acta **46**, 844 (1974).
[7]M. Fannes, H. Spohn, and A. Verbeure, J. Math. Phys. **21**, 355 (1980).
[8]J. Naudts, A. Verbeure, and R. Weder, Commun. Math. Phys. **44**, 87 (1975).
[9]O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, Berlin, 1981), Vol. II.
[10]R. Haag, E. Kastler, and E. B. Trych-Pohlmeyer, Commun. Math. Phys. **38**, 173 (1974).
[11]M. Fannes and A. Verbeure, J. Math. Phys. **19**, 558 (1978).

Evolution of SU(2) and SU(1,1) states: A further mathematical analysis

G. Dattoli, M. Richetta,[a] and A. Torre
Dip. TIB, U.S. Fisica Applicata, Centro Ricerche Energia Frascati, C.P. 65-00044 Frascati, Rome, Italy

(Received 16 June 1987; accepted for publication 17 August 1988)

The problem of the evolution of SU(2) and SU(1,1) states is analyzed from a unified point of view. A generalized Rabi matrix for a time-dependent SU(2)–SU(1,1) coherence preserving Hamiltonian is also derived.

I. INTRODUCTION

The evolution of SU(2) and SU(1,1) coherent states has been recently discussed within the framework of the theory of reduced quantum fluctuations.[1]

Two of the present authors and a co-worker[2] and Aravind[3] showed that both SU(2) and SU(1,1) dynamics can be modeled using a simple vector representation. The SU(2)-state dynamics can be understood as a rotation in an Euclidean space. Correspondingly the SU(1,1) states evolve according to an analogous rotation in a non-Euclidean space.

In both cases the equations of motion, in the Heisenberg representation, can be cast in the form of generalized Bloch-type equations.[4]

When the system is ruled by a time-independent Hamiltonian, linear combination of the SU(2) or SU(1,1) group generators, the dynamics of the Bloch vector, recently introduced also for the SU(1,1) case, can be analytically studied. If, otherwise, the Hamiltonian is a time-dependent linear combination of the generators, analytical solutions are available in a restricted number of cases only.[5,6]

In Ref. 7 algebraic methods of the Wei–Norman type[8] have been exploited to solve Schrödinger equations with time-dependent SU(2) and SU(1,1) coherence preserving Hamiltonians. It has been proved that the characteristic equations of the time-ordering procedure also can be cast in the form of a generalized torque equation.[6] This fact may not seem particularly surprising if one realizes that the average values of SU(2) or SU(1,1) generators are related to the Wei–Norman ordering functions. However, although apparently trivial, this result may be a useful guide to generalize the time-ordering procedure to generic Hamiltonian linear combinations of the generators of a Lie algebra.

We have extended the method to SU(3) coherence preserving Hamiltonians.[9] In that context it has been shown that the Wei–Norman characteristic equations can be cast in the form of a torque equation in an eight-dimensional space.[10]

In this paper we will complete the mathematical analysis of the evolution of the SU(2) and SU(1,1) states proving that a Rabi matrix can be naturally obtained even for the time-dependent case.

In fact, so far the Rabi matrix, i.e., the rotation matrix, which specializes the time evolution of the SU(2) (see Ref. 4) and SU(1,1) Bloch vector (see Refs. 2 and 3), has been explicitly derived only for time-independent Hamiltonians.

The paper is organized as follows. In Sec. II we derive the *spinorial* representation of the evolution operator, while Sec. III is devoted to obtaining the representation of the evolution operator in the SO(3) and SO(2,1) spaces, which is understood as the Rabi matrix relevant to the SU(2) and SU(1,1) Bloch vectors, respectively.

[a] ENEA student.

II. PAULI REPRESENTATION OF THE EVOLUTION OPERATOR

In this section we will discuss the Wei–Norman time-ordering method utilizing the Pauli matrix representation of the SU(2) and SU(1,1) group generators.

Since we will treat the problem from a unified point of view, we will not refer to the SU(2) or SU(1,1) group separately but to the real split three-dimensional Lie algebra with generators $\hat{F}_0$, $\hat{F}_+$, $\hat{F}_-$ obeying the commutation relations

$$[\hat{F}_0,\hat{F}_\pm] = \pm 2\hat{F}_\pm, \quad [\hat{F}_+,\hat{F}_-] = +\delta\hat{F}_0. \tag{2.1}$$

It can be easily checked that the SU(2) and SU(1,1) groups may be recovered with $\delta = 1$ and $\delta = -1$, respectively.

The Pauli matrices for the operators $\hat{F}_0$, $\hat{F}_\pm$ are readily written down as

$$\hat{F}_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \hat{F}_+ = \begin{pmatrix} 0 & \delta \\ 0 & 0 \end{pmatrix}, \quad \hat{F}_- = \begin{pmatrix} 0 & 0 \\ +1 & 0 \end{pmatrix}. \tag{2.2}$$

Let us now consider the Hamiltonian operator

$$\hat{H} = [\omega(t)/2]\,\hat{F}_0 + \Omega^*(t)\hat{F}_+ + \Omega(t)\hat{F}_-, \tag{2.3}$$

where the nonsingular time-dependent functions $\omega(t)$ and $\Omega(t)$ are real and complex, respectively. It is worth stressing that the operator $\hat{H}$ is *Hermitian* according to

$$\hat{H}\hat{M} = \hat{M}\hat{H}^\dagger, \tag{2.4}$$

where $\hat{H}^\dagger$ is the adjoint of $\hat{H}$ (transpose and complex conjugate) and $\hat{M}$ is the metric matrix

$$\hat{M} = \begin{pmatrix} 1 & 0 \\ 0 & \delta \end{pmatrix} \tag{2.5}$$

relevant to SU(2) or SU(1,1) according to the δ value.

According to the Wei–Norman technique the time-evolution operator can be written as

$$\hat{U}(t) = \exp\{h(t)\hat{F}_0\}\exp\{g(t)\hat{F}_+\}\exp\{-f(t)\hat{F}_-\}. \tag{2.6}$$

(The minus sign in the last exponential has been inserted only to make the comparison with previous papers, where the operator $\hat{F}'_- = -\hat{F}_-$ was considered, easier.) The matrix representation of the operator (2.6) can be easily obtained from (2.2) as

 0022-2488/88/122586-03$02.50

$$\hat{U}=\begin{pmatrix} e^h & 0 \\ 0 & e^{-h}\end{pmatrix}\begin{pmatrix} 1 & \delta g \\ 0 & 1\end{pmatrix}\begin{pmatrix} 1 & 0 \\ -f & 1\end{pmatrix} = \begin{pmatrix} (1-\delta fg)e^h & \delta g e^h \\ -fe^{-h} & e^{-h}\end{pmatrix}. \quad (2.7)$$

It is convenient to introduce the functions[6]

$$\mathscr{H}=e^{-h},\quad \mathscr{F}=fe^{-h},\quad \mathscr{G}=ge^h, \quad (2.8)$$

defined by the system of differential equations

$$\begin{aligned} &\dot{\mathscr{H}}=(i\omega/2)\mathscr{H}-i\delta\Omega\mathscr{G},\\ &\dot{\mathscr{G}}=-(i\omega/2)\mathscr{G}-i\Omega^*\mathscr{H},\\ &\dot{\mathscr{F}}\mathscr{H}-\dot{\mathscr{H}}\mathscr{F}=i\Omega,\\ &\mathscr{H}(0)=1,\quad \mathscr{F}(0)=\mathscr{G}(0)=0, \end{aligned} \quad (2.9)$$

inferred from the Schrödinger equation

$$\frac{i\,d\hat{U}}{dt}=\hat{H}\hat{U} \quad (2.10)$$

and from the representation (2.4) (see Ref. 6 for further details). It is a simple matter to prove that Eqs. (2.9) imply the relation

$$\mathscr{G}=\mathscr{F}^* \quad (2.11)$$

and the existence of the invariant

$$|\mathscr{H}|^2+\delta|\mathscr{F}|^2=1. \quad (2.12)$$

Consequently the expression (2.7) for $\hat{U}$ can be recast in the more compact form

$$\hat{U}=\begin{pmatrix} \mathscr{H}^* & \delta\mathscr{F}^* \\ -\mathscr{F} & \mathscr{H}\end{pmatrix}, \quad (2.13)$$

readily recognized as an element of SU(2) or SU(1,1) owing to the relation (2.12), which ensures the *unitarity* of $\hat{U}$ according to the definition

$$\hat{U}\hat{M}\hat{U}^\dagger=\hat{M}, \quad (2.14)$$

with $\hat{U}^\dagger$ being the adjont of $\hat{U}$.

III. SO(3)–SO(2,1) REALIZATION OF THE EVOLUTION OPERATOR

We are mainly interested in understanding the evolution of states ruled by the Hamiltonian (2.3) in terms of a generalized rotation. It is therefore convenient to resort to the isomorphisms between SU(2) and SO(3), and SU(1,1) and SO(2,1).

The $\hat{F}$ generators are represented by the 3×3 matrices

$$\hat{F}_0=\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0\end{pmatrix},$$
$$\hat{F}_+=\begin{pmatrix} 0 & 0 & -\delta \\ 0 & 0 & -i\delta \\ 1 & i & 0\end{pmatrix},\quad \hat{F}_-=\begin{pmatrix} 0 & 0 & \delta \\ 0 & 0 & -i\delta \\ -1 & i & 0\end{pmatrix}, \quad (3.1)$$

in the SO(3) or SO(2,1) space, according to the value of δ. Let us notice that the above representations ensure the *Hermiticity* of $\hat{H}$ as stated by (2.4), the metric matrix being now

$$\hat{M}=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \delta\end{pmatrix}. \quad (3.2)$$

Consequently the exponential operators entering the expression of $\hat{U}$ [Eq. (2.4)] take the matrix form

$$e^{h\hat{F}_0}=\begin{pmatrix} \cosh 2h & -i\sinh 2h & 0 \\ i\sinh 2h & \cosh 2h & 0 \\ 0 & 0 & 1\end{pmatrix},$$
$$e^{g\hat{F}_+}=\begin{pmatrix} 1-(\delta/2)g^2 & -i(\delta/2)g^2 & -\delta g \\ -i(\delta/2)g^2 & 1+(\delta/2)g^2 & -i\delta g \\ g & ig & 1\end{pmatrix}, \quad (3.3)$$
$$e^{-f\hat{F}_-}=\begin{pmatrix} 1-(\delta/2)f^2 & i(\delta/2)f^2 & -\delta f \\ i(\delta/2)f^2 & 1+(\delta/2)f^2 & i\delta f \\ f & -if & 1\end{pmatrix}.$$

Therefore using the relations (2.11) and (2.12) we can finally express the evolution operator as

$$\hat{U}=\begin{pmatrix} \mathrm{Re}(\mathscr{H}^{*2}-\delta\mathscr{F}^{*2}) & \mathrm{Im}(\mathscr{H}^{*2}+\delta\mathscr{F}^{*2}) & -2\delta\,\mathrm{Re}\,\mathscr{H}^*\mathscr{F}^* \\ \sqrt{\mathrm{Im}}(\mathscr{H}^{*2}-\delta\mathscr{F}^{*2}) & \mathrm{Re}(\mathscr{H}^{*2}+\delta\mathscr{F}^{*2}) & 2\delta\,\mathrm{Im}\,\mathscr{H}^*\mathscr{F}^* \\ 2\,\mathrm{Re}\,\mathscr{H}\mathscr{F}^* & -2\,\mathrm{Im}\,\mathscr{H}\mathscr{F}^* & |\mathscr{H}|^2-\delta|\mathscr{F}|^2\end{pmatrix}. \quad (3.4)$$

(It is worth pointing out that the equations for the functions h, f, and g or $\mathscr{H}$, $\mathscr{F}$, and $\mathscr{G}$ hold unchanged, being determined by the algebraic structure of the group involved and not by the dimensionality of the chosen representation.)

The unitarity of $\hat{U}$, according to (2.14), can be easily proved.

The above matrix is just the Rabi matrix for the more general time-dependent problem.

The wave function $\Psi(t)$ of the system represented by the column vector

$$\Psi(t)=\begin{pmatrix} a(t) \\ b(t) \\ c(t)\end{pmatrix} \quad (3.5)$$

is related to the function of the initial time $t_0=0$ by

$$\Psi(t)=\begin{pmatrix} a(t) \\ b(t) \\ c(t)\end{pmatrix}=\hat{U}(t)\begin{pmatrix} a(0) \\ b(0) \\ c(0)\end{pmatrix}. \quad (3.6)$$

It can be easily checked that the matrix (3.4) reduces to the rotation matrix of Refs. 2–4 for a time-independent Hamiltonian.

In fact, for Ω and ω constant the system (2.9) can be analytically solved, thus yielding

$$\begin{aligned} &\mathscr{H}=\cos\frac{Qt}{2}+i\frac{\omega}{Q}\sin\frac{Qt}{2},\\ &\mathscr{F}=i\Omega(\sin Qt/2)/(Q/2), \end{aligned} \quad (3.7)$$

with

$$Q = \sqrt{\omega^2 + 4\delta|\Omega|^2}\,. \tag{3.8}$$

In particular, assuming Ω to be real, the operator $\hat{U}$ specializes into

$$\hat{U} = \begin{pmatrix} [\delta 4\Omega^2 + \omega^2 \cos(Qt)]/Q^2 & -(\omega/Q)\sin(Qt) & (4\delta\omega\Omega/Q^2)\sin^2(Qt/2) \\ (\omega/Q)\sin(Qt) & \cos(Qt) & -(2\delta\Omega/Q)\sin(Qt) \\ (4\omega\Omega/Q^2)\sin(Qt/2) & (2\Omega/Q)\sin(Qt) & [\omega^2 + \delta 4\Omega^2 \cos(Qt)]/Q^2 \end{pmatrix}, \tag{3.9}$$

which can be immediately recognized as the Rabi matrix derived in Refs. 2–4 from a purely geometrical point of view.

The above result, in some sense, concludes the present mathematical analysis of the SU(2)- and SU(1,1)-state evolutions.

We have proved that in both cases the evolution can be visualized as a rotation defined in a suitable vector space.

Furthermore the relevant Rabi matrix has been directly deduced from the ordered form of the evolution operator and it has been expressed in terms of the Wei–Norman characteristic functions.

As a conclusive remark, let us stress that the above method can be generalized to derive the Rabi matrix for time-dependent Hamiltonian linear combinations of the SU(3) group generators. In that case, indeed, we get an 8×8 matrix, whose elements are expressed in terms of the functions characteristic of the ordering procedure,[9] although in a very intriguing form, as a consequence of the higher dimensionality of the group involved.[11]

[1]K. Wodkiewicz and J. H. Eberly, J. Opt. Soc. Am. B **2**, 458 (1985).

[2]G. Dattoli, A. Dipace, and A. Torre, Phys. Rev. A **33**, 4387 (1986).

[3]P. K. Aravind, J. Opt. Soc. Am. B **5**, 1595 (1988).

[4]L. Allen and J. H. Eberly, *Optical Resonance and Two Level Atoms* (Wiley, New York, 1975).

[5]A. Bambini and P. Berman, Phys. Rev. A **23**, 2496 (1981).

[6]G. Dattoli and A. Torre, J. Math. Phys. **28**, 618 (1987).

[7]See, e.g., F. Ciocci, G. Dattoli, A. Renieri, and A. Torre, Phys. Rep. **141**, 1 (1986).

[8]J. Wei and E. Norman, J. Math. Phys. **4**, 575 (1963).

[9]G. Dattoli, P. Di Lazzaro, and A. Torre, Phys. Rev. A **35**, 1582 (1987).

[10]See also J. N. Elgin, Phys. Lett. A **80**, 140 (1980); F. T. Hioe and J. H. Eberly, Phys. Rev. Lett. **47**, 832 (1981).

[11]G. Dattoli and A. Torre, to be published.

Finite-dimensional representations of the special linear Lie superalgebra sl(1,*n*). II. Nontypical representations

Tchavdar D. Palev
Institute for Nuclear Research and Nuclear Energy, 1184 Sofia, Bulgaria

(Received 20 April 1988; accepted for publication 20 July 1988)

All nontypical irreducible representations of the special linear Lie superalgebra sl(1,n) are constructed for any n. Explicit expressions for the transformation of the basis under the action of the generators are given. The results of this paper together with those obtained in Paper I [J. Math. Phys. **28**, 2280 (1987)] solve the problem of the finite-dimensional irreducible representations of sl(1,n).

I. INTRODUCTION

In Ref. 1 (hereafter referred to as I) we gave explicit expressions for all typical representations of the basic Lie superalgebra (LS) sl(1,n) [$=A(0,n-1)$ in the notation of Ref. 2] for any $n=2,3,\ldots$. In the present paper we solve the same problem for the nontypical representations. Throughout we use the abbreviations, the notation, and the terminology introduced in I (see, especially, Sec. II A). Here we briefly recall only some main points from that paper.

We consider sl(1,n) as a subalgebra of the general linear Lie superalgebra gl(1,n). The latter consists of all squared $(n+1)$-dimensional matrices, whose rows and columns we label with indices $A,B,C,D,\ldots=0,1,2,\ldots,n$. As a basis in gl(1,$n$) we choose all Weyl matrices e_{AB}, $A,B=0,1,\ldots,n$. Assign to each index A a degree (A), which is zero for $A=0$ and 1 for $A=1,\ldots,n$. The generator e_{AB} is even (resp. odd), if $(A)+(B)$ is an even (resp. an odd) number. The multiplication (= the supercommutator) [[,]] on gl(1,n) is given with the linear extension of the relations

$$[\![e_{AB},e_{CD}]\!]=\delta_{BC}e_{AD}-(-1)^{[(A)+(B)][(C)+(D)]}\delta_{AD}e_{CB}. \tag{1.1}$$

The LS sl(1,n) is a subalgebra of gl(1,n), consisting of all those matrices $a\in$gl(1,n), whose supertrace (= str) vanishes, i.e.,

$$\mathrm{sl}(1,n)=\left\{a\ \middle|\ a\in \mathrm{gl}(1,n),\ \mathrm{str}(a)\equiv\sum_{A=0}^{n}(-1)^{(A)}a_{AA}=0\right\}. \tag{1.2}$$

The even subalgebra

$$\mathrm{sl}(1,n)_0=\text{lin. env.}\{E_{ij}\,|\,E_{ij}=e_{ij}+\delta_{ij}e_{00},\ i,j=1,\ldots,n\} \tag{1.3}$$

is isomorphic to the general linear Lie algebra gl(n) and E_{ij} are its Weyl generators. As an ordered basis in the Cartan subalgebra of sl(1,n) and gl(n) we choose $E_{11},E_{22},\ldots,E_{nn}$ and denote by $E^1,\ldots,E^n$ the dual to its basis in H^*.

To give (one possible) definition of a nontypical representation and a nontypical module, we recall (see I) the structure of the sl(1,n) module $\overline{W}([m]_{n+1})$, induced from the gl(n) fidirmod (= finite-dimensional irreducible module) $V_0([m]_{n+1})$. By

$$[m]_{n+1}\equiv[m_{1,n+1},m_{2,n+1},\ldots,m_{n,n+1}] \tag{1.4}$$

we denote the coordinates of the gl(n) highest weight

$$\Lambda=\sum_{i=1}^{n}m_{i,n+1}E^i \tag{1.5}$$

(or the highest weight itself), corresponding to the highest weight vector from $V_0([m]_{n+1})$, where

$$m_{i,n+1}-m_{j,n+1}\in\mathbb{Z}_+,\quad \forall i<j=1,2,\ldots,n. \tag{1.6}$$

Denote by P_+ the linear envelope of all odd positive root vectors of sl(1,n),

$$P_+=\text{lin. env.}\{e_{0i}\,|\,i=1,\ldots,n\}. \tag{1.7}$$

Let $P=\mathrm{gl}(n)\oplus P_+$. To turn $V_0([m]_{n+1})$ into a P module, we set

$$P_+V_0([m]_{n+1})=0. \tag{1.8}$$

Then $\overline{W}([m]_{n+1})$ is defined to be the tensor product of the sl(1,n) universal enveloping algebra U with $V_0([m]_{n+1})$ factorized by the subspace

$$I([m]_{n+1})=\text{lin. env.}\{up\otimes v-u\otimes pv\,|\,u\in U,\ p\in P,\ v\in V_0([m]_{n+1})\}. \tag{1.9}$$

The linear space is turned into an sl(1,n) module in a natural way:

$$g(u\otimes v)=gu\otimes v,\quad g\in\mathrm{sl}(1,n),\quad u\otimes v(\in\overline{W}([m])_{n+1}). \tag{1.10}$$

Thus to every gl(n) fidirmod $V_0([m]_{n+1})$ there corresponds an induced sl(1,n) module $\overline{W}([m]_{n+1})$. Both of them have the same highest weight Λ [see (1.5)]. Every induced module $\overline{W}([m]_{n+1})$ is either irreducible [i.e., it is an sl(1,n) fidirmod] or indecomposible. The representation of sl(1,n), realized in the irreducible $\overline{W}([m]_{n+1})$ (and also the module itself), is said to be typical.[3] Each $\overline{W}([m]_{n+1})$, which is not irreducible, contains a maximal sl(1,n) invariant subspace $\overline{I}([m]_{n+1})\neq 0$. The factor module $\overline{W}([m]_{n+1})/\overline{I}([m]_{n+1})$ carries an irreducible representation of sl(1,n).

Definition (Ref. 3): If $\overline{I}([m]_{n+1})$ is a nontrivial subspace, then the representation of sl(1,n) in $\overline{W}([m]_{n+1})/\overline{I}([m]_{n+1})$ is said to be nontypical. Any irreducible module, carrying a nontypical representation, is also said to be nontypical.

 0022-2488/88/122589-10$02.50

Proposition 1 (Ref. 4): The induced sl(1,n) module $\overline{W}([m]_{n+1})$ is nontypical if and only if for a certain $k=1,2,...,n$, $m_{k,n+1}=k-1$.

II. TRANSFORMATION OF THE INDUCED MODULES (Ref. 1)

Let

$$(m)\equiv\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ \vdots\\ [m]_i\\ \vdots\\ m_{11}\end{bmatrix}\equiv\begin{bmatrix}m_{1,n+1},m_{2,n+1},...,m_{n,n+1}\\ m_{1n},m_{2n},...,m_{nn}\\ \cdots\cdots\cdots\\ m_{1i},m_{2i},...,m_{ii}\\ \cdots\cdots\cdots\\ m_{11}\end{bmatrix} \tag{2.1}$$

be a pattern of complex numbers.

Proposition 2: The basis $\overline{\Gamma}([m]_{n+1})$ in the induced sl(1,n) module $\overline{W}([m]_{n+1})$ with a highest weight (1.5) can always be chosen to consist of all those patterns (2.1), for which the following conditions hold ($\mathbb{Z}_+$ = all non-negative integers):

(1) $$m_{in}=m_{i,n+1}+\theta_i-\sum_{k=1}^{n}\theta_k,\quad \theta_1,\theta_2,...,\theta_n=0,1; \tag{2.2}$$

(2) $$m_{i,j+1}-m_{ij}\in\mathbb{Z}_+,\quad m_{ij}-m_{i+1,j+1}\in\mathbb{Z}_+,\quad \forall i\leqslant j=1,...,n-1. \tag{2.3}$$

Each pattern from $\overline{\Gamma}([m]_{n+1})$ is a weight vector. We call this basis an induced basis (I basis) and each pattern (2.1) an I pattern. Denote by $(m)_{\pm ij}$ an I pattern which is obtained from (m) after the replacement $m_{ij}\to m_{ij}\pm 1$. Introduce also the following abbreviations:

$$\begin{aligned}&[m_{1k},m_{2k},...,m_{kk}]=[m]_k,\quad k=1,...,n,\\ &[m_{1k}+c,m_{2k}+c,...,m_{kk}+c]=[m+c]_k,\quad c\in\mathbb{C},\\ &[m_{1k}\pm\delta_{1i},...,m_{kk}\pm\delta_{ki}]=[m]_{\pm i,k},\\ &l_{ij}=m_{ij}-i.\end{aligned} \tag{2.4}$$

Then the transformation of $\overline{W}([m]_{n+1})$ under the action of sl(1,n) is completely defined from the relations [see I, (3.1)–(3.3), (3.118), and (3.119)]

$$E_{kk}(m)=(m_{1k}+\cdots+m_{kk}-m_{1,k-1}-\cdots-m_{k-1,k-1})(m), \tag{2.5}$$

$$E_{k,k-1}(m)=\sum_{j=1}^{k-1}\left|\frac{\Pi_{i=1}^{k}(l_{ik}-l_{j,k-1}+1)\Pi_{i=1}^{k-2}(l_{i,k-2}-l_{j,k-1})}{\Pi_{i\neq j=1}^{k-1}(l_{i,k-1}-l_{j,k-1}+1)(l_{i,k-1}-l_{j,k-1})}\right|^{1/2}(m)_{-j,k-1}, \tag{2.6}$$

$$E_{k-1,k}(m)=\sum_{j=1}^{k-1}\left|\frac{\Pi_{i=1}^{k}(l_{ik}-l_{j,k-1})\Pi_{i=1}^{k-2}(l_{i,k-2}-l_{j,k-1}-1)}{\Pi_{i\neq j=1}^{k-1}(l_{i,k-1}-l_{j,k-1}-1)(l_{i,k-1}-l_{j,k-1})}\right|^{1/2}(m)_{j,k-1}, \tag{2.7}$$

$$e_{n0}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{bmatrix}=\sum_{i=1}^{n}(1-\theta_i)(-1)^{\theta_1+\cdots+\theta_{i-1}}\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in}-1)}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})}\right|^{1/2}\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{bmatrix}, \tag{2.8}$$

$$e_{0n}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{bmatrix}=\sum_{i=1}^{n}\theta_i(-1)^{\theta_1+\cdots+\theta_{i-1}}(l_{i,n+1}+1)\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in})}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})}\right|^{1/2}\begin{bmatrix}[m]_{n+1}\\ [m+1]_{-i,n}\\ [m+1]_{n-1}\\ \vdots\\ m_{11}+1\end{bmatrix}. \tag{2.9}$$

If for any $k=1,...,n$, $m_{k,n+1}\neq k-1$, the above relations describe the transformations of all typical sl(1,n) modules.

III. NONTYPICAL REPRESENTATIONS

If for a certain $k=1,...,n$, $m_{k,n+1}=k-1$, then the induced representation is indecomposible. The corresponding sl(1,n) module $\overline{W}([m]_{n+1})$ contains a maximal invariant subspace $\overline{I}([m]_{n+1})$ and at the same time there exists no complement to $\overline{I}([m]_{n+1})$ subspace, which is sl(1,n) invariant. The factor module $\overline{W}([m]_{n+1})/\overline{I}([m]_{n+1})$ carries an irreducible nontypical representation of sl(1,n). In order to write the formulas (2.5)–(2.9) in the corresponding factor space, we now proceed to determine the maximal invariant subspaces $\overline{I}([m]_{n+1})$.

A. Maximal invariant subspace

Let $V([m]_n)\equiv V([m_{1n},...,m_{nn}])$ be a gl(n) fidirmod with a highest weight $m_{1n}E^1+\cdots+m_{nn}E^n$. Consider $\overline{W}([m]_{n+1})$ as a gl(n) module. The following proposition was proved in I.

Proposition 3: The induced sl(1,n) module $\overline{W}([m]_{n+1})$ decomposes into a direct sum of gl(n) fidirmods as follows:

$$\overline{W}([m]_{n+1})=\sum_{[m]_n}\oplus V([m]_n), \tag{3.1}$$

where the sum is over all possible gl(n) signatures $[m]_n$, which are compatible with (2.2).

Proposition 4: For each term in the decomposition (3.1) either

$$V([m]_n)\subset\bar{I}([m]_{n+1})$$

or

$$V([m]_n)\cap\bar{I}([m]_{n+1})=0. \tag{3.2}$$

Proof: Let $U[\mathrm{gl}(n)]$ be the gl(n) universal enveloping algebra. If $0\neq x\in V([m]_n)\cap\bar{I}([m]_{n+1})$, then also $U[\mathrm{gl}(n)]x\in\bar{I}([m]_{n+1})$ and since $U[\mathrm{gl}(n)]x=V([m]_n)$, (3.2) holds.

Proposition 5: Let $[m-n+1]_{n+1}\equiv[m_{1,n+1}-n+1,\ldots,m_{n,n+1}-n+1]$. Then

$$V([m-n+1]_n)\subset\bar{I}([m]_{n+1}). \tag{3.3}$$

Proof: By construction of $\bar{W}([m]_{n+1})$ the lowest weight vector x_0 of $\bar{W}([m]_{n+1})$ belongs to $V([m-n+1]_n)$ and it can be obtained from any other $x\in\bar{W}([m]_{n+1})$ with a properly chosen element $u\in U$, i.e., $x_0=ux$. Let $0\neq x\in\bar{I}([m]_{n+1})$. Then $ux=x_0\in V([m-n+1]_n)\cap\bar{I}([m]_{n+1})$ and according to Proposition 4, (3.3) holds.

Proposition 6: The maximal invariant subspace has a zero intersection with $V([m]_{n+1})$,

$$V([m]_{n+1})\cap\bar{I}([m]_{n+1})=0. \tag{3.4}$$

Proof: We use that $\bar{W}([m]_{n+1})=U\otimes V_0([m]_{n+1})=U(1\otimes V_0([m]_{n+1}))=UV([m]_{n+1})$. In particular, for any nonzero $x\in V([m]_{n+1})$, $Ux=U\cdot U[\mathrm{gl}(n)]x=UV([m]_{n+1})=\bar{W}([m]_{n+1})$. Therefore, if $0\neq x\in V([m]_{n+1})\cap\bar{I}([m]_{n+1})$, then $Ux=\bar{W}([m]_{n+1})\subset\bar{I}([m]_{n+1})$, which is impossible, since we consider nontrivial invariant subspaces $\bar{I}([m]_n)$. Hence (3.4) holds.

Denote by $W([m]_{n+1})$ any complement to $\bar{I}([m]_{n+1})$ subspace in $\bar{W}([m]_{n+1})$, i.e.,

$$\bar{W}([m]_{n+1})=W([m]_{n+1})\oplus\bar{I}([m]_{n+1}). \tag{3.5}$$

Let $\xi_x\equiv\xi(x)\in\bar{W}([m]_{n+1})/\bar{I}([m]_{n+1})$ be the equivalence class of $x\in\bar{W}([m]_{n+1})$.

Proposition 7: The mapping f: $W([m]_{n+1})\to\bar{W}([m]_{n+1})/\bar{I}([m]_{n+1})$ defined as $f(x)=\xi_x$ is an isomorphism of $W([m]_{n+1})$ on $\bar{W}([m]_{n+1})/\bar{I}([m]_{n+1})$.

Proof: It is evident that f is linear. Suppose that $f(x)=f(y)$. Then $f(x-y)=\xi_{x-y}=0$, i.e., $x-y\in\bar{I}([m]_{n+1})$. On the other hand, $x-y\in W([m]_{n+1})$ and, since $\bar{I}([m]_{n+1})$ and $W([m]_{n+1})$ are linearly independent, $x-y=0$.

Choose

$$e_1,\ldots,e_p \text{ to be a basis in } W([m]_{n+1}), \tag{3.6}$$

$$f_1,\ldots,f_q \text{ to be a basis in } \bar{I}([m]_{n+1}). \tag{3.7}$$

If $\xi\in\bar{W}([m]_{n+1})/\bar{I}([m]_{n+1})$, according to Proposition 7 $\xi=\xi_x$, where $x\in W([m]_{n+1})$. Therefore,

$$\xi=\xi_x=\xi\Big(\sum_{i=1}^{p}\alpha_i e_i\Big)=\sum_{i=1}^{p}\alpha_i\xi_{e_i}.$$

Moreover, $\Sigma_{i=1}^{p}\alpha_i\xi_{e_i}=0$ implies $\Sigma_{i=1}^{p}\alpha_i e_i\in\bar{I}([m]_{n+1})$, which is possible only if all $\alpha_i=0$. This shows that $\xi_{e_1},\ldots,\xi_{e_p}$ constitute a basis in the factor space.

Let g be any element from sl(1,n). Since $\bar{I}([m]_{n+1})$ is sl(1,n) invariant,

$$ge_i=\sum_{j=1}^{p}A_{ji}e_j+\sum_{j=1}^{q}B_{ji}f_j,\quad gf_i=\sum_{j=1}^{q}C_{ji}f_j. \tag{3.8}$$

The transformation of the factor space under sl(1,n) is defined to be $g[\xi(x)]=\xi(gx)$. Therefore,

$$g[\xi(e_i)]=\xi(ge_i)=\xi\Big(\sum_{j=1}^{p}A_{ji}e_j+\sum_{j=1}^{q}b_{ji}f_j\Big)=\sum_{j=1}^{p}A_{ji}\xi(e_j). \tag{3.9}$$

We have used the circumstance that $\xi(f_i)=0$ in the factor space. As usual we shall identify $W([m]_{n+1})$ and $\bar{W}([m]_{n+1})/\bar{I}([m]_{n+1})$ (see Proposition 7), replacing $\xi(x)$ by x. Then the relation (3.9) reads

$$ge_i=\sum_{j=1}^{p}A_{ji}e_j. \tag{3.10}$$

We conclude the following corollary by comparing (3.10) with (3.8).

Corollary: In order to obtain the transformation of the factor space $W([m]_{n+1})\equiv\bar{W}([m]_{n+1})/\bar{I}([m]_{n+1})$ under the action of the sl(1,n) generators one has simply to replace in (3.8) all basis vectors of the maximal invariant subspace $f_1,\ldots,f_q$ by zero.

Consider the indecomposible sl(1,n) module $\bar{W}([m]_{n+1})$, corresponding to

$$m_{j,n+1}=j-1\Leftrightarrow l_{j,n+1}+1=0. \tag{3.11}$$

Since

$$l_{1,n+1}>l_{2,n+1}>\cdots>l_{k,n+1}>\cdots>l_{n,n+1}, \tag{3.12}$$

all other $l_{k,n+1}+1$ are different from zero, i.e.,

$$l_{k,n+1}+1\neq0,\quad\forall k\neq j=1,\ldots,n. \tag{3.13}$$

We recall that the gl(n) fidirmods $V([m]_n)$ of $\bar{W}([m]_{n+1})$ [see the decomposition (3.1)] are in one-to-one correspondence with all admissible θ-tuples $\{\theta_1,\ldots,\theta_n\}\equiv\{\theta\}_n$, i.e., those θ-tuples, for which

(1) $\theta_1,\ldots,\theta_n=0,1$,

(2) $[m]_n$ is lexical $\Leftrightarrow m_{in}-m_{i+1,n}\in\mathbb{Z}_+$,

$$i=1,\ldots,n-1. \tag{3.14}$$

Therefore, we set

$$V([m]_n)\equiv V(\{\theta\}_n)\equiv V(\{\theta_1,\ldots,\theta_n\}) \tag{3.15}$$

and write instead of (3.1)

$$\bar{W}([m]_{n+1})=\sum_{\theta_1,\ldots,\theta_n=0,1}{}'\oplus V(\{\theta\}_n). \tag{3.16}$$

The prime in (3.16) is to recall that the sum is only over the admissible θ-tuples.

Introduce the following two subspaces of $\bar{W}([m]_{n+1})$:

$$\bar{W}([m]_{n+1})_0=\sum_{\substack{\theta_1,\ldots,\theta_n=0,1\\ \theta_j=0}}{}'\oplus V(\{\theta\}_n), \tag{3.17}$$

$$\bar{W}([m]_{n+1})_1=\sum_{\substack{\theta_1,\ldots,\theta_n=0,1\\ \theta_j=1}}{}'\oplus V(\{\theta\}_n). \tag{3.18}$$

Then

$$\bar{W}([m]_{n+1}) = \bar{W}([m]_{n+1})_0 \oplus \bar{W}([m]_{n+1})_1. \qquad (3.19)$$

Proposition 8: The subspace $\bar{W}([m]_{n+1})_1$ is an invariant sl(1,n) submodule of $\bar{W}([m]_{n+1})$ and, therefore, $\bar{W}([m]_{n+1})_1 \subset \bar{I}([m]_{n+1})$.

Proof: Let (m) be an I pattern, corresponding to the θ-tuple $\{\theta_1,...,\theta_n\}$. From (2.2) one derives [see I, (2.61)] that for every $i=1,...,n$,

$$\theta_i = \frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m_{kn}) - (m_{i,n+1}-m_{in}). \qquad (3.20)$$

The set of all I patterns with $\theta_j = 0$ (resp. with $\theta_j = 1$) constitutes a basis $\bar{\Gamma}([m]_{n+1})_0$ in $\bar{W}([m]_{n+1})_0$ [resp. $\bar{\Gamma}([m]_{n+1})_1$ in $\bar{W}([m]_{n+1})_1$]. Let $(m)\in\bar{\Gamma}([m]_{n+1})_1$. Then

$$\theta_j = \frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m_{kn}) - (m_{j,n+1}-m_{jn}) = 1. \qquad (3.21)$$

For this particular I pattern $e_{n0}(m)$ is a linear combination of all I patterns

$$(m') = \begin{bmatrix} [m]_{n+1} \\ [m-1]_{i,n} \\ [m-1]_{n-1} \\ \vdots \\ m_{11}-1 \end{bmatrix}, \quad i\neq j = 1,...,n. \qquad (3.22)$$

Let $\{\theta'\}_n \equiv \{\theta'_1,...,\theta'_n\}$ be the θ-tuple, corresponding to (m'). Since $m'_{kn} = m_{kn} - 1 + \delta_{ki}$ and $i\neq j$,

$$\begin{aligned}\theta'_j &= \frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m'_{kn}) - (m_{j,n+1}-m'_{jn}) \\ &= \frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m_{kn}+1-\delta_{ki}) \\ &\quad - (m_{j,n+1}-m_{jn}+1) \\ &= \frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m_{kn}) \\ &\quad - (m_{j,n+1}-m_{jn}) = \theta_j = 1.\end{aligned}$$

Thus $(m')\in\bar{\Gamma}([m]_{n+1})_1$ and, therefore, $e_{n0}(m) \in\bar{W}([m]_{n+1})_1$. Similarly one shows that, if $(m)\in\bar{W}([m]_{n+1})_1$, then also $e_{0n}(m)\in\bar{W}([m]_{n+1})_1$. We conclude that $\bar{W}([m]_{n+1})_1$ is a nontrivial sl(1,n) invariant subspace of $\bar{W}([m]_{n+1})$. Since, moreover, any nontrivial invariant submodule V is contained in the maximal invariant subspace $\bar{I}([m]_{n+1})$ [otherwise $\bar{I}([m]_{n+1})$ will be a proper subspace of a larger invariant subspace $\bar{I}([m]_{n+1}) + V$], we have

$$\bar{W}([m]_{n+1})_1 \subset \bar{I}([m]_{n+1}). \qquad (3.23)$$

■

Proposition 9: Let $V_1,V_2,...,V_r$ be gl(n) fidirmods, all of them being with different signatures. Consider

$$V = V_1\oplus V_2\oplus\cdots\oplus V_r \qquad (3.24)$$

as a gl(n) module. Then for each nonzero $x\in V$ there exists $i=1,...,n$ and $u\in U[\mathrm{gl}(n)]$ such that $0\neq ux\in V_i$.

Proof: The space V is a reducible gl(n) module. Since all signatures (= highest weights) of V_i are different, V contains only r highest weight vectors $x_i\in V_i$, $i=1,...,r$ (certainly each highest weight vector is defined up to a multiplicative constant; we somehow fix this constant). Consider the subspace

$$V_0 = U[\mathrm{gl}(n)]x \equiv \{ux | u\in U[\mathrm{gl}(n)]\}. \qquad (3.25)$$

By construction V_0 is a gl(n) submodule of V. Since every finite-dimensional gl(n) module has a highest weight vector, there exists an element $u\in U[\mathrm{gl}(n)]$, such that $x_0 = ux$ is a gl(n) highest weight vector from V. Therefore, x_0 is proportional to one of the vectors $x_1,...,x_r$, i.e., $x_0 = cx_i\in V_i$, $c\in\mathbb{C}$. Thus $ux\in V_i$. ■

Proposition 10: Let $\bar{W}([m]_{n+1})$ be an indecomposible sl(1,n) module, for which (3.11) holds:

$$\text{if } V([m]_n)\subset\bar{W}([m]_{n+1})_0, \quad \text{then } V([m]_n)\cap\bar{I}([m]_{n+1}) = 0. \qquad (3.26)$$

Proof: Together with the indecomposible $\bar{W}([m]_{n+1})$ we introduce another induced module $\bar{W}([\tilde{m}]_{n+1})$, where

$$[\tilde{m}]_{n+1} = [m_{1,n+1}+c,...,m_{n,n+1}+c] \equiv [m+c]_{n+1}. \qquad (3.27)$$

We fix the constant $c\in\mathbb{C}$ in such a way that

$$\forall k=1,...,n, \quad \tilde{m}_{k,n+1}\neq k-1 \Leftrightarrow \tilde{l}_{k,n+1}+1\neq 0. \qquad (3.28)$$

Therefore, $\bar{W}([\tilde{m}]_{n+1})$ is typical and hence an irreducible sl(1,n) module. Let $(\tilde{m}) = (m+c)$ be a pattern, which is obtained from the I pattern $(m)\in\bar{W}([m]_{n+1})$ after the replacement $m_{ij}\to m_{ij}+c$. Clearly, the set of all patterns $(\tilde{m}) = (m+c)$, $(m)\in\bar{\Gamma}([m]_{n+1})$ gives the I basis $\bar{\Gamma}([\tilde{m}]_{n+1})$ in $\bar{W}([\tilde{m}]_{n+1})$.

Let $V([m]_n)\equiv V(\{\theta\}_n)\subset\bar{W}([m]_{n+1})$ be a gl(n) fidirmod of degree N [see I, Definition 2], i.e., $\theta_1+\cdots+\theta_n = N$. Assume for definiteness that in the I pattern $(m)\in V([m]_n)$ $\theta_{i_1},\theta_{i_2},...,\theta_{i_N} = 1$. Consider the I pattern $(\tilde{m}) = (m+c)\in\bar{W}([\tilde{m}]_{n+1})$ corresponding to (m). From the irreducibility of $\bar{W}([m]_{n+1})$ it follows that

$$\exists 0\neq u\in U \text{ such that } 0\neq u(\tilde{m})\in V([m]_{n+1}). \qquad (3.29)$$

Without loss of generality we can assume that u is a homogeneous polynomial of degree N on e_{0n} and depends otherwise only on the even generators, corresponding to the simple roots, namely $E_{i,i+1}$, $i=1,...,n-1$. In such a case, using Eqs. (2.7) and (2.9), it is not very difficult to show that

$$u(m) = \prod_{k=1}^{N}(l_{i_k,n+1}+1)f[(m)]x_0, \qquad (3.30)$$

$$u(\tilde{m}) = \prod_{k=1}^{N}(\tilde{l}_{i_k,n+1}+1)f[(\tilde{m})]\tilde{x}_0\neq 0, \qquad (3.31)$$

where $\tilde{x}_0\in V([\tilde{m}]_{n+1})$ and $x_0\in V([m]_{n+1})$ are the highest weight vectors of $V([m]_{n+1})$ and $V([m]_{n+1})$, respectively. The function $f[(m)]$ (resp. $f[(\tilde{m})]$) depends only on the differences $m_{ij}-m_{kl}$ (resp. $\tilde{m}_{ij}-\tilde{m}_{kl}$), $i\leqslant j=1,...,n$ and $k\leqslant l=1,...,n$. Since $(m) = (m+c)$, the latter gives

$$f[(m)] = f[(\tilde{m})]. \qquad (3.32)$$

From (3.31) we conclude that $f[(\tilde{m})]\neq 0$ and, therefore, also

$$f[(m)]\neq 0. \qquad (3.33)$$

Assume that $V([m]_n)\subset\overline{W}([m]_{n+1})_0$. Recall also that $V([m]_n)$ is of degree N. Take the I pattern $(m)\in V([m]_n)$, for which $\theta_{i_1},\theta_{i_2},...,\theta_{i_N}$ are equal to one. Since $\theta_j=0$, $j\notin(i_1,i_2,...,i_N)$. Therefore [see (3.13)], $(l_{i_k,n+1}+1)\neq 0$ for each $k=1,...,N$ and hence

$$\prod_{k=1}^{N}(l_{i_k,n+1}+1)\neq 0. \qquad (3.34)$$

Inserting (3.33) and (3.34) in (3.30), we conclude that for any I pattern $(m)\in V([m]_n)\subset\overline{W}([m]_{n+1})_0$ there exists u from the sl(1,n) universal enveloping algebra U, such that $0\neq u(m)\in V([m]_{n+1})$. Therefore also

$$\begin{aligned}&\forall 0\neq x\in V([m]_n)\subset\overline{W}([m]_{n+1})_0,\\ &\exists 0\neq u\in U \text{ such that } 0\neq ux\in V([m]_{n+1}).\end{aligned} \qquad (3.35)$$

Suppose that for a certain $V([m]_n)\subset\overline{W}([m]_{n+1})_0$

$$V([m]_n)\cap\overline{I}([m]_{n+1})\neq 0 \qquad (3.36)$$

and let $0\neq x\in V([m]_n)\cap\overline{I}([m]_{n+1})$. Choose $u\in U$ to be such that (3.35) holds. Then $0\neq ux\in V([m]_{n+1})\cap\overline{I}([m]_{n+1})$, which contradicts Proposition 6. Hence (3.36) is impossible, i.e., (3.26) holds. ■

Theorem: $\overline{I}([m]_{n+1})=\overline{W}([m]_{n+1})_1$.

Proof: Proposition 8 asserts that

$$\overline{W}([m]_{n+1})_1\subset\overline{I}([m]_{n+1}). \qquad (3.37)$$

Suppose that $x\in\overline{I}([m]_{n+1})$. Then [see (3.19)] x can be uniquely represented as $x=x_0+x_1$, where $x_0\in\overline{W}([m]_{n+1})_0$ and $x_1\in\overline{W}([m]_{n+1})_1$. Since [see (3.37)] $x_1\in\overline{I}([m]_{n+1})$, also $x-x_1=x_0\in\overline{I}([m]_{n+1})$, i.e.,

$$x_0\in\sum_{\substack{\theta_1,...,\theta_n=0,1\\ \theta_j=0}}{}'\oplus V(\{\theta\}_n). \qquad (3.38)$$

Suppose that

$$\begin{aligned}&\exists x=x_0+x_1\in\overline{I}([m]_{n+1}),\\ &\qquad\text{such that } 0\neq x_0\in\overline{W}([m]_{n+1})_0.\end{aligned} \qquad (3.39)$$

Since $\overline{W}([m]_{n+1})_0$ is a direct sum of gl(n) fidirmods $V(\{\theta\}_n)$, all of which have different signatures, we can apply Proposition 9. Let $u_0\in U[\text{gl}(n)]$ be such that $0\neq u_0x_0\in V(\{\theta\}_n)\equiv V([m]_n)\subset\overline{W}([m]_{n+1})_0$. In the same time $u_0x_0\in\overline{I}([m]_{n+1})$. Thus we have

$$V([m]_n)\subset\overline{W}([m]_{n+1})_0$$

and

$$V([m]_n)\cap\overline{I}([m]_{n+1})\neq 0, \qquad (3.40)$$

which according to Proposition 10 is impossible. Hence the assumption (3.39) cannot hold, i.e., $x_0=0$. Therefore, $x=x_1\in\overline{W}([m]_{n+1})_1$. We have shown that, if $x\in\overline{I}([m]_{n+1})$, then $x\in\overline{W}([m]_{n+1})_1$, i.e.,

$$\overline{W}([m]_{n+1})_1\supset\overline{I}([m]_{n+1}). \qquad (3.41)$$

The inclusions (3.37) and (3.41) yield

$$\overline{I}([m]_{n+1})=\overline{W}([m]_{n+1})_1. \qquad (3.42)$$

■

B. Transformation of the nontypical modules. First basis

We recall that we are considering an indecomposible sl(1,n) module $\overline{W}([m]_{n+1})$, corresponding to the case $m_{j,n+1}=j-1$ or, which is the same, to $l_{j,n+1}+1=0$. In order to apply the Corollary we have first to choose a complement to $\overline{I}([m]_{n+1})$ space in $\overline{W}([m]_{n+1})$. To this end we combine Eqs. (3.19) and (3.42), writing

$$\overline{W}([m]_{n+1})=\overline{W}([m]_{n+1})_0\oplus\overline{I}([m]_{n+1}). \qquad (3.43)$$

Therefore, as a complement space $W([m]_{n+1})$ [see (3.5)] we take the subspace $\overline{W}([m]_{n+1})_0$, i.e., we set

$$W([m]_{n+1})=\overline{W}([m]_{n+1})_0. \qquad (3.44)$$

In order to obtain the transformation of the factor space under the action of the sl(1,n) generators (see the Corollary) we have to replace everywhere in (2.5)–(2.9) the basis vectors $(m)\in\overline{W}([m]_{n+1})_1$, i.e., all I patterns, corresponding to $\theta_j=1$, by zero.

Proposition 11: Let $\overline{W}([m]_{n+1})$ be an indecomposible sl(1,n) module, corresponding to the case (3.11). Consider an I pattern $(m)\in\overline{W}([m]_{n+1})_0$, for which $\theta_i=0$. Then

$$(\tilde{m})=\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{bmatrix}\in\overline{W}([m]_{n+1})_0$$

$$\text{if and only if } i\neq j. \qquad (3.45)$$

Proof: For (m) both $\theta_i=0$ and $\theta_j=0$. Therefore [see (3.20)], we have

$$\theta_j=\frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m_{kn})-(m_{j,n+1}-m_{jn})=0. \qquad (3.46)$$

Taking into account that $\tilde{m}_{kn}=m_{kn}-1+\delta_{ki}$, and using (3.46), we obtain for $\tilde{\theta}_j$ of $(\tilde{m})$,

$$\begin{aligned}\tilde{\theta}_j&=\frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-\tilde{m}_{kn})\\ &\quad-(m_{j,n+1}-\tilde{m}_{jn})\\ &=\frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m_{kn}+1-\delta_{ki})\\ &\quad-(m_{j,n+1}-m_{jn}+1-\delta_{ji})\\ &=\frac{1}{n-1}\sum_{k=1}^{n}(m_{k,n+1}-m_{kn})-(m_{j,n+1}-m_{jn})\\ &\quad+\delta_{ij}=\delta_{ij}.\end{aligned} \qquad (3.47)$$

Thus $\tilde{\theta}_j=0$ if and only if $i\neq j$, i.e., (3.45) holds.

In a similar way one proves the next statement.

Proposition 12: Let $\overline{W}([m]_{n+1})$ be an indecomposible sl(1,n) module, corresponding to the case (3.11). Consider an I pattern $(m)\in\overline{W}([m]_{n+1})_0$, for which $\theta_i=1$. Then

$$(\tilde{m}) = \begin{bmatrix} [m]_{n+1} \\ [m+1]_{-i,n} \\ [m+1]_{n-1} \\ \vdots \\ m_{11}+1 \end{bmatrix} \in \overline{W}([m]_{n+1})_0, \tag{3.48}$$

if and only if $i \neq j$.

Clearly, the subspace $\overline{W}([m]_{n+1})_0$ is invariant under the action of all even generators of sl$(1,n)$. Therefore, applying the Corollary, we conclude that the relations (2.5)–(2.7) remain unaltered in the factor space. From (3.48) and (2.9) it follows also that $\overline{W}([m]_{n+1})_0$ is closed under the action of e_{0n} and, hence, it is closed under the action of all odd positive roots. Therefore, the relations (2.9) and [I, (3.117)], considered as transformations of the factor space, remain also unaltered.

Turn now to the transformation of $\overline{W}([m]_{n+1})_0$ under e_{n0} [see (2.8)]. Let $\overline{W}([m]_{n+1})$ be the indecomposible module, corresponding to the case (3.11). Then (Proposition 11) all terms in the right-hand side of (2.9), corresponding to $i \neq j$, are vectors from $\overline{W}([m]_{n+1})_0$. However, the I pattern

$$\begin{bmatrix} [m]_{n+1} \\ [m+1]_{j,n} \\ [m+1]_{n-1} \\ \vdots \\ m_{11}+1 \end{bmatrix} \in \overline{W}([m]_{n+1})_1,$$

and, therefore, the term in the right-hand side of (2.9), corresponding to $i=j$, is zero in the factor space. Hence the transformation of the factor space under the action of e_{n0} reads

$$e_{n0} \begin{bmatrix} [m]_{n+1} \\ [m]_n \\ [m]_{n-1} \\ \vdots \\ m_{11} \end{bmatrix} = \sum_{i \neq j=1}^{n} (1-\theta_i)(-1)^{\theta_1+\cdots+\theta_{i-1}} \times \left| \frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in}-1)}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})} \right|^{1/2} \times \begin{bmatrix} [m]_{n+1} \\ [m-1]_{i,n} \\ [m-1]_{n-1} \\ \vdots \\ m_{11}-1 \end{bmatrix}. \tag{3.49}$$

We underline that (3.49) holds only as a transformation of the factor module, when $\overline{W}([m]_{n+1})$ is an indecomposible sl$(1,n)$ module, corresponding to (3.11). Observe that (3.49) can be written also in the following form:

$$e_{n0} \begin{bmatrix} [m]_{n+1} \\ [m]_n \\ [m]_{n-1} \\ \vdots \\ m_{11} \end{bmatrix} = \sum_{i=1}^{n} (1-\theta_i)(-1)^{\theta_1+\cdots+\theta_{i-1}}[1-\delta(l_{i,n+1}+1)] \times \left| \frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in}-1)}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})} \right|^{1/2} \times \begin{bmatrix} [m]_{n+1} \\ [m-1]_{i,n} \\ [m-1]_{n-1} \\ \vdots \\ m_{11}-1 \end{bmatrix}. \tag{3.50}$$

The advantage of the last relation stems from the observation that it gives the right law of transformation both for the typical and for the nontypical sl$(1,n)$ irreducible modules. Indeed, if $\overline{W}([m]_{n+1})$ is a typical module, then for any $i=1,\ldots,n$, $l_{i,n+1}+1 \neq 0$ and, therefore, $1-\delta(l_{i,n+1}+1) = 1$. Thus (3.50) reduces to (2.8). If $\overline{W}([m]_{n+1})$ is indecomposible and corresponds to the case (3.11), then

$$1-\delta(l_{i,n+1}+1) = \begin{cases} 1, & \forall i \neq j = 1,\ldots n, \\ 0, & \text{for } i=j, \end{cases} \tag{3.51}$$

and (3.50) reduces to (3.49). Exactly in the same way one proceeds in order to write the expressions for all odd generators [see I, (3.117) and (3.118)] in the factor space $\overline{W}([m]_{n+1})/\overline{I}([m]_{n+1})$. Assuming that in the typical modules $\overline{W}([m]_{n+1})$ the maximal invariant subspace $\overline{I}([m]_{n+1}) = 0$, we can write in a unified form the transformations of the typical and the nontypical fidirmods under the action of the algebra. We formulate the results, obtained so far, as a separate statement.

Proposition 13: The finite-dimensional irreducible modules $W([m]_{n+1})$ of the Lie superalgebra sl$(1,n)$ are in one-to-one correspondence with the set of all complex n-tuples

$$[m]_{n+1} \equiv [m_{1,n+1}, m_{2,n+1}, \ldots, m_{n,n+1}],$$

$$\text{such that } m_{i,n+1} - m_{j,n+1} \in \mathbb{Z}_+, \tag{3.52}$$

for all $i<j=1,2,\ldots,n$. These numbers are the coordinates of the highest weight

$$\Lambda = \sum_{i=1}^{n} m_{i,n+1} E^i \tag{3.53}$$

of the corresponding fidirmod, which we denote by $\overline{W}([m]_{n+1})$. The basis $\Gamma([m]_{n+1})$ in $W([m]_{n+1})$ is given with all patterns

$$(m) \equiv \begin{bmatrix} [m]_{n+1} \\ [m]_n \\ \vdots \\ [m]_i \\ \vdots \\ m_{11} \end{bmatrix} \equiv \begin{bmatrix} m_{1,n+1}, m_{2,n+1}, \ldots, m_{n,n+1} \\ m_{1n}, m_{2n}, \ldots, m_{nn} \\ \cdots\cdots\cdots \\ m_{1i}, m_{2i}, \ldots, m_{ii} \\ \cdots\cdots\cdots \\ m_{11} \end{bmatrix}, \tag{3.54}$$

consistent with the following conditions:

(1) the numbers $m_{1,n+1}, m_{2,n+1}, \ldots, m_{n,n+1}$ are fixed and label the module $W([m]_{n+1})$;

(2) $m_{in} = m_{i,n+1} + \theta_i - \sum_{k=1}^{n} \theta_k, \quad \theta_1, \theta_2, \ldots, \theta_n = 0,1;$

(3) if $m_{j,n+1} = j-1$, then $\theta_j = 0$;

(4) $m_{i,j+1} - m_{ij} \in \mathbf{Z}_+, \qquad m_{ij} - m_{i+1,j+1} \in \mathbf{Z}_+, \quad \forall i \leqslant j = 1, \ldots, n-1.$ (3.55)

The transformation of $W([m]_{n+1})$ under the action of the even generators is completely determined from relations (2.5)–(2.7). The odd generators e_{p0} and e_{0p}, $p=1,\ldots,n$ transform the fidirmod $W([m]_{n+1})$ as follows:

$$
e_{p0}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ \vdots\\ [m]_p\\ [m]_{p-1}\\ \vdots\\ m_{11}\end{bmatrix} = \sum_{i_n=1}^{n}\sum_{i_{n-1}=1}^{n-1}\cdots\sum_{i_p=1}^{p}(1-\theta_{i_n})(-1)^{\theta_1+\cdots+\theta_{i_n-1}}[1-\delta(l_{i_n,n+1}+1)]
$$
$$
\times\prod_{r=p+1}^{n} S(i_r,i_{r-1})\left|\frac{\Pi_{k\neq i_{r-1}=1}^{r-1}(l_{k,r-1}-l_{i_r,r}-1)\Pi_{k\neq i_r=1}^{r}(l_{kr}-l_{i_{r-1},r-1})}{\Pi_{k\neq i_r=1}^{r}(l_{kr}-l_{i_r,r})\Pi_{k\neq i_{r-1}=1}^{r-1}(l_{k,r-1}-l_{i_{r-1},r-1}-1)}\right|^{1/2}
$$
$$
\times\left|\frac{\Pi_{k\neq i_n=1}^{n}(l_{kn}-l_{i_n n})\Pi_{k=1}^{p-1}(l_{k,p-1}-l_{i_p p}-1)}{\Pi_{k\neq i_n=1}^{n}(l_{k,n+1}-l_{i_n,n+1})\Pi_{k\neq i_p=1}^{p}(l_{kp}-l_{i_p p})}\right|^{1/2}\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i_n n}\\ \vdots\\ [m-1]_{i_p p}\\ [m-1]_{p-1}\\ \vdots\\ m_{11}-1\end{bmatrix}, \tag{3.56}
$$

$$
e_{0p}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ \vdots\\ [m]_p\\ [m]_{p-1}\\ \vdots\\ m_{11}\end{bmatrix} = \sum_{i_n=1}^{n}\sum_{i_{n-1}=1}^{n-1}\cdots\sum_{i_p=1}^{p}\theta_{i_n}(-1)^{\theta_1+\cdots+\theta_{i_n-1}}(l_{i_n,n+1}+1)
$$
$$
\times\prod_{r=p+1}^{n} S(i_r,i_{r-1})\left|\frac{\Pi_{k\neq i_{r-1}=1}^{r-1}(l_{k,r-1}-l_{i_r,r})\Pi_{k\neq i_r=1}^{r}(l_{kr}-l_{i_{r-1},r-1}+1)}{\Pi_{k\neq i_r=1}^{r}(l_{kr}-l_{i_r,r})\Pi_{k\neq i_{r-1}=1}^{r-1}(l_{k,r-1}-l_{i_{r-1},r-1}+1)}\right|^{1/2}
$$
$$
\times\left|\frac{\Pi_{k\neq i_n=1}^{n}(l_{kn}-l_{i_n n})\Pi_{k=1}^{p-1}(l_{k,p-1}-l_{i_p p})}{\Pi_{k\neq i_n=1}^{n}(l_{k,n+1}-l_{i_n,n+1})\Pi_{k\neq i_p=1}^{p}(l_{kp}-l_{i_p p})}\right|^{1/2}\begin{bmatrix}[m]_{n+1}\\ [m+1]_{-i_n n}\\ \vdots\\ [m+1]_{-i_p p}\\ [m+1]_{p-1}\\ \vdots\\ m_{11}+1\end{bmatrix}. \tag{3.57}
$$

In the above relations

$$
S(i,j)=\begin{cases}1, & \text{for } i\leqslant j,\\ -1, & \text{for } i>j.\end{cases} \tag{3.58}
$$

It is worth writing the transformations (3.56) and (3.57) in the case $p=n$, because they, together with (2.5)–(2.7), determine all other generators through the supercommutation relations. These relations were already given [see (2.9) and (3.50)], but for completeness we write them once more here,

$$e_{n0}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{bmatrix}=\sum_{i=1}^{n}(1-\theta_i)(-1)^{\theta_1+\cdots+\theta_{i-1}}[1-\delta(l_{i,n+1}+1)]\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in}-1)}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})}\right|^{1/2}\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{bmatrix}, \tag{3.59}$$

$$e_{0n}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{bmatrix}=\sum_{i=1}^{n}\theta_i(-1)^{\theta_1+\cdots+\theta_{i-1}}(l_{i,n+1}+1)\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in})}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})}\right|^{1/2}\begin{bmatrix}[m]_{n+1}\\ [m+1]_{-i,n}\\ [m+1]_{n-1}\\ \vdots\\ m_{11}+1\end{bmatrix}. \tag{3.60}$$

C. Transformation of the irreducible modules. Second basis

We now proceed to introduce a new basis in the fidirmods $W([m]_{n+1})$, which leads to more symmetric expressions for the odd generators, leaving the expressions (2.5)–(2.7), for the even generators unchanged.

Consider first a typical sl(1,n) module $\overline{W}([m]_{n+1})$. Then $l_{i,n+1}+1\neq 0$, $\forall i=1,...,n$. We set

$$\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{bmatrix}=\prod_{k=1}^{n}(l_{k,n+1}+1)^{\theta_k/2}\left(\begin{matrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{matrix}\right), \tag{3.61}$$

where $\{\theta_1,...,\theta_n\}\equiv\{\theta\}_n$ is the θ-tuple, corresponding to the gl(n) submodule $V([m]_n)\subset W([m]_{n+1})$ [see (3.20)]. The $\tilde{\theta}$-tuple of $V([\tilde{m}]_n)\equiv V([m-1]_{in})$ can be easily obtained from $\{\theta\}_n$,

$$\tilde{\theta}_k=\theta_k+\delta_{ki},\quad k=1,...,n. \tag{3.62}$$

Therefore

$$\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{bmatrix}=(l_{i,n+1}+1)^{\theta_i/2+1/2}\prod_{k\neq i=1}^{n}(l_{k,n+1}+1)^{\theta_k/2}\times\left(\begin{matrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{matrix}\right). \tag{3.63}$$

Taking into account (3.61) and (3.63), we obtain from (3.59),

$$e_{n0}\left(\begin{matrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{matrix}\right)=\sum_{i=1}^{n}(1-\theta_i)(-1)^{\theta_1+\cdots+\theta_{i-1}}\times(l_{i,n+1}+1)^{1/2}\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in}-1)}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})}\right|^{1/2}\times\left(\begin{matrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{matrix}\right). \tag{3.64}$$

Similarly,

$$\begin{bmatrix}[m]_{n+1}\\ [m+1]_{-i,n}\\ [m+1]_{n-1}\\ \vdots\\ m_{11}+1\end{bmatrix}=(l_{i,n+1}+1)^{\theta_i/2-1/2}\prod_{k\neq i=1}^{n}(l_{k,n+1}+1)^{\theta_k/2}\times\left(\begin{matrix}[m]_{n+1}\\ [m+1]_{-i,n}\\ [m+1]_{n-1}\\ \vdots\\ m_{11}+1\end{matrix}\right). \tag{3.65}$$

Substituting (3.61) and (3.65) in (3.60), we have

$$e_{0n}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{pmatrix}$$

$$=\sum_{i=1}^{n}\theta_i(-1)^{\theta_1+\cdots+\theta_{i-1}}$$

$$\times(l_{i,n+1}+1)^{1/2}\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in})}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})}\right|^{1/2}$$

$$\times\begin{bmatrix}[m]_{n+1}\\ [m+1]_{-i,n}\\ [m+1]_{n-1}\\ \vdots\\ m_{11}+1\end{pmatrix}. \quad (3.66)$$

Consider now a nontypical module $W([m]_{n+1})$, corresponding to

$$l_{j,n+1}+1=0\Rightarrow l_{k,n+1}+1\neq 0,\quad \forall k\neq j=1,\ldots,n. \quad (3.67)$$

In this case we set

$$\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{bmatrix}=\prod_{\kappa\neq j=1}^{n}(l_{k,n+1}+1)^{\theta_k/2}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{pmatrix}. \quad (3.68)$$

Then for $i\neq j$

$$\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{bmatrix}$$

$$=(l_{i,n+1}+1)^{1/2}\prod_{k\neq j=1}^{n}(l_{k,n+1}+1)^{\theta_k/2}$$

$$\times\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{pmatrix}. \quad (3.69)$$

Inserting (3.68) and (3.69) in (3.59) we obtain

$$e_{n0}\begin{bmatrix}[m]_{n+1}\\ [m]_n\\ [m]_{n-1}\\ \vdots\\ m_{11}\end{pmatrix}$$

$$=\sum_{i\neq j=1}^{n}(1-\theta_i)(-1)^{\theta_1+\cdots+\theta_{i-1}}$$

$$\times(l_{i,n+1}+1)^{1/2}\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{in}-1)}{\Pi_{k\neq i=1}^{n}(l_{k,n+1}-l_{i,n+1})}\right|^{1/2}$$

$$\times\begin{bmatrix}[m]_{n+1}\\ [m-1]_{i,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{pmatrix}. \quad (3.70)$$

Adding to the right-hand side of (3.70) the zero term

$$(l_{j,n+1}+1)^{1/2}\left|\frac{\Pi_{k=1}^{n-1}(l_{k,n-1}-l_{jn}-1)}{\Pi_{k\neq j=1}^{n}(l_{k,n+1}-l_{j,n+1})}\right|^{1/2}$$

$$\times\begin{bmatrix}[m]_{n+1}\\ [m-1]_{j,n}\\ [m-1]_{n-1}\\ \vdots\\ m_{11}-1\end{pmatrix}, \quad (3.71)$$

we end up with the same relation (3.64) as for the typical case.

In a similar way, taking into account that for $i\neq j$

$$\begin{bmatrix}[m]_{n+1}\\ [m+1]_{-i,n}\\ [+1]_{n-1}\\ \vdots\\ m_{11}+1\end{bmatrix}$$

$$=(l_{i,n+1}+1)^{-1/2}\prod_{k\neq j=1}^{n}(l_{k,n+1}+1)^{\theta_k/2}$$

$$\times\begin{bmatrix}[m]_{n+1}\\ [m+1]_{-i,n}\\ [m+1]_{n-1}\\ \vdots\\ m_{11}+1\end{pmatrix}, \quad (3.72)$$

we derive (3.66), i.e., also in this case we obtain no difference between the relations, corresponding to typical and nontypical modules.

We summarize the results, obtained for the second basis.

Proposition 14: The finite-dimensional irreducible modules $W([m]_{n+1})$ of the Lie superalgebra sl$(1,n)$ are in one-to-one correspondence with the set of all complex n-tuples

$$[m]_{n+1} \equiv [m_{1,n+1}, m_{2,n+1}, \ldots, m_{n,n+1}],$$
$$m_{i,n+1} - m_{j,n+1} \in \mathbb{Z}_+, \tag{3.73}$$

for all $i < j = 1,2,\ldots,n$. The basis in $W([m]_{n+1})$ can be chosen to consist of all patterns

$$|m) \equiv \left[\begin{matrix} [m]_{n+1} \\ [m]_n \\ \vdots \\ [m]_i \\ \vdots \\ m_{11} \end{matrix}\right) \equiv \left[\begin{matrix} m_{1,n+1}, m_{2,n+1}, \ldots, m_{n,n+1} \\ m_{1n}, m_{2n}, \ldots, m_{nn} \\ \cdots\cdots\cdots\cdots \\ m_{1i}, m_{2i}, \ldots, m_{ii} \\ \cdots\cdots\cdots\cdots \\ m_{11} \end{matrix}\right),$$

which are consistent with the conditions (1)–(4) in (3.55). The transformation of the basis under the action of the superalgebra $sl(1,n)$ is completely determined from the relations (3.64) and (3.66) and the expressions for the even generators (2.5)–(2.7), which hold also in the new basis.

[1]T. D. Palev, J. Math. Phys. **28**, 2280 (1987).
[2]V. G. Kac, Adv. Math. **26**, 8 (1977).
[3]V. G. Kac, Lect. Notes Math. **626**, 597 (1978).
[4]T. D. Palev and O. Ts. Stoychev, Preprint JINR E5-82-54, Dubna 1982.

Infinite-dimensional symmetry algebras and an infinite number of conserved quantities of the (2+1)-dimensional Davey–Stewartson equation

Minoru Omote
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan

(Received 8 April 1988; accepted for publication 27 July 1988)

An infinite-dimensional symmetry algebra of the Davey–Stewartson equation is explicitly presented. It is shown that this algebra is a gauge generalization of the symmetry transformation for the Schrödinger equation, and that the Virasoro algebra appears as the subalgebra. An infinite number of conserved quantities associated with the transformations are also obtained.

I. INTRODUCTION

It is known that the Davey–Stewartson (DS)[1] equation can be solved by the inverse scattering method[2] and the Bäcklund transformations.[3] The soliton solution and the multisoliton solutions have been explicitly found.

The quantum theory of the DS equation has been discussed by using the quantum inverse scattering method, and it is shown that the system has the generalized Yang–Baxter algebra.[4]

Recently Champagne and Winternitz[5] have shown, using the numerical method, that the DS equation has an infinite-dimensional symmetry transformation group.

In this paper we will show that the infinite-dimensional symmetry group can be easily obtained as the gauge generalization of the symmetry transformation for the Schrödinger equation, and that the DS equation has an infinite number of conserved quantities associated with the tranformations.

In Sec. II we will review the symmetry transformations of the Schrödinger equation and show that the equation has linearly time-dependent, dilational symmetry transformations. In Sec. III we will show that in the DS equation the symmetry transformations of the Schrödinger equation can be generalized to the infinite-dimensional symmetry group with arbitrary functions of time. Explicit forms of an infinite number of conserved quantities will be given. In Sec. IV we show that the infinite-dimensional symmetry group is a gauge transformation with time-dependent gauge functions, and it is pointed out that the gauge fixing is necessary in the discussion of the DS equation.

II. THE SYMMETRY ALGEBRA OF THE SCHRÖDINGER EQUATION

A. The symmetry algebra of the classical free particle

First of all we will discuss the symmetry algebra of the classical free particle defined by the Lagrangian

$$L=\frac{m}{2}\sum_{i=1}^{d}\left(\frac{dx^i}{dt}\right)^2, \tag{2.1}$$

where d is a dimension of the space. As is well known this system has the following five kinds of symmetry transformations:

(i) space translation,

$$x'^i=x^i+a_0^i,\quad t'=t; \tag{2.2}$$

(ii) time translation,

$$x'^i=x^i,\quad t'=t+\varepsilon_0; \tag{2.3}$$

(iii) rotation,

$$x'^i=x^i-\sum_{j=1}^{d}\omega^i_j x^j,\quad t'=t,\qquad \omega^i_j=-\omega^j_i; \tag{2.4}$$

(iv) the Galilei transformation,

$$x'^i=x^i+a_1^i t,\quad t'=t; \tag{2.5}$$

and (v) dilatation,

$$x'^i=(1+\tfrac{1}{2}\varepsilon_1)x^i,\quad t'=(1+\varepsilon_1)t, \tag{2.6}$$

where a_0^i, a_1^i, ε_0, ε_1, and ω^i_j are arbitrary infinitesimal constants.

Constants of motions associated with the symmetry transformations are given, respectively, by

(i) space translation,

$$p^i=m\dot{x}^i; \tag{2.7}$$

(ii) time translation,

$$H=\frac{m}{2}\sum_{i=1}^{d}\left(\frac{dx^i}{dt}\right)^2=\frac{1}{2m}\sum_{i=1}^{d}(p^i)^2; \tag{2.8}$$

(iii) rotation,

$$L^{[i,j]}=\tfrac{1}{2}(p^i x^j-p^j x^i); \tag{2.9}$$

(iv) the Galilei transformation,

$$G^i=p^i t-mx^i; \tag{2.10}$$

and (v) dilatation,

$$D=\frac{1}{2m}\sum_{i=1}^{d}\{t(p^i)^2-mx^ip^i\}. \tag{2.11}$$

Furthermore it can be easily shown that the equation of motion for the free particle is covariant under a new transformation defined by

(vi) time-dependent dilatation,

$$x'^i=(1+\tfrac{1}{2}\varepsilon_2 t)x^i,\quad t'=(1+\tfrac{1}{2}\varepsilon_2 t)t, \tag{2.12}$$

where ε_2 is an infinitesimal constant. A constant of motion associated with the new transformation is found to be given by

$$F=\frac{1}{4m}\sum_{i=1}^{d}\{t^2(p^i)^2+m^2(x^i)^2-4mtx^ip^i\}. \tag{2.13}$$

Vector fields that generate the transformations are given by

(i) space translation,

$$\hat{p}^i = \frac{\partial}{\partial x^i};$$

(ii) time translation,

$$\hat{H} = \frac{\partial}{\partial t};$$

(iii) rotation,

$$\hat{L}^{[i,j]} = \frac{1}{2}\left(x^i \frac{\partial}{\partial x^j} - x^j \frac{\partial}{\partial x^i}\right); \tag{2.14}$$

(iv) the Galilei transformation,

$$\hat{G}^i = t\frac{\partial}{\partial x^i};$$

(v) dilatation,

$$\hat{D} = t\frac{\partial}{\partial t} + \frac{1}{2}x^i\frac{\partial}{\partial x^i};$$

and (vi) time-dependent dilatation,

$$\hat{F} = \frac{1}{2}t^2\frac{\partial}{\partial t} + \frac{1}{2}tx^i\frac{\partial}{\partial x^i}.$$

These vector fields are shown to satisfy the Lie algebra

$$\begin{aligned}
&[\hat{p}^i,\hat{p}^j] = 0, \quad [\hat{p}^i,\hat{H}] = 0, \quad [\hat{p}^i,\hat{L}^{[j,k]}] = \tfrac{1}{2}(\delta^{ij}\hat{p}^k - \delta^{ik}\hat{p}^j),\\
&[\hat{p}^i,\hat{G}^j] = 0, \quad [\hat{p}^i,\hat{D}] = \tfrac{1}{2}\hat{p}^i, \quad [\hat{p}^i,\hat{F}] = \tfrac{1}{2}\hat{G}^i,\\
&[\hat{H},\hat{L}^{[i,j]}] = 0, \quad [\hat{H},\hat{G}^i] = \hat{p}^i, \quad [\hat{H},\hat{D}] = \hat{H},\\
&[\hat{H},\hat{F}] = \hat{D},\\
&[\hat{L}^{[i,j]},\hat{L}^{[k,l]}] = \tfrac{1}{2}(\delta^{jk}\hat{L}^{[i,l]} - \delta^{jl}\hat{L}^{[i,k]} - \delta^{ik}\hat{L}^{[j,l]}\\
&\qquad + \delta^{il}\hat{L}^{[j,l]}),\\
&[\hat{L}^{[i,j]},\hat{G}^k] = \tfrac{1}{2}(-\delta^{ik}\hat{G}^j + \delta^{jk}\hat{G}^i), \quad [\hat{L}^{[i,j]},\hat{D}] = 0,\\
&[\hat{L}^{[i,j]},\hat{F}] = 0,\\
&[\hat{G}^i,\hat{G}^j] = 0, \quad [\hat{G}^i,\hat{D}] = -\tfrac{1}{2}\hat{G}^i, \quad [\hat{G}^i,\hat{F}] = 0,\\
&[\hat{D},\hat{F}] = \hat{F}.
\end{aligned} \tag{2.15}$$

B. The symmetry algebra of the Schrödinger equation

The Schrödinger equation of the free particle is given by

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\sum_{i=1}^{d}\left(\frac{\partial}{\partial x^i}\right)^2\psi. \tag{2.16}$$

It can be shown that this equation is covariant under the above-mentioned six kinds of transformations,

$$\begin{aligned}
x'^i &= x^i + a_0^i - \omega_j^i x^j + a_1^i t + \tfrac{1}{2}\varepsilon_1 x^i + \tfrac{1}{2}\varepsilon_2 t x^i,\\
t' &= t + \varepsilon_0 + \varepsilon_1 t + \tfrac{1}{2}\varepsilon_2 t^2,
\end{aligned} \tag{2.17}$$

provided that the wave function ψ transforms under (2.17) as

$$\psi'(x',t') = \left[1 + \frac{i}{\hbar}m\sum_{i=1}^{d}a_1^i x^i - \frac{\varepsilon_1}{4}d + \frac{\varepsilon_2}{4}\left\{-dt + \frac{i}{\hbar}m\sum_{i=1}^{d}(x^i)^2\right\}\right]\psi(x,t). \tag{2.18}$$

On the right-hand side of (2.18) the third term represents the transformation property of ψ under the dilatation. It is well known that ψ transforms with a linearly coordinate-dependent phase under the Galilei transformation and that it gives a ray representation of the Galilei transformation group.[6] The last term defines the way ψ should transform under the time-dependent dilatation. Futhermore it is easily shown that (2.16) is covariant under a constant phase transformation of ψ,

$$\psi' = (1 + i\lambda)\psi.$$

Now we will define functions of time $a^i(t)$ and $\varepsilon(t)$ as

$$\begin{aligned}
a^i(t) &= a_0^i + a_1^i t,\\
\varepsilon(t) &= \varepsilon_0 + \varepsilon_1 t + \tfrac{1}{2}\varepsilon_2 t^2;
\end{aligned} \tag{2.19}$$

then infinitesimal transformations (2.17) and (2.18) (including the constant phase transformation) are expressed in terms of $a^i(t)$ and $\varepsilon(t)$ as follows:

$$\begin{aligned}
x'^i &= x^i - \sum_{j=1}^{d}\omega_j^i x^j + a^i(t) + \tfrac{1}{2}\dot{\varepsilon}(t)x^i,\\
t' &= t + \varepsilon(t),
\end{aligned} \tag{2.20}$$

and

$$\psi'(x',t') = \left[1 - \frac{d}{4}\dot{\varepsilon}(t) + \frac{i}{\hbar}\lambda + \frac{i}{\hbar}m \times\left\{\dot{a}^i(t)x^i + \frac{\ddot{\varepsilon}(t)}{4}\sum_{i=1}^{d}(x^i)^2\right\}\right]\psi(x,t), \tag{2.21}$$

where $\dot{a}^i$, $\dot{\varepsilon}$, and so on, represent derivatives of $a^i(t)$, $\varepsilon(t)$ with t.

Vector fields that generate transformations (2.20) and (2.21) are given by

$$\begin{aligned}
&\hat{p}^i = \frac{\partial}{\partial x^i}, \quad \hat{H} = \frac{\partial}{\partial t}, \quad \hat{L}^{[i,j]} = \frac{1}{2}\left(x^i\frac{\partial}{\partial x^j} - x^j\frac{\partial}{\partial x^i}\right),\\
&\hat{G}^i = t\frac{\partial}{\partial x^i} + \frac{i}{\hbar}mx^i\left(\psi\frac{\partial}{\partial\psi} - \psi^*\frac{\partial}{\partial\psi^*}\right),\\
&\hat{D} = t\frac{\partial}{\partial t} + \frac{1}{2}x^i\frac{\partial}{\partial x^i} - \frac{d}{4}\left(\psi\frac{\partial}{\partial\psi} + \psi^*\frac{\partial}{\partial\psi^*}\right),\\
&\hat{F} = \frac{1}{2}t^2\frac{\partial}{\partial t} + \frac{1}{2}tx^i\frac{\partial}{\partial x^i} - \frac{d}{4}t\left(\psi\frac{\partial}{\partial\psi} + \psi^*\frac{\partial}{\partial\psi^*}\right)\\
&\qquad + \frac{i}{4\hbar}m\sum_{i=1}^{d}(x^i)^2\left(\psi\frac{\partial}{\partial\psi} - \psi^*\frac{\partial}{\partial\psi^*}\right),\\
&\hat{\Lambda} = \frac{i}{\hbar}\left(\psi\frac{\partial}{\partial\psi} - \psi^*\frac{\partial}{\partial\psi^*}\right).
\end{aligned} \tag{2.22}$$

It follows that these vector fields satisfy the Lie algebra

$$\begin{aligned}
&[\hat{p}^i,\hat{L}^{[j,k]}] = \tfrac{1}{2}(\delta^{ij}\hat{p}^k - \delta^{ik}\hat{p}^j), \quad [\hat{p}^i,\hat{G}^j] = m\delta^{ij}\hat{\Lambda},\\
&[\hat{p}^i,\hat{D}] = \tfrac{1}{2}\hat{p}^i, \quad [\hat{p}^i,\hat{F}] = \tfrac{1}{2}\hat{G}^i, \quad [\hat{H},\hat{G}^i] = \hat{p}^i,\\
&[\hat{H},\hat{D}] = \hat{H}, \quad [\hat{H},\hat{F}] = \hat{D},\\
&[\hat{L}^{[i,j]},\hat{L}^{[k,l]}] = \tfrac{1}{2}(\delta^{jk}\hat{L}^{[i,l]} - \delta^{il}\hat{L}^{[i,k]}\\
&\qquad - \delta^{ik}\hat{L}^{[j,l]} + \delta^{il}\hat{L}^{[j,l]}),\\
&[\hat{L}^{[i,j]},\hat{G}^k] = \tfrac{1}{2}(-\delta^{ik}\hat{G}^j + \delta^{jk}\hat{G}^i),\\
&[\hat{G}^i,\hat{D}] = -\tfrac{1}{2}\hat{G}^i,\\
&[\hat{D},\hat{F}] = \hat{F},
\end{aligned} \tag{2.23}$$

and the others are zero.

The commutation relations of vector fields in (2.15) and (2.23) have same forms except for $[\hat{p}^i,\hat{G}^i] = m\delta^{ij}\hat{\Lambda}$. Since $\hat{\Lambda}$ commutes all other vector fields, it plays the role of the center term for the Lie algebra.

Next we will obtain conserved quantities of the Schrödinger equation associated with the symmetry transformations (2.20) and (2.21). The Lagrangian of the Schrödinger equation is given by

$$L=\frac{i}{2}\hbar(\psi^*\partial_t\psi-\partial_t\psi^*\psi)-\frac{\hbar^2}{2m}\sum_{i=1}^{d}\frac{\partial\psi^*}{\partial x^i}\frac{\partial\psi}{\partial x^i}. \tag{2.24}$$

Then using the Noether theorem it follows that the conserved quantities are given by

$$\begin{aligned}
&Q=\int J^0(\mathbf{x},t)dV,\quad H=\int T_0^0\,dV,\quad p^i=\int T_0^i\,dV,\\
&L^{[i,j]}=\int (x^iT_j^0-x^jT_i^0\,dV,\\
&G^i=\int (mx^iJ^0+tT_i^0)dV,\\
&D=\int\left(tT_0^0+\frac{1}{2}x^iT_i^0\right)dV,\\
&F=\int\left\{t^2T_0^0+tx^iT_i^0+\frac{m}{2}\sum_{i=1}^{d}(x^i)^2J^0\right\}dV,
\end{aligned} \tag{2.25}$$

where

$$\begin{aligned}
&J^0=\psi^*\psi,\quad T_0^0=\frac{\hbar^2}{2m}\sum_{i=1}^{d}\frac{\partial\psi^*}{\partial x^i}\frac{\partial\psi}{\partial x^i},\\
&T_i^0=-\frac{i\hbar}{2}\left(\frac{\partial\psi^*}{\partial x^i}\psi-\psi^*\frac{\partial\psi}{\partial x^i}\right).
\end{aligned} \tag{2.26}$$

III. THE SYMMETRY ALGEBRA OF THE DAVEY–STEWARTSON EQUATION

The Davey–Stewartson (DS) equation is the generalized Schrödinger equation in the $(2+1)$-dimensional space-time which has a nonlinear self-interaction and couples with a scalar field ϕ.

In this section we will discuss symmetry transformations of the DS equation, which is defined by the Lagrangian

$$\begin{aligned}
L=&\frac{i\hbar}{2}(\psi^*\partial_t\psi-\partial_t\psi^*\psi)-\frac{\hbar^2}{2m}(\partial_x\psi^*\partial_x\psi+\partial_y\psi^*\partial_y\psi)\\
&-\frac{1}{2}(\partial_x\phi\,\partial_x\phi-\partial_y\phi\,\partial_y\phi)-\frac{\kappa}{2}|\psi|^4-\mu|\psi|^2\partial_y\phi,
\end{aligned} \tag{3.1}$$

where κ and μ are coupling constants. From (3.1) it follows that ψ and ϕ satisfy the following differential equation:

$$\begin{aligned}
&i\hbar\,\partial_t\psi=-(\hbar^2/2m)(\partial_x^2+\partial_y^2)\psi+\kappa|\psi|^2\psi+\mu\psi\,\partial_y\phi,\\
&(\partial_x^2-\partial_y^2)\phi=-\mu\,\partial_y|\psi|^2.
\end{aligned} \tag{3.2}$$

It can be shown that the DS equation is covariant under (2.20) (without rotations) and (2.21), that is,

$$\begin{aligned}
&x'=x+a_x(t)+\tfrac{1}{2}\dot\varepsilon(t)x,\\
&y'=y+a_y(t)+\tfrac{1}{2}\dot\varepsilon(t)y,\\
&t'=t+\varepsilon(t),
\end{aligned} \tag{3.3}$$

and

$$\begin{aligned}
\psi'(x',t')=\Big[1-\frac{1}{2}\dot\varepsilon+\frac{i}{\hbar}\lambda+\frac{im}{\hbar}&\\
\times\Big\{\dot a_xx+\dot a_yy+\frac{\ddot\varepsilon}{4}(x^2+y^2)\Big\}\Big]\psi(x,y,t),&
\end{aligned} \tag{3.4}$$

provided that $\phi(x,y,t)$ transforms as

$$\phi'(x',y',t')=\{1-\tfrac{1}{2}\dot\varepsilon(t)\}\phi(x,y,t). \tag{3.5}$$

In (3.3)–(3.5), $a_x(t)$, $a_y(t)$, and $\varepsilon(t)$ are given by

$$\begin{aligned}
&a_x(t)=a_{x0}+a_{x1}t,\quad a_y(t)=a_{y0}+a_{y1}t,\\
&\varepsilon(t)=\varepsilon_0+\varepsilon_1t+\tfrac{1}{2}\varepsilon_2t^2.
\end{aligned} \tag{3.6}$$

As explicitly expressed by (3.6) we have, so far, considered symmetry transformations with infinitesimal quantities a_x, a_y, and ε which are linear or quadratic functions of t. Now we will consider generalized symmetry transformations which are defined by (3.3) and (3.4) with arbitrary infinitesimal functions of time $a_x(t)$, $a_y(t)$, $\varepsilon(t)$, and $\lambda(t)$. Under the generalized transformations we will assume that ϕ transforms as

$$\phi'(x',y',t')=(1-\tfrac{1}{2}\dot\varepsilon)\phi(x,y,t)+\Delta\phi(x,y,t). \tag{3.7}$$

Here let us consider the conditions on $\Delta\phi$ under which the DS equation transforms covariantly. Then it follows that $\Delta\phi$ has to satisfy

$$\begin{aligned}
\partial_y(\Delta\phi)=&-(1/\mu)\{\dot\lambda+m(\ddot a_xx+\ddot a_yy)\\
&+(m/4)\dddot\varepsilon(x^2+y^2)\},\\
\partial_x^2(\Delta\phi)=&-(m/\mu)\ddot a_y-(m/2\mu)\dddot\varepsilon y.
\end{aligned} \tag{3.8}$$

From (3.8) we find that

$$\begin{aligned}
\phi'(x',y',t)=&\left(1-\frac{1}{2}\dot\varepsilon\right)\phi(x,y,t)\\
&-\frac{\dot\lambda}{\mu}y-\frac{m}{2\mu}\{2xy\ddot a_x+(x^2+y^2)\ddot a_y\}\\
&-\frac{m}{4\mu}\dddot\varepsilon\left(x^2y+\frac{1}{3}y^3\right)\\
&+\frac{x}{\mu}\alpha(t)+\frac{1}{\mu}\beta(t),
\end{aligned} \tag{3.9}$$

where $\alpha(t)$ and $\beta(t)$ are arbitrary infinitesimal functions of t. Thus we have shown that the DS equation is covariant under the generalized transformations (3.3), (3.4), and (3.9). We can easily check that these transformations are equivalent to those found by Champagne and Winternitz.[5]

Vector fields that generate the symmetry transformations are given by

$$\begin{aligned}
\hat X(a_x)=&a_x\frac{\partial}{\partial x}+\frac{i}{\hbar}mx\dot a_x\left(\psi\frac{\partial}{\partial\psi}-\psi^*\frac{\partial}{\partial\psi^*}\right)\\
&-\frac{m}{\mu}xy\ddot a_x\frac{\partial}{\partial\phi},\\
\hat Y(a_x)=&a_y\frac{\partial}{\partial y}+\frac{i}{\hbar}my\dot a_y\left(\psi\frac{\partial}{\partial\psi}-\psi^*\frac{\partial}{\partial\psi^*}\right)\\
&-\frac{m}{2\mu}(x^2+y^2)\ddot a_y\frac{\partial}{\partial\phi},
\end{aligned}$$

$$\hat{D}(\varepsilon) = \frac{1}{2}\dot{\varepsilon}\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right) + \varepsilon\frac{\partial}{\partial t}$$
$$-\frac{1}{2}\dot{\varepsilon}\left(\psi\frac{\partial}{\partial\psi} + \psi^*\frac{\partial}{\partial\psi^*} + \phi\frac{\partial}{\partial\phi}\right)$$
$$+\frac{im}{4\hbar}\ddot{\varepsilon}\,(x^2+y^2)\left(\psi\frac{\partial}{\partial\psi} - \psi^*\frac{\partial}{\partial\psi^*}\right)$$
$$-\frac{m}{4\mu}\ddot{\varepsilon}\left(x^2y + \frac{1}{3}y^3\right)\frac{\partial}{\partial\phi}, \tag{3.10}$$
$$\hat{\Lambda}(\lambda) = \frac{i}{\hbar}\lambda\left(\psi\frac{\partial}{\partial\psi} - \psi^*\frac{\partial}{\partial\psi^*}\right) - \frac{\dot{\lambda}}{\mu}y\frac{\partial}{\partial\phi},$$
$$\hat{\Phi}(\alpha) = \alpha(t)\frac{x}{\mu}\frac{\partial}{\partial\phi}, \quad \hat{\Theta}(\beta) = \frac{\beta(t)}{\mu}\frac{\partial}{\partial\phi}.$$

It is easy to see that these vector fields satisfy the following commutation relations

$$[\hat{X}(a_x^{(1)}),\hat{X}(a_x^{(2)})] = m\hat{\Lambda}(a_x^{(1)}\dot{a}_x^{(2)} - \dot{a}_x^{(1)}a_x^{(2)}),$$
$$[\hat{X}(a_x),\hat{Y}(a_y)] = -m\hat{\Phi}(a_x\ddot{a}_y - \ddot{a}_x a_y),$$
$$[\hat{Y}(a_y^{(1)}),\hat{Y}(a_y^{(2)})] = m\hat{\Lambda}(a_y^{(1)}\dot{a}_y^{(2)} - \dot{a}_y^{(1)}a_y^{(2)}),$$
$$[\hat{X}(a_x),\hat{D}(\varepsilon)] = \hat{X}(\tfrac{1}{2}a_x\dot{\varepsilon} - \varepsilon\dot{a}_x),$$
$$[\hat{Y}(a_y),\hat{D}(\varepsilon)] = \hat{Y}(\tfrac{1}{2}a_y\dot{\varepsilon} - \varepsilon\dot{a}_y),$$
$$[\hat{D}(\varepsilon_1),\hat{D}(\varepsilon_2)] = \hat{D}(\varepsilon_1\dot{\varepsilon}_2 - \dot{\varepsilon}_1\varepsilon_2), \tag{3.11}$$
$$[\hat{y}(a_y),\hat{\Lambda}(\lambda)] = -\hat{\theta}(\dot{\lambda}a_y),$$
$$[\hat{D}(\varepsilon),\hat{\Lambda}(\lambda)] = \hat{\Lambda}(\varepsilon\dot{\lambda}),$$
$$[\hat{D}(\varepsilon),\hat{\Phi}(\alpha)] = \Phi(\alpha\dot{\varepsilon}),$$
$$[\hat{D}(\varepsilon),\hat{\Theta}(\beta)] = \tfrac{1}{2}\hat{\Theta}(\beta\dot{\varepsilon}),$$

and the other commutation relations are zero.

Since the above-mentioned symmetry transformations have arbitrary functions of t, they constitute an infinite-dimensional algebra. In order to see this it is convenient to expand $a_x(t)$, $a_y(t)$, $\varepsilon(t)$, $\lambda(t)$, $\alpha(t)$, and $\beta(t)$ into the Laurent series of t. For example, if we expand $D(\varepsilon)$ as

$$D(\varepsilon) = \sum_{n=-\infty}^{\infty} t^{1-n}\overset{(n)}{D}, \tag{3.12}$$

then from (3.11) it follows that

$$\left[\overset{(n)}{D},\overset{(m)}{D}\right] = (n-m)\overset{(n+m)}{D}. \tag{3.13}$$

Thus we find that the DS equation has as its symmetry algebra the Virasoro algebra, which has played the important role in the quantum field theory of the $(1+1)$-dimensional conformally invariant system.

We turn next to the problem of finding conserved quantities associated with the infinite-dimensional transformations. These quantities are found to be given by

$$Q(\lambda) = \int \lambda(t)J^0(x,y,t)dx\,dy,$$
$$P_x(a_x) = \int \{a_x(t)T^0_x + m\dot{a}_x(t)xJ^0\}dx\,dy,$$
$$P_y(a_y) = \int \{a_y(t)T^0_y + m\dot{a}_y(t)yJ^0\}dx\,dy, \tag{3.14}$$
$$D(\varepsilon) = \int \left\{\varepsilon T^0_0 + \frac{1}{2}\dot{\varepsilon}(xT^0_x + yT^0_y) + \frac{m}{4}\ddot{\varepsilon}(x^2+y^2)J^0\right\}dx\,dy,$$

where $J^0 = \psi^*\psi$, and

$$T^0_0 = (\hbar^2/2m)(\partial_x\psi^*\,\partial_x\psi + \partial_y\psi^*\,\partial_y\psi)$$
$$+\tfrac{1}{2}(\partial_x\phi\,\partial_x\phi - \partial_y\phi\,\partial_y\phi)$$
$$+\tfrac{1}{2}\kappa\,|\psi|^4 + \mu\,|\psi|^2\,\partial_y\phi, \tag{3.15}$$
$$T^0_x = (-i\hbar/2)(\partial_x\psi^*\psi - \psi^*\,\partial_x\psi),$$
$$T^0_y = -(i\hbar/2)(\partial_y\psi^*\psi - \psi^*\,\partial_y\psi).$$

From (3.14) we see that the DS equation has an infinite number of conserved quantities.

Here we have noticed that all of the infinite number of conserved quantities can be expressed in terms of the number density J^0 and the energy-momentum density T^i_j $(i,j = 0,x,y)$. On the other hand, as is well known, completely integrable models in the $(1+1)$-dimensional space-time have another kind of infinite number of conserved quantities that play the role of the action variables and have various functional forms of field variables. In the previous paper we have shown that the Thirring model is completely integrable and has the conformal invariance, and then that the model has the above-mentioned two kinds of conserved quantities.[7]

In the DS equation, which has been shown to be solved by the inverse scattering method, it is an interesting problem to find explicit forms of the conserved quantities corresponding to the action variables.

IV. GAUGE SYMMETRIES OF THE DS EQUATION

By integrating the differential equation of ϕ given by (3.2), we find that ϕ can be expressed as

$$\phi(x,y,t) = \phi_0(x,y,t) - \mu\int dx'\,dy'\,G(x,y;x'y')\partial'_y|\psi'|^2, \tag{4.1}$$

where $G(x,y,x',y')$ is the Green's function that satisfies

$$(\partial_x^2 - \partial_y^2)G(x,y;x',y') = \delta(x-x')\delta(y-y'), \tag{4.2}$$

and $\phi_0(x,y,t)$ is an arbitrary solution of the differential equation

$$(\partial_x^2 - \partial_y^2)\phi_0(x,y,t) = 0. \tag{4.3}$$

Next let us consider $\Delta\phi$ defined by

$$\Delta\phi = -\frac{\dot{\lambda}}{\mu}y - \frac{m}{2\mu}\{2xy\ddot{a}_x + (x^2+y^2)\ddot{a}_y\}$$
$$-\frac{m}{4\mu}\dddot{\varepsilon}\left(x^2y + \frac{1}{3}y^3\right) + \alpha(t)\frac{x}{\mu} + \frac{\beta(t)}{\mu}, \tag{4.4}$$

which represents the variation of ϕ generated by the generalized symmetry transformation. It can be easily shown that $\Delta\phi$, with arbitrary functions λ, a_x, a_y, ε, α, and β of t, satisfies

$$(\partial_x^2 - \partial_y^2)\Delta\phi = 0. \tag{4.5}$$

From (4.3) and (4.5) we see that $\phi_0(x,y,t)$ transforms under the generalized transformations as

$$\phi_0'(x',y',t') = (1 - \tfrac{1}{2}\dot{\varepsilon})\phi_0(x,y,t) + \Delta\phi. \tag{4.6}$$

In other words, the freedom of the generalized symmetry transformations (3.3) and (3.4) comes from the arbitrariness of ϕ_0. From this point of view we can say that the generalized symmetry transformation of the DS equation is a gauge transformation with time-dependent gauge functions, and that ϕ_0 plays the role of gauge field.

In order to represent generators of the gauge transformations in terms of field variables we will introduce the canonical conjugate momentum field $\pi(x,y,t)$ of ϕ_0. This field $\pi(x,y,t)$ satisfies a first-kind constraint $\pi(x,y,t) = 0$, since the Lagrangian (3.1) does not have a $\dot{\phi}$ term. Then we find that generators of the gauge transformations are given by

$$\hat{\Lambda}(\lambda) = -\int \left\{\lambda J^0 + \frac{\dot{\lambda}}{\mu} y\pi\right\} dx\, dy,$$

$$\hat{X}(a_x) = -\int \left\{a_x T_x^0 + m\dot{a}_x x J^0 + \frac{m}{\mu} xy\ddot{a}_x \pi + a_x \partial_x \phi_0 \pi\right\} dx\, dy,$$

$$\hat{Y}(a_y) = -\int \left\{a_y T_y^0 + m\dot{a}_y y J^0 - \frac{m}{2\mu}(x^2 + y^2)\ddot{a}_y \pi + a_y \partial_y \phi_0 \pi\right\} dx\, dy,$$

$$\hat{D}(\varepsilon) = -\int \left\{\varepsilon T_0^0 + \frac{1}{2}\dot{\varepsilon}(xT_x^0 + yT_y^0) + \frac{m}{4}\ddot{\varepsilon}(x^2+y^2)J^0 + \frac{1}{2}\dot{\varepsilon}\phi_0\pi - \frac{m}{4\mu}\dddot{\varepsilon}\left(x^2 y + \frac{1}{3}y^3\right)\pi + \varepsilon\phi_0\pi + \frac{\dot{\varepsilon}}{2}(x\partial_x\phi_0 + y\partial_y\phi_0)\pi\right\} dx\, dy, \tag{4.7}$$

$$\hat{\Phi}(\alpha) = \frac{1}{\mu}\int \alpha(t) x\pi\, dx\, dy,$$

$$\hat{\Theta}(\beta) = \frac{1}{\mu}\int \beta(t)\pi\, dx\, dy.$$

Since the DS equation has the gauge symmetry, we have to fix the gauge freedom in order to obtain solutions of the classical DS equation, or in order to discuss the quantum theory of the system. As the simplest choice of gauge we can take the gauge fixing defined by

$$\phi_0(x,y,t) = 0. \tag{4.8}$$

In this gauge the DS equation can be shown to be reduced to the system that has the nonlocal-nonlinear interaction of ψ and ψ^*.

[1]A. Davey and K. Stewartson, Proc. R. Soc. London Ser. A **338**, 101 (1974).
[2]D. Anker and N. C. Freeman, Proc. R. Soc. London Ser. A **360**, 592 (1978).
[3]S. Ganesan and M. Lakshmanan, J. Phys. A. Math. Gen. **20**, 1143 (1987).
[4]C. L. Schultz and M. J. Ablowitz, Phys. Rev. Lett. **59**, 2825 (1987).
[5]B. Champagne and P. Winternitz, J. Math. Phys. **29**, 1 (1988).
[6]V. Bargmann, Ann. Math. **59**, 1 (1954).
[7]K. Inoue, K. Odaka, and M. Omote, J. Math. Phys. **29**, 476 (1988).

Adiabatic switching for time-dependent electric fields

Márcia A. G. Scialom
Instituto de Matemática, Estatística e Ciência da Computação—Universidade Estadual de Campinas, Caixa Postal 6065, CEP. 13.081—Campinas, São Paulo, Brazil

Rafael J. Iório, Jr.
Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina, 110, CEP. 22.460—Rio de Janeiro, Rio de Janeiro, Brazil

(Received 30 April 1987; accepted for publication 29 June 1988)

In this work the scattering theory associated with the differential equation $i(\partial\psi/\partial t) = (-\Delta + e^{-\varepsilon|t|}g(t)x_1 + q(x))\psi$ is considered, where $x = (x_1,x^\perp)\in\mathbb{R}\times\mathbb{R}^2$, $\varepsilon\geqslant 0$, $\omega > 0$, $\alpha\in\mathbb{R}$, $g(t)$, $t\in\mathbb{R}$ is continuous, periodic with mean value zero over a period, and $q(x)$ approaches to zero sufficiently fast as $|x|\to\infty$. In the case $\varepsilon > 0$, it is shown that the usual theory is adequate; however, a limit does not exist when $\varepsilon\downarrow 0$. A modified theory is developed where the limit does exist as $\varepsilon\downarrow 0$. Furthermore, the concepts of bound states and scattering states for $\varepsilon\geqslant 0$ are discussed.

I. INTRODUCTION

In this paper we will discuss the scattering theory associated with the Cauchy problem,

$$i\,\partial_t\psi = (-\Delta + e^{-\varepsilon|t|}g(t)x_1 + q(x))\psi, \qquad \psi(x,s) = \psi_s(x)\in L^2(\mathbb{R}^3), \tag{1.1}$$

where $x = (x_1,x^\perp)\in\mathbb{R}\times\mathbb{R}^2$, t, $s\in\mathbb{R}$, $\varepsilon\geqslant 0$, $g(t)$, $q(x)$ are both real valued, $g(t)$ is continuous and bounded, and $q(x)$ has the form

$$q(x) = q_1(x) + q_2(x), \qquad q_1\in L^\infty_\infty(\mathbb{R}^3),\quad q_2\in L^2(\mathbb{R}^3). \tag{1.2}$$

Here $L^\infty_\infty(\mathbb{R}^3)$ denotes the set of $f\in L^\infty(\mathbb{R}^3)$ that tend to zero at infinity. Further assumptions on q and g will be introduced as we proceed. Under these conditions, the operator defined by

$$\circ A^\varepsilon(t) = -\Delta + e^{-\varepsilon|t|}g(t)x_1 + q(x), \quad D(\circ A^\varepsilon(t)) = C_0^\infty(\mathbb{R}^3), \tag{1.3}$$

is essentially self-adjoint (see Sec. 3 of Ref. 1 and references therein). We denote its closure by $A^\varepsilon(t)$ and write $A_0^\varepsilon(t)$ for the case $q = 0$. As is well known, (1.1) describes the interaction of a quantum-mechanical particle in the semiclassical approximation with a potential $q(x)$ and the electric field $e^{-\varepsilon|t|}g(t)$ (1,0,0). The case $\varepsilon = 0$ was studied in Ref. 1 where existence and uniqueness of solutions for (1.1) was proved assuming that q_1 is also continuous. As pointed out by Kato[2] this assumption is not needed. It should be stressed, however, that the hypotheses in Ref. 1 already cover the Coulomb potential case, as far as existence and uniqueness are concerned. From now on we will assume that $g(t)$ is periodic with period $\tau > 0$ and

$$\int_0^\tau g(t)dt = 0. \tag{1.4}$$

In this case a satisfactory scattering theory was established in Ref. 1 (see also Ref. 3) under the assumptions

$$q(x) = (1 + |x|^2)^{-\rho}(W_1(x) + W_2(x)), \tag{1.5}$$

$$\rho > \tfrac{1}{2}, \quad W_1\in L^\infty(\mathbb{R}^3), \quad W_2\in L^2(\mathbb{R}^3),$$

$$\frac{\partial W_1}{\partial x_1}\in L^2(\mathbb{R}^3), \tag{1.6}$$

where the derivative in (1.6) is computed in the sense of distributions. More precisely, if $U_{A^0}(t,s)$ is the propagator associated to (1.1) (with $\varepsilon = 0$) and $\Theta(s) = U_{A^0}(s+\tau,s)$ is the Floquet operator of the system, then

$$L^2(\mathbb{R}^3) = \mathscr{H}_{\rm ac}(\Theta(s)) \oplus \mathscr{H}_p(\Theta(s)), \tag{1.7}$$

$$\mathscr{R}(\Omega_\pm(A^0,A_0^0;s)) = \mathscr{H}_{\rm ac}(\Theta(s)), \tag{1.8}$$

where $\mathscr{H}_p(U)$ and $\mathscr{H}_{\rm ac}(U)$ are, respectively, the pure point and absolutely continuous subspaces associated with the unitary operator U, and the wave operators are defined by

$$\Omega_\pm(A^0,A_0^0;s) = \operatorname*{s-lim}_{t\to\pm\infty} U_{A^0}(t,s)^* U_{A_0^0}(t,s). \tag{1.9}$$

It can also be shown[1] that $\mathscr{H}_p(\Theta(s))$ and $\mathscr{H}_{\rm ac}(\Theta(s))$ are precisely the bound state and scattering state subspaces in the time-dependent sense (see Sec. IV). In particular the "free" dynamics in this formulation is determined by the Hamiltonian $A_0^0(t)$. Although this is a very pleasing theory from the mathematical point of view, physically one would expect to be able to compare the dynamics generated by $A(t)$ with the one determined by $H_0 = -\Delta$ [the Laplacian in $L^2(\mathbb{R}^3)$] since, after all, the mean value of $A^0(t)$ over a period is simply $H = H_0 + q$ and there is a very well established scattering theory for the pair (H,H_0). That this can in fact be done by suitably modifying the wave operators is shown in Sec. IV. This was one of the main motivations for this work.

We were also interested in the so-called adiabatic switching of the field which is often used in physics (see Refs. 4–6 and the references therein). Roughly speaking, this procedure consists in introducing a "regularizing factor" depending continuously on some parameter $\varepsilon > 0$ (in our case $e^{-\varepsilon|t|}$), developing the theory in this situation and taking limits as $\varepsilon\downarrow 0$ in the hope of being able to handle the (in principle) more difficult case $\varepsilon = 0$. In connection with this, one should note that Dollard[7] has studied adiabatic switching in the usual theory of scattering. More precisely, he introduces the Hamiltonian $H(t) = H_0 + e^{-\varepsilon|t|}q$ and shows

 0022-2488/88/122604-07$02.50

that if $q(x)$ is a short range potential, the usual wave operators with $\varepsilon > 0$ exist and are unitary and, in the limit, they coincide with the wave operators for the pair (H,H_0). On the other hand, if q is the Coulomb potential the same result holds in the case $\varepsilon > 0$, but the limit does not exist. Dollard also shows how to modify the theory in order to obtain the right wave operators as $\varepsilon \downarrow 0$. Note that in both situations there are no bound states if $\varepsilon > 0$. In the electric field case the situation is different. In Sec. III we show that if $\varepsilon > 0$ and $H = H_0 + q$ has a bound state then there are solutions of (1.1) that behave as bound states as $t \to \pm\infty$. We also prove that the usual wave operators exist. In the following section we show that these operators do not have a limit as $\varepsilon \downarrow 0$. The definitions are then modified and a satisfactory scattering theory is obtained in the limit, as mentioned above. Section II contains some notation and various technical results that will be used in the remainder of this work.

II. PRELIMINARIES

We begin by introducing several auxiliary functions which will be needed in the next three sections. Assume that $g: \mathbb{R} \to \mathbb{R}$ is continuous periodic with period $\tau > 0$ and satisfies (1.4). In this case it is easy to see that we can choose h and G such that for all $t \in \mathbb{R}$,

$$\begin{aligned} &h'(t) = g(t), \quad G'(t) = h(t), \\ &h(t+\tau) = h(t), \quad G'(t+\tau) = G(t), \\ &\int_0^\tau h(t)dt = \int_0^\tau G(t)dt = 0. \end{aligned} \tag{2.1}$$

Moreover, we will also need $k(t)$ such that

$$k'(t) = h(t)^2. \tag{2.2}$$

Next, if $\varepsilon \geqslant 0$, we define functions g^ε, h^ε, G^ε, and k^ε as follows. If $\varepsilon = 0$ let g^0, h^0, G^0, k^0 be the functions just introduced. If $\varepsilon > 0$, choose

$$g^\varepsilon(t) = \exp(-\varepsilon|t|)g(t), \tag{2.3}$$

$$h^\varepsilon(t) = \begin{cases} -\int_t^\infty g^\varepsilon(s)ds, & t \geqslant 0, \\ \int_{-\infty}^t g^\varepsilon(s)ds, & t < 0, \end{cases} \tag{2.4}$$

$$G^\varepsilon(t) = \begin{cases} -\int_t^\infty h^\varepsilon(s)ds, & t \geqslant 0, \\ \int_{-\infty}^t h^\varepsilon(s)ds, & t < 0, \end{cases} \tag{2.5}$$

$$k^\varepsilon(t) = \begin{cases} -\int_t^\infty (h^\varepsilon(s))^2 ds, & t \geqslant 0, \\ \int_{-\infty}^t (h^\varepsilon(s))^2 ds, & t < 0. \end{cases} \tag{2.6}$$

Now assume that $q(x)$ satisfies (1.2) and let $\psi(x,t)$ be the solution of (1.1) with $\varepsilon \geqslant 0$ fixed (which exists globally and is unique; see Theorem 2.1 below), and introduce

$$\varphi(x,t) = \exp(ih^\varepsilon(t)x_1)\psi(x,t), \tag{2.7}$$

$$\chi(x,t) = \exp(ik^\varepsilon(t))\varphi(x_1 - 2G^\varepsilon(t),x^\perp). \tag{2.8}$$

Then an easy computation shows that φ and χ are solutions of the equations

$$i\,\partial_t\varphi = [((1/i)\partial_{x_1} - h^\varepsilon(t))^2 - \Delta^\perp]\varphi + q\varphi, \tag{2.9}$$

$$i\,\partial_t\chi = (-\Delta + q(x_1 - 2G^\varepsilon(t),x^\perp))\chi, \tag{2.10}$$

where $\Delta^\perp$ denotes the Laplacian with respect to the $x^\perp$ variable.

Let $\circ A^\varepsilon(t)$, $\circ B^\varepsilon(t)$, and $\circ H^\varepsilon(t)$, be the Hamiltonians that occur on the right-hand sides of (1.1), (2.9), and (2.10) with domain $C_0^\infty(\mathbb{R}^3)$. These operators are essentially self-adjoint and we will denote their self-adjoint realizations in $L^2(\mathbb{R}_3)$ by $A^\varepsilon(t)$, $B^\varepsilon(t)$, and $H^\varepsilon(t)$ (see Ref. 1 and the references therein). In case $q = 0$ we will write $A_0^\varepsilon(t)$, $B_0^\varepsilon(t)$, and H_0. Applying Kato's theory of existence and uniqueness for linear "hyperbolic" evolution equations,[8] it was shown in Ref. 1 that the following theorem holds.

Theorem 2.1: Let $K(t)$ denote any one of the three operators $A^\varepsilon(t)$, $B^\varepsilon(t)$, $H^\varepsilon(t)$. Then there exists a unique evolution operator (propagator) $U_K(t,s)$, $(t,s) \in \mathbb{R}^2$, solving

$$i\frac{d\theta}{dt} = K(t)\theta(t), \quad \theta(s) = \theta_s \in Y, \tag{2.11}$$

where

$$Y = \{f \in L^2(\mathbb{R}^3) | \Delta f, (1 + x_1^2)^{1/2} f \in L^2(\mathbb{R}^3)\} \tag{2.12}$$

in the case of (1.1) and $Y = D(H_0) = H^2(\mathbb{R}^3)$ for the other two equations. Moreover

$$U_K(t,s)(Y) \subseteq Y \tag{2.13}$$

in all three cases and the propagators are related by

$$\begin{aligned} U_{A^\varepsilon}(t,s) &= T^\varepsilon(t)^{-1} U_{B^\varepsilon}(t,s) T^\varepsilon(s) \\ &= T^\varepsilon(t)^{-1} V^\varepsilon(t)^{-1} U_{H^\varepsilon}(t,s) V^\varepsilon(s) T^\varepsilon(s), \end{aligned} \tag{2.14}$$

with $T^\varepsilon(t)$, $V^\varepsilon(t)$, $t \in \mathbb{R}$, given by

$$(T^\varepsilon(t)f)(x) = \exp(ih^\varepsilon(t)x_1)f(x), \tag{2.15}$$

$$(V^\varepsilon(t)f)(x) = \exp(ik^\varepsilon(t))f(x_1 - 2G^\varepsilon(t),x^\perp), \tag{2.16}$$

for all $f \in L^2(\mathbb{R}^3)$.

Finally in the remainder of this paper we will need the following limiting properties of the auxiliary functions introduced at the beginning of this section.

Lemma 2.2: Let g, h, G, k, g^ε, h^ε, G^ε, k^ε be as above. Then, (i) for each fixed $\varepsilon > 0$ we have

$$\lim_{t \to \pm\infty} h^\varepsilon(t) = \lim_{t \to \pm\infty} G^\varepsilon(t) = \lim_{t \to \pm\infty} k^\varepsilon(t) = 0; \tag{2.17}$$

(ii) for each fixed $t \in \mathbb{R}$, we have

$$\lim_{\varepsilon \downarrow 0} h^\varepsilon(t) = h(t), \quad \lim_{\varepsilon \downarrow 0} G^\varepsilon(t) = G(t), \tag{2.18}$$

$$\lim_{\varepsilon \downarrow 0} (k^\varepsilon(t) - k^\varepsilon(s)) = k(t) - k(s), \tag{2.19}$$

$$\lim_{\varepsilon \downarrow 0} k^\varepsilon(t) = \begin{cases} -\infty, & \text{if} \quad t > 0, \\ \infty, & \text{if} \quad t < 0. \end{cases} \tag{2.20}$$

Proof: We will concentrate on the case $t \geqslant 0$. Similar arguments hold for $t < 0$. The limits in (2.17) and (2.18) follow by combining (2.3) and (2.4) in order to obtain the estimate

$$|h^\varepsilon(t)| \leqslant \varepsilon^{-1} e^{-\varepsilon t} \|g\|_\infty, \quad \forall t \geqslant 0, \tag{2.21}$$

where $\|\cdot\|_\infty$ denotes the L^∞ norm.

Next, using $h' = g$, $G' = h$, and integrating by parts twice we obtain

$$h^\varepsilon(t) = -\int_t^\infty \exp(-\varepsilon s)h'(s)ds$$
$$= e^{-\varepsilon t}h(t) + \varepsilon \int_t^\infty \exp(-\varepsilon s)G'(s)ds$$
$$= e^{-\varepsilon t}h(t) + \varepsilon e^{-\varepsilon t}G(t) + \varepsilon^2 \int_t^\infty \exp(-\varepsilon s)G(s)ds$$
$$= e^{-\varepsilon t}h(t) + \varepsilon e^{-\varepsilon t}G(t) + \varepsilon \int_{\varepsilon t}^\infty e^{-\theta}G\left(\frac{\theta}{\varepsilon}\right)d\theta. \quad (2.22)$$

Since G is a bounded function, the integral in the last member of (2.22) can be estimated by $\|G\|_\infty \exp(-\varepsilon t)$ and the first limit in (2.18) follows at once. In order to prove the second, note that since $G' = h$ the fourth equality in (2.22) implies

$$G^\varepsilon(t) = e^{-\varepsilon t}G(t) + \varepsilon^2 \int_t^\infty ds \int_s^\infty du \exp(-\varepsilon u)G(u)$$
$$= e^{-\varepsilon t}G(t) + \varepsilon^2 \int_t^\infty du(u-t)\exp(-\varepsilon u)G(u), \quad (2.23)$$

and the result follows in the same way as the previous one. The only difference is that to control the integral of $u\exp(-\varepsilon u)G(u)$, we must use another function H, periodic with mean value zero such that $H' = G$, and integrate by parts in order to get the factor ε^3 where we need it. Equation (2.19) is an easy consequence of the dominated convergence theorem. We now turn to (2.20), which is by far the hardest part. From the third equality in (2.22) we get

$$h^\varepsilon(t)^2 = e^{-2\varepsilon t}h(t)^2 + 2\varepsilon e^{-\varepsilon t}h(t)\int_t^\infty \exp(-\varepsilon s)h(s)ds + \varepsilon^2\left(\int_t^\infty \exp(-\varepsilon s)h(s)ds\right) \times\left(\int_t^\infty \exp(-\varepsilon u)h(u)du\right). \quad (2.24)$$

It is easy to see that after integration the last two terms of the right-hand side of (2.24) tend to zero as $\varepsilon \downarrow 0$. Thus it remains to show that

$$\int_t^\infty e^{-2\varepsilon s}(h(s))^2 ds \to \infty, \quad \text{as } \varepsilon\downarrow 0. \quad (2.25)$$

To do this let $\alpha = \sup_{s\in\mathbf{R}} h(s)^2$ and write

$$X_e = \{s\in[t,\infty)\,|\,h(s)^2 < \alpha/2\},$$
$$X_r = \{s\in[t,\infty)\,|\,h(s)^2 \geqslant \alpha/2\}, \quad (2.26)$$

so that $X_e \cap X_r$ is empty and $[t,\infty) = X_e \cup X_r$. Then,

$$\int_t^\infty e^{-\varepsilon s}(h(s))^2 ds \geqslant \int_{X_r} e^{-\varepsilon s}(h(s))^2 ds \geqslant \frac{\alpha}{2}\int_{X_r} e^{-\varepsilon s} ds. \quad (2.27)$$

But as $\varepsilon\downarrow 0$ the integral in the right-hand side of (2.27) tends to the Lesbegue measure $|X_r|$ of X_r, which is infinite since $(h(s))^2$ is periodic, non-negative, and nontrivial. This completes the proof. Q.E.D.

III. SCATTERING THEORY WITH $\varepsilon > 0$

The purpose of this section is to relate the asymptotic behavior of $U_{A^\varepsilon}(t,s)$ and $\exp(-i(t-s)H)$ as $t \to \pm\infty$, with both $\varepsilon > 0$ and $s \in \mathbf{R}$ kept fixed. In order to accomplish this it is convenient to establish a series of preliminary results, the first one of which is the following theorem.

Theorem 3.1: Assume that $q(x)$ satisfies

$$q(x) = (1+|x|^2)^{-\rho}q_1(x) + q_2(x), \quad \rho > \tfrac{1}{2},$$
$$q_1 \in L^\infty(\mathbf{R}^3), \quad q_2 \in L^2(\mathbf{R}^3). \quad (3.1)$$

Then the wave operators

$$\Omega^\varepsilon_\pm(A^\varepsilon, A_0^\varepsilon; s) = \operatorname*{s-lim}_{t\to\pm\infty} U_{A^\varepsilon}(t,s)^* U_{A_0^\varepsilon}(t,s) \quad (3.2)$$

exist, where the right-hand side of (3.2) denotes, as usual, the strong limit in $L^2(\mathbf{R}^3)$.

Proof: We will consider only the limit as $t\to\infty$. The existence of the other follows from similar arguments. Furthermore, since the main idea involved here, namely the Cook–Kuroda method, is by now standard, we will just indicate the estimates involved. Note that in order to prove (3.2) (with $t \to +\infty$), it is enough to show that

$$\int_a^\infty \|qU_{A_0^\varepsilon}(t,s)\phi\|_{L^2}\, dt < \infty, \quad (3.3)$$

for some $a > s$ (which is fixed) and all $\phi \in C_0^\infty(\mathbf{R}^3)$. To do this we use (2.14)–(2.16) to write $U_{A_0^\varepsilon}(t,s)$ in terms of $\exp(-i(t-s)H_0)$ as follows:

$$U_{A_0^\varepsilon}(t,s) = \exp[-i(h^\varepsilon(t)x_1 + k^\varepsilon(t))] \times\exp(-i(t-s)H_0)S_{2G^\varepsilon(t)-2G^\varepsilon(s)} \times\exp(ih^\varepsilon(s)(x_1)). \quad (3.4)$$

If $q\in L^2(\mathbf{R}^3)$, (3.4) implies that

$$|q(x)(U_{A_0^\varepsilon}(t,s)\phi)(x)| \leqslant C|t-s|^{-3/2}\|\phi\|_{L^1}|q(x)|, \quad (3.5)$$

where $C > 0$ is a constant. Integrating this inequality over $\mathbf{R}^3$, we obtain the estimate needed to prove (2.3) in this case. Next if $q(x) = (1+|x|^2)^{-\rho}q_1(x)$, $q_1 \in L^\infty(\mathbf{R}^3)$, it is enough to consider $\frac{1}{2} < \rho \leqslant \frac{3}{4}$, since otherwise $q\in L^2(\mathbf{R}^3)$ and there is nothing to prove in view of the previous remarks. Using the fact that the first factor on the right-hand side of (3.4) commutes with multiplications and choosing p, $\tilde{p}$, r such that $r\in(3/2\rho,3)$, $\tilde{p}^{-1} + r^{-1} = 2^{-1}$, $p^{-1} + \tilde{p}^{-1} = 1$, we can apply the Riesz–Thorin theorem[9] to conclude

$$\|qU_{A_0^\varepsilon}(t,s)\phi\|_2 \leqslant C\|q_1\|_{L^\infty}|t-s|^{-3/r}\|\phi\|_p, \quad (3.6)$$

where $C > 0$ is again a constant. Since $3/r > 1$, the result follows also in this case and the proof is complete. Q.E.D.

Note that the estimates in (3.5) and (3.6) are independent of ε, and therefore they also hold in the case $\varepsilon = 0$ [in particular the wave operators $\Omega_\pm(A^0, A_0^0; s)$ exist under the assumptions made in Theorem 3.1; this result is stronger than the corresponding existence theorem in Sec. 5 of Ref. 1]. This remark will be used in Sec. IV.

Next we prove a technical lemma that will be useful in the remainder of this section and has some interest of its own.

Lemma 3.2: Assume that $q(x)$ satisfies (1.5) and let $H = H_0 + q$. Then

$$\underset{t\to\pm\infty}{\text{s-lim}}\ e^{i(t-s)H}T^\varepsilon(t)e^{-i(t-s)H} = 1, \tag{3.7}$$

for all $s\in\mathbb{R}$, where 1 denotes the identity operator in $L^2(\mathbb{R}^3)$.

Proof: Without loss of generality we will assume that $s = 0$. It is well known that under assumption (1.5) the following decomposition holds[10,11]:

$$L^2(\mathbb{R}^3) = \mathscr{H}_p(H) \oplus \mathscr{H}_{ac}(H), \tag{3.8}$$

where $\mathscr{H}_p(H)$ [resp. $\mathscr{H}_{ac}(H)$] denotes the pure point (resp. absolutely continuous) subspace of $L^2(\mathbb{R}^3)$ with respect to H (for the definition of these objects see Refs. 11 or 12). In order to prove the results we will show that the limit exists in the two subspaces on the right-hand side of (3.8). We start with $\mathscr{H}_p(H)$. Let $f\in\mathscr{D}(H) = \mathscr{D}(H_0)$ be such that $Hf = \lambda f$ for some $\lambda\in\mathbb{R}$. Then

$$\|e^{itH}T^\varepsilon(t)e^{-itH}f - f\|_{L^2} = \|T^\varepsilon(t)f - f\|_{L^2}, \tag{3.9}$$

and the right-hand side of (3.9) tends to zero as $t\to\infty$, since by the dominated convergence theorem we have

$$\underset{t\to\pm\infty}{\text{s-lim}}\ T^\varepsilon(t) = 1. \tag{3.10}$$

Using a simple approximation argument we obtain the result in $\mathscr{H}_p(H)$. Next we turn to $\mathscr{H}_{ac}(H)$. We will consider only $t\to+\infty$. The other case can be treated similarly. Recall from usual potential scattering that given $f\in\mathscr{H}_{ac}(H)$ there exists a unique $\varphi_+\in L^2(\mathbb{R}^3)$ such that

$$\lim_{t\to\infty}\|e^{-itH}f - e^{-itH_0}\varphi_+\|_{L^2} = 0. \tag{3.11}$$

For a proof of this statement we refer the reader to Ref. 10 and/or Ref. 11. Adding and subtracting the appropriate quantities, using the triangle inequality and the unitarity of e^{itH} and $T^\varepsilon(t)$ for each $t\in\mathbb{R}$, we obtain

$$\|e^{itH}T^\varepsilon(t)e^{-itH}f - f\|_{L^2} \leqslant 2\|e^{-itH}f - e^{-itH_0}\varphi_+\|_{L^2} + \|(T^\varepsilon(t) - 1)e^{-itH_0}\varphi_+\|_{L^2}. \tag{3.12}$$

In view of (3.11) it remains to show that the second term on the right-hand side of (3.12) tends to zero as $t\to\infty$. Let $\hat\theta$ denote the Fourier transform of $\theta\in L^2(\mathbb{R}^3)$ (for details, see Ref. 9). Given $\delta > 0$ choose $\hat\theta\in C_0^\infty(\mathbb{R}^3)$ such that $\|\theta - \varphi_+\|_{L^2} < \delta$. Then

$$\|(T^\varepsilon(t) - 1)e^{-itH_0}\varphi_+\|^2_{L^2} \leqslant (\|(T^\varepsilon(t) - 1)e^{-itH_0}\theta\|_{L^2} + \delta)^2, \tag{3.13}$$

so that it suffices to prove that $\|(T^\varepsilon(t) - 1)e^{-itH_0}\theta\|_{L^2}$ tends to zero as $t\to\infty$. Applying Parseval's identity[9] we obtain

$$\|(T^\varepsilon(t) - 1)e^{-itH_0}\theta\|^2_{L^2} = \int_{\mathbb{R}^3} |E(\xi_1 - h^\varepsilon(t),\xi^\perp,t) - E(\xi,t)|^2\, d\xi, \tag{3.14}$$

where $E(\xi,t) = \exp(-it\xi^2)\hat\theta(\xi)$. Since $h^\varepsilon(t)$ is a bounded function, it is easy to see that the integrand in the rhs of (3.14) has compact support. But then the dominated convergence theorem implies the result because according to (2.21) both $h^\varepsilon(t)$ and $th^\varepsilon(t)$ tend to zero as $t\to\infty$. Q.E.D.

As a consequence of Lemma 3.2 we obtain the following important result which relates the asymptotic behaviors of the propagators $U_{A_0^\varepsilon}(t,s)$ and $\exp(-i(t-s)H_0)$.

Corollary 3.3: Let $U_{A_0^\varepsilon}(t,s)$ and H_0 be as in Theorem 2.1 (with $q = 0$). Then

$$\begin{aligned}\Omega^\varepsilon_\pm(A_0^\varepsilon,H_0;s) &= \underset{t\to\pm\infty}{\text{s-lim}}\ U_{A_0^\varepsilon}(t,s)^* e^{-i(t-s)H_0}\\ &= V^\varepsilon(s)^{-1}T^\varepsilon(s)^{-1}\\ &= \exp(-ik^\varepsilon(s))\exp(-ih^\varepsilon(s)x_1)S_{2G^\varepsilon(s)},\end{aligned} \tag{3.15}$$

where S_a, $a\in\mathbb{R}$, is given by

$$(S_a f)(x) = f(x_1 + a,x^\perp),\quad f\in L^2(\mathbb{R}^3). \tag{3.16}$$

In particular the operators $\Omega_\pm(A_0^\varepsilon,H_0;s)$ are unitary.

Proof: Applying (2.14) with $q = 0$ and noting that $V^\varepsilon(t)$ commutes with $\exp(-i(t-s)H_0)$ we obtain

$$U_{A_0^\varepsilon}(t,s)^*\exp(-i(t-s)H_0) = T^\varepsilon(s)^{-1}V^\varepsilon(s)^{-1}V^\varepsilon(t)e^{i(t-s)H_0}T^\varepsilon(t)e^{-i(t-s)H_0}, \tag{3.17}$$

and the result follows at once from Lemma 3.2 and part (i) of Lemma 2.2 which implies that $\text{s-lim}_{t\to\pm\infty} V^\varepsilon(t) = 1$. Q.E.D.

Combining Theorem 3.1 and Corollary 3.3 it is easy to prove the following Corollary.

Corollary 3.4: Let q be as in Theorem 3.1. Then the wave operators

$$\Omega^\varepsilon_\pm(A^\varepsilon,H_0;s) = \underset{t\to\pm\infty}{\text{s-lim}}\ U_{A^\varepsilon}(t,s)^* e^{-i(t-s)H_0} \tag{3.18}$$

exist.

We are now in position to state and prove the main result of this section, namely, the following Theorem.

Theorem 3.5: Assume that q satisfies condition (1.5). Let $H = H_0 + q$ and $A^\varepsilon(t)$ be as in Theorem 2.1. Then the limits

$$\Gamma_\pm(A^\varepsilon,H;s) = \underset{t\to\pm\infty}{\text{s-lim}}\ U_{A^\varepsilon}(t,s)^* e^{-i(t-s)H} \tag{3.19}$$

exist and are unitary. Moreover the following intertwining relations holds:

$$U_{A^\varepsilon}(t,s)\Gamma_\pm(A^\varepsilon,H;s) = \Gamma_\pm(A^\varepsilon,H;t)e^{-i(t-s)H}. \tag{3.20}$$

Proof: In view of the first equality in (2.14) we may write $\Gamma(t) = U_{A^\varepsilon}(t,s)^*e^{-i(t-s)H}$ as

$$T(t) = T^\varepsilon(s)^{-1}(U_{B^\varepsilon}(t,s)^*e^{-i(t-s)H}) \times e^{i(t-s)H}T^\varepsilon(t)e^{-i(t-s)H}, \tag{3.21}$$

where $B^\varepsilon(t)$ is as in Theorem 2.1. Due to Lemma 3.2 [and the uniform boundness of all the factors in (3.21) with respect to t], it is enough to show that the limits

$$\Gamma_\pm(B^\varepsilon,H;s) = \underset{t\to\pm\infty}{\text{s-lim}}\ U_{B^\varepsilon}(t,s)^* e^{-i(t-s)H} \tag{3.22}$$

exist and are unitary, since, in this case,

$$\lim_{t\to\pm\infty} \Gamma(t) = \Gamma_{\pm}(A^{\epsilon},H;s) = T^{\epsilon}(s)^{-1}\Gamma_{\pm}(B^{\epsilon},H;s), \tag{3.23}$$

which are obviously unitary. In order to obtain (3.22), we remark first that, as is well known (Sec. 3 of Ref. 1), we have

$$D(B^{\epsilon}(t)) = D(H) = D(H_0), \tag{3.24}$$

for all $t\in\mathbb{R}$. Let $\mathscr{G}$ denote $D(H_0)$ provided with the graph norm $|\,\|f\|\,| = \|f\|^2_{L^2} + \|H_0 f\|^2_{L^2}$ and let $\mathscr{B} = B(\mathscr{G},L^2(\mathbb{R}^3))$ denote the set of all bounded operators from $\mathscr{G}$ into $L^2(\mathbb{R}^3)$. Then it is easy to verify that

$$\int_{\mathbb{R}} \|B^{\epsilon}(t) - H\|_{\mathscr{B}}\, dt < \infty, \tag{3.25}$$

$$\operatorname*{Var}_{\mathbb{R}}(B(\cdot)) = \sup \sum_{0\leqslant j\leqslant n-1} \|B^{\epsilon}(t_{j+1}) - B^{\epsilon}(t_j)\|_{\mathscr{B}} < \infty, \tag{3.26}$$

where the supremun is taken over all finite real sequences $t_0 < t_1 < t_2 < \cdots < t_n$. Under these conditions Theorem 6 of Ref. 13 implies that the operators in (3.22) exist and have the stated properties. The proof of the intertwining relations is standard and will be omitted (see Chap. X of Ref. 12, where the proof is presented in the case of time-independent Hamiltonians; the same idea works in our case). Q.E.D.

A few remarks are now in order. Let $\varphi\in\mathscr{H}_p(H)$. Then if $f_{\pm} = \Gamma_{\pm}(A^{\epsilon},H;s)\varphi$, we have

$$\lim_{t\to\pm\infty} \|U_{A^{\epsilon}}(t,s)f_{\pm} - e^{-i(t-s)H}\varphi\|_{L^2} = 0, \tag{3.27}$$

and it is easy to see that the wave functions $\psi_{\pm}(t) = U_{A^{\epsilon}}(t,s)f_{\pm}$ behave as bound states as $t\to\pm\infty$. More precisely, the probability of finding the particle in $\{|x|\geqslant R\}$ at time t can be estimated as follows:

$$\begin{aligned} P(t,\{|x|\geqslant R\};f_{\pm}) &= \|f_{\pm}\|^{-2}\int_{\mathbb{R}^3} |\chi_{\{|x|>R\}}(x) \\ &\quad\times (U_{A^{\epsilon}}(t,s)f_{\pm})(x)|^2\, dx \\ &\leqslant \|f_{\pm}\|^{-2}(\|\chi_{\{|x|>R\}} e^{-i(t-s)H}\varphi\| \\ &\quad + \|U_{A^{\epsilon}}(t,s)f_{\pm} - e^{-i(t-s)H}\varphi\|)^2, \end{aligned} \tag{3.28}$$

where χ_s is the characteristic function of the set S. Thus, given $\eta>0$, there exist $t_0>0$ and $R_0>0$ such that if $|t|>t_0$ and $R>R_0$ then $P(t;\{|x|\geqslant R\};f_{\pm}) < \eta$. This means that the particle is asymptotically (as $t\to\pm\infty$) in a bound state. Moreover, it can also easily be shown that if $\varphi\in\mathscr{H}_{ac}(H)$, then $f_{\pm} = \Gamma_{\pm}(A^{\epsilon},H;s)\varphi$ are such that $\psi_{\pm} = U_{A^{\epsilon}}(t,s)f_{\pm}$ behave as scattering states as $t\to\pm\infty$.

In view of the remarks just made, Theorem 3.5 and Eq. (3.8) imply two decompositions of $L^2(\mathbb{R}^3)$ into (asymptotic) bound state and scattering subspaces, namely,

$$\begin{aligned} L^2(\mathbb{R}^3) &= \Gamma_{\pm}(A^{\epsilon},H;s)(\mathscr{H}_{ac}(H)) \\ &\quad \oplus \Gamma_{\pm}(A^{\epsilon},H;s)(\mathscr{H}_p(H)). \end{aligned}$$

It should be remarked, however, that, as far as we know, it is an open question whether or not the above decompositions coincide.

IV. THE ADIABATIC LIMIT

In this section we will be concerned with the asymptotic behavior (in time) of the solution (1.1) as $\epsilon\downarrow 0$. The first thing to notice is that it is hopeless to take the "limit of the theory" established for $\epsilon>0$. This is already apparent in Corollary 3.3. Indeed, in view of (3.15) and the behavior of $G^{\epsilon}(s)$, $h^{\epsilon}(s)$, and $k^{\epsilon}(s)$, described in Lemma 2.2, it follows that $\Omega_{\pm}(A^{\epsilon}_0,H_0;s)$ does not have a limit as $\epsilon\downarrow 0$. This also indicates what the problem is and points the way to the correct definitions. Let $\epsilon\geqslant 0$ and introduce

$$\begin{aligned} \Lambda^{\epsilon}(t,s) &= e^{ik^{\epsilon}(s)}T^{\epsilon}(t)^{-1}V^{\epsilon}(t)^{-1} \\ &= \exp[-i(k^{\epsilon}(t) - k^{\epsilon}(s))] \\ &\quad\times \exp[-ih^{\epsilon}(t)x_1]S_{2G^{\epsilon}(t)}. \end{aligned} \tag{4.1}$$

Define the modified wave operators for the pair $(A^{\epsilon}(\cdot),H_0)$ by

$$W_{\pm}(A^{\epsilon},H_0;s) = \operatorname*{s-lim}_{t\to\pm\infty} U_{A^{\epsilon}}(t,s)^{*}\Lambda^{\epsilon}(t,s)e^{-i(t-s)H_0}, \tag{4.2}$$

if the limits exist.

In what follows we will show that they indeed exist for $\epsilon\geqslant 0$ and that (4.2) is continuous in ϵ up to $\epsilon=0$. We begin with the case $q=0$, which is trivial. Applying (2.15) to write $U_{A^{\epsilon}_0}(t,s)$ in terms of $\exp(-i(t-s)H_0)$ and using the definition of $\Lambda^{\epsilon}(t,s)$, we obtain

$$\begin{aligned} &U_{A^{\epsilon}_0}(t,s)^{*}\Lambda^{\epsilon}(t,s)e^{-i(t-s)H_0} \\ &\quad = T^{\epsilon}(s)^{-1}(e^{k^{\epsilon}(s)}V^{\epsilon}(s))^{-1} = T^{\epsilon}(s)^{-1}S_{-2G^{\epsilon}(s)}, \end{aligned} \tag{4.3}$$

for all $\epsilon\geqslant 0$. Note that this expression is independent of t! This means that the modification just introduced cancels out the oscillations responsible for nonexistence of the limit of $\Omega_{\pm}(A^{\epsilon}_0,H_0;s)$ as $\epsilon\downarrow 0$, uniformly in t. It should also be noted that $\Lambda^{\epsilon}(t,s)e^{-i(t-s)H_0}$ is a "modified free evolution" in the sense that

$$\lim_{t\to\pm\infty}\int_S |\Lambda^{\epsilon}(t,s)e^{-i(t-s)H_0}f(x)|^2\, dx = 0,$$

for all bounded measurable $S\subseteq\mathbb{R}^3$ and $f\in L^2(\mathbb{R}^3)$.

In order to proceed, we will assume from now on that q satisfies (1.5) and (1.6). In this case, as shown in Sec. 5 of Ref. 1, the wave operators

$$\Omega_{\pm}(A^0,A^0_0;s) = \operatorname*{s-lim}_{t\to\pm\infty} U_{A^0}(t,s)^{*}U_{A^0_0}(t,s) \tag{4.4}$$

exist and are complete in the sense of (1.7) and (1.8) for all $s\in\mathbb{R}$, where $\Theta(s)$ is the Floquet (or period) operator of the system, namely,

$$\Theta(s) = U_{A^0}(s+\tau,s), \quad s\in\mathbb{R}. \tag{4.5}$$

With these remarks in mind, we have the following theorem.

Theorem 4.2: Let q satisfy (1.5) and (1.6). Then

$$\Omega_{\pm}(A^0,A^0_0;s) = \operatorname*{s-lim}_{\epsilon\downarrow 0}\Omega_{\pm}(A^{\epsilon},A^{\epsilon}_0;s). \tag{4.6}$$

Proof: We will consider the case of $\Omega_{+}(A^0,A^0_0;s)$. The other limit can be handled similarly. Moreover, since all operators involved are uniformly bounded with respect to $\epsilon\geqslant 0$, it is enough to prove that the limit exists in $C_0^\infty(\mathbb{R}^3)$. Thus if φ is any such function, we have

$$\|\Omega_+(A^0,A_0^0;s)\varphi - \Omega_+(A^\varepsilon,A_0^\varepsilon;s)\varphi\|_{L^2} \leqslant \|\Omega_+(A^0,A_0^0;s)\varphi - U_{A^0}(t,s)^*U_{A_0^0}(t,s)\varphi\|_{L^2} + \|U_{A^0}(t,s)^*U_{A_0^0}(t,s)\varphi - U_{A^\varepsilon}(t,s)^*U_{A_0^\varepsilon}(t,s)\varphi\|_{L^2} + \|U_{A^\varepsilon}(t,s)^*U_{A_0^\varepsilon}(t,s)\varphi - \Omega_+(A^\varepsilon,A_0^\varepsilon;s)\varphi\|_{L^2}, \qquad (4.7)$$

for all $t\in\mathbb{R}$. According to the remark following the proof of Theorem 3.1, the first and third terms in the right-hand side of (4.7) can be estimated as follows:

$$\|\Omega_+(A^\varepsilon,A_0^\varepsilon;s)\varphi - U_{A^\varepsilon}(t,s)^*U_{A_0^\varepsilon}(t,s)\varphi\|_{L^2} \leqslant \int_t^\infty \|qU_{A_0^\varepsilon}(r,s)\varphi\|_{L^2}\,dr \leqslant C\Big(\|q_1\|_{L^\infty}\|\varphi\|_p \int_t^\infty |u-s|^{-3/r}\,du + \|q_2\|_2\|\varphi\|_{L^1}\int_t^\infty |u-s|^{-3/2}\,du\Big), \qquad (4.8)$$

where $\varepsilon\geqslant 0$, $t>s$, and C is a constant independent of ε. Since the last member of (4.8) tends to zero as $t\to\infty$, it remains to show that the second term on the rhs of (4.7) tends to zero as $\varepsilon\downarrow 0$. In order to do this note that the differential equation satisfied by the propagators in question implies

$$U_{A^\varepsilon}(t,s)^*\varphi = U_{A^0}(t,s)^*\varphi + i\int_s^t U_{A^\varepsilon}(r,s)^* \times (e^{-\varepsilon r}-1)\cdot g(t)x_1U_{A^0}(t,r)dr. \qquad (4.9)$$

Before proceeding it should be remarked that $x_1U_{A^0}(t,r)\varphi$ belongs to $L^2(\mathbb{R}^3)$ and depends continuously in t because of (2.13). Then

$$\|U_{A^\varepsilon}(t,s)^*\varphi - U_{A^0}(t,s)^*\varphi\| \leqslant \int_s^t |e^{-\varepsilon r}-1|\,\|x_1U_{A^0}(t,r)^*\varphi\|dr \qquad (4.10)$$

and the rhs tends to zero as $\varepsilon\downarrow 0$ by the dominated convergence theorem. This completes the proof. Q.E.D.

We now turn to the main result of this section, namely, the following theorem.

Theorem 4.3: Let q satisfy (1.5) and (1.6). Then the wave operators $W_\pm(A^\varepsilon,H_0;s)$ exist of all $\varepsilon\geqslant 0$. If $\varepsilon>0$, they are given by

$$W_+(A^\varepsilon,H_0;s) = \Omega_\pm(A^\varepsilon,A_0^\varepsilon;s)T^\varepsilon(s)^{-1}S_{-2G^\varepsilon(s)}, \qquad (4.11)$$

while if $\varepsilon=0$, we have

$$W_\pm(A^0,H_0;s) = \lim_{\varepsilon\downarrow 0} W_\pm(A^\varepsilon,A_0^\varepsilon;s) = \Omega_\pm(A^0,A_0^0;s)T(s)^{-1}S_{-2G(s)}. \qquad (4.12)$$

In particular,

$$\mathscr{R}(W_\pm(A^0,H_0;s)) = \mathscr{R}(\Omega_\pm(A^0,A_0^0;s)) = \mathscr{H}_{ac}(\Theta(s)), \qquad (4.13)$$

where $\Theta(s)$ is the Floquet operator defined in (4.5).

Proof: Due to (4.3), we can write

$$U_{A^\varepsilon}(t,s)^*\Lambda^\varepsilon(t,s)e^{-i(t-s)H_0} = U_{A^\varepsilon}(t,s)^*U_{A_0^\varepsilon}(t,s)T^\varepsilon(s)^{-1}S_{-2G^\varepsilon(s)}, \qquad (4.14)$$

for all $t\in\mathbb{R}$ and $\varepsilon\geqslant 0$. Taking the limit as $t\to\pm\infty$, we obtain (4.11) and the second equality in (4.12). Next recall that in the proof of Theorem 4.2 we have shown that

$$\text{s-lim}_{\varepsilon\downarrow 0}\, U_{A^\varepsilon}(t,s)^*U_{A_0^\varepsilon}(t,s) = U_{A^0}(t,s)^*U_{A_0^0}(t,s)$$

[see the second term on the rhs of (4.7)]. Therefore

$$\text{s-lim}_{t\to\pm\infty}\,\text{s-lim}_{\varepsilon\downarrow 0}\, U_{A^\varepsilon}(t,s)^*U_{A_0^\varepsilon}(t,s)T^\varepsilon(s)^{-1}S_{-2G^\varepsilon(s)} = \text{s-lim}_{t\to\pm\infty}\, U_{A^0}(t,s)^*U_{A_0^0}(t,s)T(s)^{-1}S_{-2G(s)} = \Omega_\pm(A^0,A_0^0;s)T(s)^{-1}S_{-2G(s)} = W_\pm(A^0,A_0^0;s), \qquad (4.15)$$

since we already know that the last equality in (4.15) holds. The statement about the Floquet operator and the ranges of the wave operators follows from (4.6) and the proof is complete. Q.E.D.

We will now make some final remarks on the results presented above. First of all, it is natural to ask what is the relation between the modified and usual theories when $\varepsilon>0$. The answer, which is not difficult to obtain, is given by the relation

$$\widetilde{\Gamma}_\pm(A^\varepsilon,H;s) = \text{s-lim}_{t\to\pm\infty}\, U_{A^\varepsilon}(t,s)^*\Lambda^\varepsilon(t,s)e^{-i(t-s)H} = e^{-ik^\varepsilon(s)}\Gamma_\pm(A^\varepsilon,H;s), \qquad (4.16)$$

where $\Gamma_\pm(A^\varepsilon,H;s)$ is the operator defined in Theorem 3.5. Thus we obtain two decompositions of $L^2(\mathbb{R}^3)$ into scattering and (asymptotic) bound states which are exactly the same as before except for a phase [which does not have a limit as $\varepsilon\downarrow 0$; see (2.20)]. In particular, we do not know if the decompositions coincide. In the limit, however, the results of this section show that we can construct a satisfactory scattering theory. In this case we may have bound states in the usual time-dependent sense[1,11] and the set of scattering states is exactly the same as those obtained in Ref. 1 using $U_{A^0}(t,s)$ as the free evolution.

ACKNOWLEDGMENTS

The authors would like to thank Professor L. Davidovich and Professor D. Marchesin of Catholic University of Rio de Janeiro (PUC/RJ) for several enlightening conversations.

[1]R. J. Iorio, Jr. and D. Marchesin, Proc. R. Soc. Edinburgh A **96**, 117 (1984).

[2]T. Kato (Private communication).

[3]H. Kitada and K. Yajima, Duke Math. J. **49**, 341 (1982).

[4]H. S. Antunes Neto, L. Davidovich, and D. Marchesin, Preprint PUC/RJ, to appear.

[5]H. S. Brandi, L. Davidovich, and N. Zaguri, Phys. Rev. A **24**, 2044 (1981).

[6]H. S. Antunes Neto and L. Davidovich, Phys. Rev. Lett. **53**, 2238, 2241 (1984).

[7]J. D. Dollard, J. Math. Phys. 7, 809 (1966).
[8]T. Kato, J. Fac. Sci. Univ. Tokyo, Sec. I, **17**, 241 (1970).
[9]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[10]S. Agmon, Ann. Scuola Norm. Sup. Pisa Cl. Sci. **II**, 151 (1975).
[11]W. Amrein, J. Jauch, and K. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, New York, 1977).
[12]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[13]G. Schmidt, Indiana Univ. Math. J. **24**, 925 (1975).

Padé oscillators and a new formulation of perturbation theory

M. Znojil
Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Řež, Czechoslovakia

(Received 3 May 1988; accepted for publication 10 August 1988)

Being guided by the problem of bound states in potentials close to their Padé approximants, a new Rayleigh–Schrödinger-type perturbation theory is developed. The unperturbed system is understood here in a broader sense: its solutions are not needed, but merely the related nondiagonal unperturbed propagator R. In particular, all the chain models $H_0\psi = ES_0\psi$ (H_0, S_0 = band matrices) with arbitrary perturbations are then perturbatively solvable, with R constructed in terms of auxiliary matrix continued fractions f_n. Alternatively, a "generalized unperturbed spectrum" $\hat{f}_n$ may be required as an input: The algebraically constructed asymptotics of the f_n's play this role in our Padé examples. Due to $S \neq I$, the "Sturmians" may also be constructed. In the test evaluations of the binding energies and/or couplings, the simultaneous upper and lower bounds of high precision are shown to be numerically obtainable.

I. INTRODUCTION

Usually, the textbook[1] Rayleigh–Schrödinger (RS) perturbation theory is applied to the Hamiltonian matrices that are almost diagonal,

$$\langle m|H|n\rangle = \begin{cases} \text{large}, & m = n, \\ \text{small}, & m \neq n. \end{cases} \tag{1.1}$$

Recently, a weaker assumption

$$\langle m|H|n\rangle = \begin{cases} \text{large}, & |m-n| \leqslant t, \\ \text{small}, & |m-n| > t, \quad t \geqslant 0, \end{cases} \tag{1.2}$$

has been shown to be tractable in the same spirit.[2] Here, we intend to describe a generalizaton of the latter modified RS (MRS) perturbation theory to an even more flexible class of Hamiltonians.

As an inspiration, let us recall the radial Schrödinger one-body equation

$$\left[-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r) \right]\psi(r) = E\psi(r), \quad l = 0,1,..., \tag{1.3}$$

with the often-studied potential[3]

$$V(r) = \frac{a + br^2 + cr^4}{1 + dr^2} \sim r^2 + \frac{\mu r^2}{1+dr^2}, \quad c > 0, \quad d > 0, \tag{1.4}$$

all of the matrix elements of which are nonzero in the standard harmonic oscillator basis[4] [i.e., $t \gg 1$ in (1.2)]. Nevertheless, an elementary modification of Eqs. (1.3) and (1.4),

$$[S(r)H^{(\text{oscillator})} + \mu r^2]\psi(r) = ES(r)\psi(r), \quad S(r) = 1 + dr^2 > 0, \tag{1.5}$$

acquires the infinite tridiagonal-matrix form

$$\sum_{n=m-1}^{m+1} [\langle m|(SH^{(\text{oscillator})} + \mu r^2)|n\rangle - E\langle m|S|n\rangle]\langle n|\psi\rangle = 0, \quad m = 0,1,..., \tag{1.6}$$

in the same basis, with $S \neq I$ but, at the same time, with the extremely simple form of matrix elements (Whitehead *et al.*[5]).

Our present paper will be devoted to all the generalized eigenvalue problems[6] of similar type:

$$(H - ES)|\psi\rangle = 0, \quad S \neq I. \tag{1.7}$$

For the sake of definiteness, we shall assume that

$$H = T + \lambda W, \quad S = S_0 + \lambda S_1, \tag{1.8}$$

where both T and S_0 are band matrices and $\lambda \ll 1$.

In the first, general part of our study (Sec. II), we shall introduce a rearrangement of Eq. (1.8),

$$H = H_0 + \lambda H_1, \quad H_0 = T + U, \quad \lambda H_1 = \lambda W - U, \tag{1.9}$$

where $U = O(\lambda)$ will be specified in a "self-consistent" way. Then we shall describe the related generalized RS (GRS) construction of expansions,

$$\begin{aligned} E &= E_0 + \lambda E_1 + \lambda^2 E_2 + \cdots, \\ |\psi\rangle &= |\psi_0\rangle + \lambda|\psi_1\rangle + \lambda^2/\psi_2\rangle + \cdots, \end{aligned} \tag{1.10}$$

the complexity of which remains, roughly speaking, on the MRS level.

In Sec. III, an illustration of applicability of the overall GRS scheme may be found: We consider arbitrary $V(r)$ in Eq. (1.3), and an arbitrary Padé approximant[7] of this function is incorporated into the band-matrix component T of the Hamiltonian. A $(2t+1)$-diagonal equation of the type (1.6) is obtained as a simple special case of Eq. (1.7).

In Sec. IV, a matrix continued fractional (MCF) construction of the propagator R is described, with a particular emphasis laid upon the MCF convergence for $V(r) = V_{\text{Padé}}(r)$. Nevertheless, only Sec. V makes our perturbative formalism complete: It resorts to the use of $\hat{f}_n$'s (the "capped" input MCF-like quantities) and simplifies the whole GRS construction considerably. Its final form may be characterized by the capped rearrangement of Eq. (1.9),

$$\begin{aligned} H &= H_0 + \lambda H_1, \\ H_0 &= T + U + \hat{\Theta} = \hat{T} + U, \\ \lambda H_1 &= \lambda W - U - \hat{\Theta} = H - H_0, \\ U &= O(\lambda), \quad \hat{\Theta} = \hat{T} - T = O(\lambda), \end{aligned} \tag{1.11}$$

 0022-2488/88/122611-07$02.50

and makes the traditional RS unperturbed solutions redundant.

In the Padé-oscillator examples, the role of the $\hat{f}_n$ "capped input" is assigned to the so-called fixed point asymptotic MCF estimates. These quantities define the propagator $R = \hat{R}$ as well as the capped operators $\hat{T}$ and $\hat{\Theta}$ in Eq. (1.11). In a way, they are analogs to the ordinary RS unperturbed spectrum. An example of their self-consistent variation is studied numerically and shown to be able to provide the upper and lower bonds to the energies.

II. THE GRS PERTURBATION THEORY

An insertion of the RS-type ansatz (1.9) and (1.10) in the perturbed chain model (1.7) leads immediately to the RS hierarchy of requirements arranged as the separate $O(\lambda^k)$ identities. The first one is a $(2t+1)$-diagonal chain model[8]

$$\sum_{n=m-t}^{m+t} (\langle m|H_0|n\rangle - E_0\langle m|S_0|n\rangle)\langle n|\psi_0\rangle = 0, \quad m = 0,1,\dots, \quad (2.1)$$

while the rest of them reads

$$(H_0 - E_0S_0)|\psi_k\rangle + H_1|\psi_{k-1}\rangle = \sum_{m=1}^{k} (E_mS_0 + E_{m-1}S_1)|\psi_{k-m}\rangle, \quad k = 1,2,\dots . \quad (2.2)$$

A. $k=0$

In contrast to the standard RS theory that requires the operator $H_0 - E\,S_0$ diagonal, the $t\geqslant 1$ Eq. (2.1) is not solvable in general. Fortunately, we may circumvent this difficulty in full analogy with the $S=I$ MRS case,[2] i.e., via an introduction of an auxiliary separable field U,

$$H_0 = T + U, \quad \lambda H_1 = \lambda W - U, \quad U = |0\rangle g\langle 0|. \quad (2.3)$$

With the sufficiently small coupling strength,

$$g = O(\lambda), \quad (2.4)$$

this split of the Hamiltonian remains still compatible with (1.10).

In the same MRS spirit, we may introduce also the projectors $P = |0\rangle\langle 0|$ and $Q = 1 - P$ and define the auxiliary unperturbed propagator as an inverse matrix,

$$R = Q(P - QD_0Q)^{-1}Q, \quad D_0 = T - E_0S_0. \quad (2.5)$$

Our (assumed) knowledge of this operator enables us to eliminate the "irrelevant" components of the wave functions

$$Q|\psi_0\rangle = RD_0|0\rangle\langle 0|\psi_0\rangle, \quad \langle 0|\psi_0\rangle \neq 0 \quad (2.6)$$

from (2.1). This reduces our unperturbed $k=0$ Schrödinger equation to mere scalar relation

$$g = -\langle 0|D_0|0\rangle - \langle 0|D_0RD_0|0\rangle. \quad (2.7)$$

Now a key point of the construction is a reinterpretation of this chain-model eigenvalue condition as a definition of $g = g(E_0)$. Then Eq. (2.1) becomes an identity, the zeroth-order quantities $\langle 0|\psi_0\rangle$ and E_0 remain free parameters, and only the condition of smallness (2.4) must be taken into account as a restriction of variability of the latter parameter.

B. $k\geqslant 1$

We may pick up the most natural normalization

$$\langle 0|\psi_0\rangle = 1, \quad \langle 0|\psi_k\rangle = 0, \quad k = 1,2,\dots, \quad (2.8)$$

and rewrite the P projection of Eq. (2.2) as a definition of energies

$$\begin{aligned} E_k\langle 0|S_0|\psi_0\rangle = {} & \langle 0|D_0|\psi_k\rangle + \langle 0|W|\psi_{k-1}\rangle \\ & - \delta_{k,1}\cdot h_0 - E_0\langle 0|S_1|\psi_{k-1}\rangle \\ & - \sum_{m=1}^{k-1} E_m(\langle 0|S_0|\psi_{k-m}\rangle \\ & + \langle 0|S_1|\psi_{k-m-1}\rangle), \quad k = 1,2,\dots . \end{aligned} \quad (2.9)$$

Here, the symbol δ_{ij} is a Kronecker delta ($\sim\langle 0|\psi_{k-1}\rangle$) and $g/\lambda = h_0 = O(1)$ [cf. Eq. (2.4)]. Similarly, the Q-projected remainder of Eq. (2.2) defines the wave functions

$$|\psi_k\rangle = RW|\psi_{k-1}\rangle - \sum_{m=1}^{k} E_mRS_0|\psi_{k-m} - \sum_{m=1}^{k} E_{m-1}RS_1|\psi_{k-m}\rangle, \quad k = 1,2,\dots . \quad (2.10)$$

C. The iterative interpretation of formulas

In principle, the coupled pair of the $t\geqslant 1$ equations (2.9) and (2.10) defines the explicit RS-like multiple series representation of corrections E_k and $|\psi_k\rangle$ in (1.10). For example, we may eliminate, say, $|\psi_k\rangle$ from (2.9) and get

$$\begin{aligned} E_k = {} & \kappa(\langle\varphi_0|W|\psi_{k-1}\rangle - h_0\delta_{k1} \\ & - \sum_{m=1}^{k-1} E_m\langle\varphi_0|S_0|\psi_{k-m}\rangle \\ & - \sum_{m=0}^{k-1} E_m\langle\varphi_0|S_1|\psi_{k-m-1}\rangle), \quad k = 1,2,\dots, \\ & \kappa = 1/\langle\varphi_0|S_0|\psi_0\rangle, \quad \langle\varphi_0| = \langle 0| + \langle 0|D_0R, \end{aligned} \quad (2.11)$$

etc.

In practice, we may recommend a direct numerical use of Eqs. (2.10) and (2.11) as recurrences. Thus our knowledge of $|\psi_0\rangle$ [(2.6) and (2.8)] and $\langle\varphi_0|$ [(2.11)] enables us to define the first nontrivial contribution

$$E_1 = \kappa(\langle\varphi_0|W|\psi_0\rangle - h_0 - E_0\langle\varphi_0|S_1|\psi_0\rangle) \quad (2.12)$$

to the energies as well as the subsequent wave function correction

$$|\psi_1\rangle = R(W|\psi_0\rangle - E_1S_0|\psi_0\rangle - E_0S_1|\psi_0\rangle). \quad (2.13)$$

For $k\geqslant 2$, we may proceed further in an iterative manner and, in a preparatory step, abbreviate

$$\begin{aligned} & |\tilde{r}_1\rangle = QS_0|0\rangle, \quad |\tilde{r}_2\rangle = QS_1|0\rangle, \\ & \langle\tilde{r}_3| = \langle\varphi_0|S_0Q, \quad \langle\tilde{r}_4| = \langle\varphi_0|S_1Q, \\ & \langle\tilde{r}_5| = \langle 0|WQ, \quad |\tilde{r}_a\rangle = E_0|\tilde{\psi}_1\rangle + E_1|\tilde{\psi}_0\rangle, \quad k = 2, \end{aligned} \quad (2.14)$$

where a restriction to the Q-projected subspace is marked by a tilde: this denotation is reasonable due to our normalization (2.8) (with $|\psi_k\rangle = |\tilde{\psi}_k>$, $k\geqslant 1$, etc.)

In the forthcoming steps, our formulas (2.10) and (2.11) may be understood as an algorithm:

(A1) define

$$|\tilde r_b\rangle = \sum_{m=1}^{k-1} E_m |\tilde\psi_{k-m}\rangle; \tag{2.15}$$

(A2) compute

$$E_k = \kappa(\langle \tilde r_5|\tilde\psi_{k-1}\rangle + \langle\tilde\varphi_0|\widetilde W|\tilde\psi_{k-1}\rangle - \langle\tilde r_3|\tilde r_b\rangle - \langle\tilde r_4|\tilde r_a\rangle; \tag{2.16}$$

(A3) redefine

$$|\tilde r_c\rangle = |\tilde r_b\rangle + E_k|\tilde\psi_0\rangle; \tag{2.17}$$

(A4) compute

$$|\tilde\psi_k\rangle = \widetilde R(\widetilde W|\tilde\psi_{k-1}\rangle - \widetilde S_0|\tilde r_c\rangle - \widetilde S_1|\tilde r_a\rangle - E_k|\tilde r_1\rangle - E_{k-1}|\tilde r_2\rangle; \tag{2.18}$$

(A5) redefine

$$|\tilde r_a\rangle := |\tilde r_c\rangle + E_0|\tilde\psi_k\rangle, \quad k := k+1; \tag{2.19}$$

and return to (A1) if the higher-order corrections are required.

In fact, our present GRS prescription is just an extension of the MRS formalism,[2] where $S = I$, i.e.,

$$|\tilde r_1\rangle = |\tilde r_2\rangle = |\tilde r_a\rangle = 0, \quad \langle\tilde r_3| = \langle\tilde\varphi_0|. \tag{2.20}$$

Similarly, we may even return to the standard RS $t=0$ formalism with

$$\langle\tilde r_3| = \langle\tilde\varphi_0| = 0, \quad |\tilde\psi_0\rangle = 0, \quad |\tilde r_c\rangle = |\tilde r_b\rangle. \tag{2.21}$$

III. APPLICATION: THE PADÉ APPROXIMATION OF POTENTIALS

A phenomenological use of the various central potentials $V(r)$ in Eq. (1.3) ranges from nuclear physics up to molecular physics and quantum chemistry. The functions $V(r)$ may differ in their $r\to\infty$ asymptotics as well as in their (possibly even weakly singular) behavior at $r\to 0$. In general we may assume that an ansatz

$$V_{\text{Padé}}(r) = (\text{a polynomial in } z)/(\text{polynomial in } z),$$
$$z = z(r) = r^{\text{integer/integer}}, \tag{3.1}$$
$$\lim_{r\to 0} r^2 V_{\text{Padé}}(r) > -\tfrac14,$$

will represent a sufficiently broad class of the necessary approximants, making an arbitrary "realistic" input force expressible in the form

$$V(r) = V_{\text{Padé}}(r) + \lambda V_c(r), \tag{3.2}$$

where $\lambda V_c(r)$ is arbitrarily small.[7]

In Eq. (1.3) with $V = V_{\text{Padé}}$, a change of variables $r\to r^{\text{integer/integer}}$ may be used for its conversion into the same equation with the $z(r) = r^2$ "canonical" form of the Padé force with $\lambda V_c(r) = 0$,

$$V(r) = G(r^2) + A(r^2)/B(r^2),$$
$$A(x) = \sum_{m=0}^{p-1} a_m x^m, \quad G(x) = \sum_{m=1}^{q} g_m x^m, \quad g_q > 0,$$
$$B(x) = \sum_{m=0}^{p} b_m x^m > 0, \quad b_p > 0, \quad p \geqslant 0, \quad q \geqslant 1, \tag{3.3}$$

and with an inessential reinterpretation of the parameters: the angular momentum $l > -\frac12$ becomes shifted and ceases to be an integer, and the original energy becomes transformed into one of the new couplings (say, g_{m_0}). Vice versa, the new energy variable E will represent just one of the original couplings. As an illustration, we may recall the screened Coulomb (SC) force

$$V_{\text{SC}}(r) = \alpha/r + \beta/(r+\gamma^2) \tag{3.4}$$

with the canonical representation (1.4).[9]

Among the merits of the Padé forces (3.3) we may mention not only their flexibility and universality, but also their elementary solvability for some particular couplings,[10] useful for performing easily the numerical tests.[11] Nevertheless, their main merit lies in a possibility to use the harmonic oscillator basis $|n\rangle$ and analogy with Eq. (1.6). Thus our GRS approach to the $\lambda V_c = 0$ force (3.2) may be based on an arbitrary product decomposition of the denominator in (3.3),

$$B(x) = B_L(x)\cdot B_R(x),$$
$$B_i(x) = \sum_{m=0}^{p_i} b_m^{(i)} x^m, \quad p_i \geqslant 0, \quad i = L,R, \quad p_L + p_R = p, \tag{3.5}$$

followed by a multiplicative redefinition of our $\lambda V_c = W = 0$ equation of the type (1.5),

$$\left\{B_L(r^2)\left[-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + G(r^2) - E\right]B_R(r^2) + A(r^2)\right\}\varphi(r) = 0, \tag{3.6}$$
$$\varphi(r) = \psi(r)/B_R(r^2),$$

and, finally, on an expansion of $\varphi(r)$ in the basis $|n\rangle$. As a result, we obtain our $W=0$ chain-model generalization (2.1) of Eq. (1.6),

$$\sum_{n=m-t}^{m+t} \langle m|H|n\rangle\langle n|\varphi\rangle = \sum_{n=m-s}^{m+s} E\langle m|S|n\rangle\langle n|\varphi\rangle, \quad m = 0,1,\dots, \tag{3.7}$$

where the operators

$$H = B_L(X)[H^{(\text{oscillator})} - X + G(X)]B_R(X) + A(X) = T,$$
$$S = B_L(X)B_R(X) = B(X) = S_0, \tag{3.8}$$

are band matrices since $X = r^2$ is tridiagonal. Obviously, we have $s = p$ and $t = p+q$ in general. For $q = 1$, we may even put $t = p$ after a rescaling of the coordinate such that $g_1 = 1$. Similarly, we may also incorporate $\lambda V_C \neq 0$.

In the conclusion, we have to reemphasize that our GRS series (1.10) need not represent the energy: For the $z(r)\neq r^2$ forces (3.1), $-E$ denotes, in fact, just a strength of the coupling g_0 that creates a bound state at a prescribed energy $-g_{m_0}$, $m_0 > 0$. Hence we are forced to denote $G_0(r) = G(r) - g_{m_0} r^{2m_0} + g_0$, $E = -g_{m_0}$ and change the corresponding matrices in Eq. (3.7),

$$H = B_L(X)[H^{(\text{oscillator})} - X + G_0(X)]B_R(X) + A(X) = T, \quad (3.9)$$
$$S = B_L(X)X^{m_0}B_L(X) = X^{m_0}B(X) = S_0.$$

Of course, we have to redefine also the integer $s = p + m_0$ and add the corrections $\lambda V_c(r) \neq 0$ and $\lambda S_1 \neq 0$ if needed.

IV. THE CONSTRUCTION OF PROPAGATORS R IN TERMS OF THE MATRIX CONTINUED FRACTIONS

A. The formalism

In accord with Sec. II, a knowledge of R (2.5) not only leads to the zeroth-order solution of the unperturbed problem, but it also renders possible an explicit specification of all the higher-order corrections. Thus a practical efficiency of the GRS expansions will mainly be determined by the inversion in (2.5).

Keeping this in mind, let us partition the unperturbed Schrödinger operator into its $(t \times t)$-dimensional submatrices,

$$D_0 = T - E_0S_0 = \begin{pmatrix} a_0 & b_0 & & \\ c_1 & a_1 & b_1 & \\ & c_2 & a_2 & b_2 \\ & & \cdots & \end{pmatrix}, \quad (4.1)$$

and also introduce the partitioned auxiliary matrices,

$$F_U = \begin{pmatrix} I & b_0f_1 & & \\ & I & b_1f_2 & \\ & & I & b_2f_3 \\ & & & \cdots \end{pmatrix},$$
$$F_L = \begin{pmatrix} I & & \\ f_1c_1 & I & \\ & f_2c_2 & I \\ & \cdots & \end{pmatrix}, \quad (4.2)$$
$$F_D = \begin{pmatrix} 1/f_0 & \\ & 1/f_1 \\ & \cdots \end{pmatrix}.$$

Then we may postulate an explicit factorization formula

$$D_0 = T - E_0S_0 = F_UF_DF_L. \quad (4.3)$$

This becomes an algebraic identity if and only if

$$1/f_k = a_k - b_kf_{k+1}c_{k+1}, \quad k = 0,1,\ldots . \quad (4.4)$$

Thus, whenever D_0 may be treated as an infinite-dimensional limit of its truncated forms, the latter relations may be initialized by $f_{M+1} = 0$, $M \to \infty$ and define just a matrix form of the analytic continued fractions.[12]

A motivation of our matrix continued fractional (MCF) factorization (4.3) of D_0 lies in the related simple form of the inverse matrices (e.g.,

$$F_U^{-1} = \begin{pmatrix} I & -b_0f_1 & b_0f_1b_1f_2 & -b_0f_1b_1f_2b_2f_3\cdots \\ & I & -b_1f_2 & -b_1f_2b_2f_3\cdots \\ & & I & -b_2f_3\cdots \\ & & & \cdots \end{pmatrix}, \quad (4.5)$$

etc.) and also in the easy projection of the triangular matrices (e.g., $F_LQ = QF_LQ$, etc.). Both these consequences of our ansatz (4.2) lead quickly to the explicit MCF definition

$$R = F_L^{-1}Q\begin{pmatrix} f_0 & & & \\ & f_1 & & \\ & & f_2 & \\ & & & \cdots \end{pmatrix}QF_U^{-1} \quad (4.6)$$

of our "input" unperturbed propagator R.

B. The MCF convergence

For the systems of the simple type (3.9), an important merit lies in a possibility of their semi- or non-numerical analysis. In particular, a study of convergence of the present infinite MCF expansions

$$f_k = \cfrac{1}{a_k - b_k\cfrac{1}{a_{k+1} - \cdots}c_{k+1}} \quad (4.7)$$

becomes feasible[13] due to a validity of the same leading-order asymptotic behavior of the matrices H and S,

$$\langle M+m|D_0|M+n\rangle = \text{const}\, M^{p+q}\binom{2t}{t+m-n} + O(M^{p+q-1}), \quad M \gg 1. \quad (4.8)$$

This may quite easily be derived from the well known formula[2,5]

$$\langle m|r^2|n\rangle \begin{cases} 2m + l + \frac{3}{2}, & n = m, \\ [(m+1)(m+l+\frac{3}{2})]^{1/2}, & n = m+1, \\ [m(m+l+\frac{1}{2})]^{1/2}, & n = m-1, \end{cases} \quad (4.9)$$

and, up to the irrelevant constant term, leads to the leading-order MCF estimate

$$f_{M+k} \sim f_M^{(\text{FP}0)} \sim M^{-p-q}\cfrac{1}{a^{(0)} - b^{(0)}\cfrac{1}{a^{(0)}\cdots}c^{(0)}},$$
$$a^{(0)} = SS^T + S^TS, \quad b^{(0)} = c^{(0)T} = S^TS^T, \quad (4.10)$$
$$S_{mn} = \binom{t}{n-m}, \quad m,n = 1,2,\ldots,t,$$

which may be summed up in the closed form

$$f_M^{(\text{FP}0)} \sim M^{-p-q}(SS^T)^{-1}. \quad (4.11)$$

In a higher-order analysis, we may recall the specialized literature[14] and create the more general asymptotics [the so-called fixed-point MCF expansions] of the type

$$f_{M+k} = f_M^{(\text{FP}0)} + f_M^{(\text{FP}1)} + \cdots + f_M^{(\text{FP}k)} + f'_{M+k}. \quad (4.12)$$

Here the remainder term is a MCF-type quantity again: for a sufficient number k_{FP} of subtractions of the type (4.12), the corresponding remainder recurrences may be linearized for $M \gg 1$ and the rigorous proof of convergence becomes trivial.[13–15]

In the present context, let us illustrate the higher-order MCF analysis on a particular $q > 2$ and $\gamma > 0$ extension of the asymptotic estimate (4.8),

$$\text{const}\langle M+m|D_0|M+n\rangle = M^{p+q}\binom{2t}{t+m-n} + \gamma M^{p+q-1}\binom{2t-2}{t+m-n-1} + O(M^{p+q-2}),$$

$$\gamma = b_{p-1}/b_p + g_{q-1}/g_q,\quad M \gg 1. \tag{4.13}$$

Then, in accord with the Appendix of Ref. 13, we may postulate

$$f_{M_0-k} = \frac{\text{const}\, M_0^{-p-q}}{SS^T + S^T|\beta\rangle\rho_k\langle\beta|S + \text{small terms}},$$
$$\langle\beta| = (1,-1,1,-1,...,(-1)^{t+1}),$$
$$M_0 \gg 1,\quad k = 0,1,.... \tag{4.14}$$

Such an approximation dominates the second-order behavior and reduces the $(t\times t)$-dimensional MCF recurrences to mere one-dimensional mapping $\rho_k \to \rho_{k+1}$, $k \ll M_0$,

$$\rho_{k+1} = (\rho_k + \gamma M_0^{-1} + O(M_0^{-2}))/(1+t\rho_k),$$
$$k = 0,1,.... \tag{4.15}$$

An analysis of convergence of the latter reduced MCF mapping is not difficult. In accord with Fig. 1, we may conclude that it has a pair of the real fixed points,

$$\rho_k \approx \rho_{k+1} \approx \pm\rho^{(\text{FP})} = \pm[\gamma/M_0 t + O(M_0^{-2})]^{1/2},\quad k \gg 1, \tag{4.16}$$

the positive one of which is a point of accumulation of the sequence

$$\rho_0 \to \cdots \to \rho_k \to \rho_{k+1} \to \cdots,\quad k \ll M_0, \tag{4.17}$$

for an almost arbitrary initial choice ρ_0. In particular, a convergence from above and from below may be expected for $|\rho_0| > \rho^{(\text{FP})}$ and $|\rho_0| < \rho_1^{(\text{FP})}$, respectively.

V. THE NON-NUMERICAL CONSTRUCTION OF $R = \hat{R}$

A. The second auxiliary rearrangement operator $\hat{\Theta}$ and the final form of the GRS formalism

For the sufficiently large number of terms in the fixed point formula (4.12), we may define

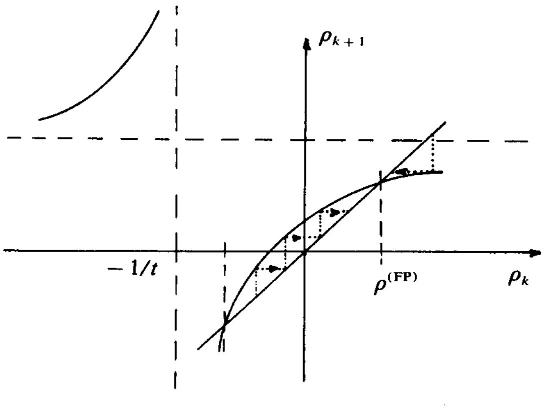


FIG. 1. The auxiliary $(t\times t)$-dimensional continued fraction convergence in the leading-order asymptotic one-dimensional approximation (4.15) $[\rho^{(\text{FP})} = (\gamma/Mt)^{1/2}]$.

$$\hat{f}_M = f_M^{(\text{FP}0)} + f_M^{(\text{FP}1)} + \cdots + f_M^{(\text{FP}k)},\quad M \geq M_0, \tag{5.1}$$

as a sufficiently reliable approximant to the uncapped MCF quantities. Vice versa, the capped quantities $\hat{f}_M$ have to correspond to some unknown operator $\hat{D}_0$ via the factorization (4.3). Now, an assumption $\hat{b}_m = b_m$, $\hat{c}_{m+1} = c_{m+1}$ may be combined with the necessary and sufficient condition (4.4) into an explicit definition

$$\hat{a}_m = \hat{f}_m^{-1} + b_m\hat{f}_{m+1}c_{m+1},\quad m \geq M_0, \tag{5.2}$$

of the missing matrix elements of $\hat{D}_0$.

For the sufficiently large M_0 or k_{FP} [= number of terms in (5.1)], the difference between the capped and uncapped operators will be small,

$$\hat{\Theta} = \hat{D}_0 - D_0 = O(\lambda), \tag{5.3}$$

and may be treated as another component of the perturbation. In more detail, we may replace Eq. (1.9) by Eq. (1.11), i.e., (a) start from the FP formula (5.1) and define the capped operator $\hat{D}_0$ by means of Eq. (5.2); (b) define the capped, solvable band matrix $\hat{T} = \hat{T}(\hat{f}_i, i = 0,1,...)$,

$$\hat{T} = T + \hat{D}_0 - D_0, \tag{5.4}$$

and preserve the overlap matrix S unchanged; (c) use Eq. (4.6) and replace R [the inversion (2.5)] by the capped operator $\hat{R} = \hat{R}(\hat{f}_i, i = 0,1,...)$; (d) replace all D_0 by $\hat{D}_0$ and also substitute

$$\lambda\hat{W} = H - \hat{T} = \lambda W - \hat{\Theta},\quad \hat{\Theta} = O(\lambda), \tag{5.5}$$

for all the uncapped λW's; and (e) use the GRS formalism of Sec. II with the capped quantities: $T \to \hat{T}$, $R \to \hat{R}$, $D_0 \to \hat{D}_0$, and $W \to \hat{W}$.

B. The Padé oscillator illustration

Let us put $V_c(r) = 0$ in Eq. (3.2) and define the related uncapped unperturbed operator R in terms of the MCF quantities f_n. Then, our perturbation $\lambda H_1 = -U \neq 0$ is separable and reflects merely our "bad" choice of E_0. Of course, we are tempted to define now the "best" value of $E_0 = E_{\text{exact}}$ by the requirement $g(E_0) = 0$, i.e., $\lambda H_1 = 0$, i.e.,

$$\det 1/f_0 = 0 \tag{5.6}$$

[cf. Eq. (2.7)]. This is a Brillouin–Wigner-type formula and enables us to determine E_{exact} numerically by the trial and error technique (cf., e.g., Graffi and Grecchi,[16] etc.).

In light of Sec. IV B, acceleration of the MCF convergence in Eq. (5.6) may be achieved by the FP subtractions. Alternatively, we may ignore the higher-order MCF corrections and use Eq. (5.1) in an approximative evaluation of the energies. In the present GRS context, we may combine both these ideas: We may start from the uncapped $W = 0$ and matrix D_0 and redefine the corrections ($\lambda\hat{W} \neq 0$),

$$\text{const}\langle M_0+m|\lambda\hat{W}|M_0+n\rangle = \gamma M_0^{p+q-1}\binom{2t-2}{t+m-n-1} + O(M_0^{p+q-2}),$$
$$m,n \geq 0,$$

via the explicit algebraic formula (5.1). No infinite MCF expansions of the type (4.7) are needed anymore.

C. Perturbative determination of the upper and lower estimates of energies

An introduction of the auxiliary operator $\hat{\Theta} = O(\lambda)$ and of the secondary rearrangement (1.11) of the full Hamiltonian H has been motivated formally: It enables us to reconstruct the complete GRS input (namely, the operator $R = \hat{R}$, approximant $\hat{T}$, and the split matrices $H_0 = \hat{T} + U$ and $\lambda H_1 = \lambda \hat{W} - U$) in an entirely non-numerical manner.

Now, let use pay attention to the corresponding GRS perturbation formulas at a given and fixed order of precision $O(\lambda^k)$. For the sake of simplicity, we shall analyze just the first-order example of the preceding subsection and notice that the leading-order components (5.7) of the perturbation $\lambda \hat{W}$ are positive semidefinite. Indeed, they may be interpreted as certain matrix elements of the powers of $r^2 > 0$.

The most important consequence of the above observation lies in an easy majorization and minorization of the whole perturbation $\lambda \hat{W} > 0$ by the operators

$$\lambda W^{\binom{\rm maj}{\rm min}} = \lambda z^{\binom{\rm maj}{\rm min}} \hat{W}, \quad z^{(\rm maj)} > 1 + O(1/M_0), \qquad z^{(\rm min)} < 1 - O(1/M_0). \tag{5.8}$$

The corresponding replacement of the parameter $\gamma > 0$ in Eq. (5.7) by $\gamma^{(\rm maj)} = z^{(\rm maj)}\gamma > 0$ and $\gamma^{(\rm min)} = z^{(\rm min)}\gamma > 0$, respectively, leads also to an explicit change of ρ_0 in (4.14),

$$\begin{aligned} &\rho_0^{\binom{\rm maj}{\rm min}} \in (\rho_a^{\binom{\rm maj}{\rm min}}, \rho_b^{\binom{\rm maj}{\rm min}}),\\ &\rho_a^{\binom{\rm maj}{\rm min}} = (z^{\binom{\rm maj}{\rm min}}\gamma/M_0 t - |{\rm const}|/M_0^2)^{1/2},\\ &\rho_b^{\binom{\rm maj}{\rm min}} = (z^{\binom{\rm maj}{\rm min}}\gamma/M_0 t + |{\rm const}|/M_0^2)^{1/2}. \end{aligned} \tag{5.9}$$

Here, the length of the uncertainty interval need not be negligible since the parameters themselves are small.

For the purely numerical purposes, let us now treat ρ_0 as a free parameter, and eliminate also the GRS convergence questions by using the Brillouin–Wigner-type formula (5.6). We may expect that the resulting energy roots $E = E(M_0, \rho_0)$ will also follow the majorization and minorization change of the Hamiltonians,

$$E(M_0, \rho_0^{(\rm min)}) < E_{\rm exact} < E(M_0, \rho_0^{(\rm maj)}). \tag{5.10}$$

As a consequence, the best results will lie somewhere within the interval of parameters

$$\rho_0 \in (\rho_b^{(\rm min)}, \ \rho_a^{(\rm maj)}). \tag{5.11}$$

This is our first important observation.[17]

In the computations, a precise localization of the optimal interval (5.11) may prove difficult both algebraically (the higher-order FP corrections are complicated[14]) and computationally (whenever we get $\rho_b^{(\rm min)} \approx \rho_a^{(\rm maj)}$ for some small M_0). Nevertheless, there is an easy way out of this difficulty: The dependence of energies $E(M_0, \rho_0)$ weakens with the increasing M_0 so that the two curves $E(M_0^{(\alpha)}, \rho_0)$ and $E(M_0^{(\beta)}, \rho_0)$ must intersect somewhere in the interval (5.11). The point of intersection specifies the best values of ρ_0 and E_0 geometrically.

Numerically, the geometric "sandwiching" of energies of this type has been tested on the various $t \leqslant 2$ examples with $M_0^{(\alpha)} = 400$ and $M_0^{(\beta)} = 500$. Schematically, the results are summarized in Fig. 2 where a common pattern of the M_0 and ρ_0 dependence of $E(M_0, \rho_0)$ is depicted in arbitrary units. The scheme worked even beyond its natural limits given by its present derivation: A large initial choice of $\rho_0 \to \pm \infty$ simulates the MCF initialization $f_{m_0} = 0$ and exhibits also the expected variational behavior typical for the standard increasing truncation of Hamiltonians. For the smaller M_0's, a use of the third curve $E(M_0^{(\gamma)}, \rho_0)$ may sometimes be recommended for an estimate of the error bars.[18]

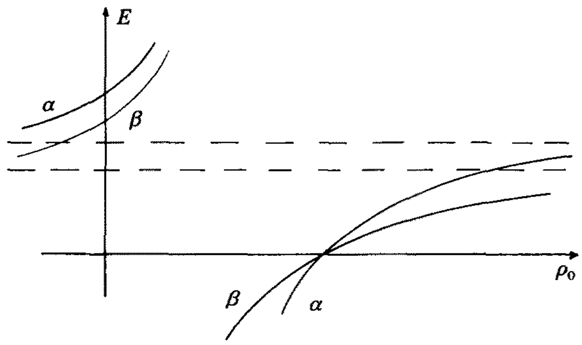


FIG. 2. A geometric, numerical sandwiching of the energies.

VI. SUMMARY

In the paper, a generalized Rayleigh–Schrödinger perturbation theory has been proposed and described as a means of solving the Schrödinger bound-state problem when formulated as a matrix equation in some nonorthogonal basis,

$$H\psi = ES\psi, \quad S \neq I. \tag{6.1}$$

As an illustrative example, we have chosen the determination of bound states in a general central potential approximated by a ratio of two polynomials of an arbitrary degree. The specific features of such a Padé oscillator example proved extremely useful.

(1) In the first stage, a band-matrix $(2t + 1)$-diagonal structure of the related Schrödinger operator enabled us to satisfy the main and strongest GRS requirement (namely, an *a priori* knowledge of the unperturbed propagator R) in the manner proposed recently for the $S = I$ case [namely, via a construction of R in terms of certain auxiliary $(t \times t)$-dimensional continued fractions].

(2) An asymptotic smoothness of the matrix elements of D_0 proved essential for a subsequent elimination of the infinite (i.e., numerical) MCF expansions. In the GRS formulas, we have replaced them by certain algebraic (so-called fixed point, capped) finite expressions that generalized directly the standard RS input unperturbed spectrum.

(3) The $S \neq I$ formulation of the Padé-oscillator examples with the FP elimination of the MCF expansions has been shown to contain the positive semidefinite perturbations $\lambda \hat{W}$. The subsequent easy majorization and minorization (sandwiching) of Hamiltonians also leads quickly to the GRS formulas that majorize and minorize the energies of a $O(\lambda^k)$ level of precision in principle.

(4) A modified interpretation of the operator S renders possible a GRS generation of the perturbation formulas for the couplings

$$G = G_0 + \lambda G_1 + \lambda^2 G_2 + \cdots . \qquad (6.2)$$

Such a completion of the expansions of energies may also be complemented by the perturbative construction of the so-called "Sturmians" [5] if needed.

In a broader methodical sence, all our Padé-oscillator results exemplify the main merits of our GRS expansions: (1) a feasibility of the non-numerical $S \neq I$ expansions, (2) a generalization and weakening of the standard RS "solvability" requirement, (3) a possible sandwiching of the exact values in a variational and "antivariational" manner, and (4) an extension (6.2) of the standard "changing-energy" picture of vicinity of the fixed "unperturbed" system to a full analysis [cf. Eq. (3.9) with $m_0 = 0,1,...,q$] of mutual relations between the variable couplings and the energies.

[1]E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961).

[2]M. Znojil, Phys. Rev. A **35**, 2448 (1987); Phys. Lett. A **125**, 443 (1987); preprint JINR Dubna E5-87-634, 1987; M. Znojil, M. Flynn, and R. F. Bishop, proceedings "Hadron Structure 87," in *Physics and Applications*, edited by D. Krupa (Institute of Physics, Bratislava, 1988), Vol. 14, p. 252.

[3]See, e.g., the references listed by Y. P. Varshni, Phys. Rev. A **36**, 3009 (1987).

[4]N. Bessis and G. Bessis, J. Math. Phys. **21**, 2780 (1980).

[5]R. R. Whitehead, A. Watt, G. Flessas, and M. A. Nagarajan, J. Phys. A: Mat. Gen. **15**, 1217 (1982).

[6]H. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford U.P., Oxford, 1965).

[7]G. A. Baker, Jr., and P. Graves-Morris, *Padé Approximants* (Addison–Wesley, Reading, MA, 1981).

[8]M. Znojil, J. Math. Phys. **25**, 2979 (1984).

[9]M. Znojil, Phys. Lett. A **94**, 120 (1983).

[10]M. Znojil, J. Phys. A: Math. Gen. **16**, 279 (1983).

[11]M. H. Blecher and P. G. L. Leach, J. Phys. A: Math. Gen. **20**, 5923 (1987).

[12]H. S. Wall, *Analytic Theory of the Continued Fractions* (Van Nostrand, New York, 1948).

[13]M. Znojil, J. Phys. A: Math. Gen. **17**, 1611 (1984).

[14]See Ref. 8 and further papers listed in M. Znojil, Czech. J. Phys. B **37**, 1072 (1987).

[15]M. Znojil, Phys. Rev. D **24**, 903 (1981).

[16]S. Graffi and V. Grecchi, Lett. Nuovo Cimento **12**, 425 (1975).

[17]M. Znojil, preprint JINR Dubna E4-86-60, 1986.

[18]In fact, there exist many other majorizations and minorizations of energies in perturbation theory, cf., e.g., R. C. Young, L. C. Biedenharn, and E. Feenberg, Phys. Rev. **106**, 1151 (1957), as one of the typical early references also using continued fractions in a somewhat different setting.

The exterior metric approach to a charged axially symmetric celestial body—the fourth-order approximate solutions of Einstein–Maxwell equations

Zhou Qi-huang
Department of Physics, Hunan Normal University, Changsha, Hunan, People's Republic of China

(Received 1 June 1988; accepted for publication 10 August 1988)

Starting with the general expression of a static state axisymmetric metric and using the principle of equivalence and the Maccullagh formula, the Einstein–Maxwell equations of a charged axisymmetric celestial body are obtained. Next, using the method of undetermined coefficients these equations are solved up to fourth-order approximate. These sets of solutions are generally appropriate for all kinds of charged axisymmetric celestial bodies.

I. THE GENERAL EXPRESSION OF THE METRIC OF A STATIC STATE AXISYMMETRIC GRAVITATIONAL FIELD AS WELL AS ITS CONNECTION AND RICCI TENSORS

The general expression for the metric of a static axisymmetric gravitational field is[1]

$$d\tau^2 = -e^{2u}(dx^0)^2 + e^{2(\kappa-u)} \times[(dx^1)^2 + (dx^2)^2] + \rho^2 e^{-2u} d\varphi^2, \quad (1)$$

where u, κ, and ρ are the functions of x^1 and x^2. Therefore the nonzero components of affine connection are

$$\begin{aligned}
&\Gamma^0_{01} = \Gamma^0_{10} = \frac{\partial u}{\partial x^1}, \quad \Gamma^0_{02} = \Gamma^0_{20} = \frac{\partial u}{\partial x^2},\\
&\Gamma^1_{00} = e^{4u-2\kappa}\frac{\partial u}{\partial x^1}, \quad \Gamma^1_{11} = \frac{\partial \kappa}{\partial x^1} - \frac{\partial u}{\partial x^1},\\
&\Gamma^1_{21} = \Gamma^1_{12} = \frac{\partial \kappa}{\partial x^2} - \frac{\partial u}{\partial x^2}, \quad \Gamma^1_{22} = \frac{\partial u}{\partial x^1} - \frac{\partial \kappa}{\partial x^1},\\
&\Gamma^1_{33} = e^{-2\kappa}\left(\rho^2\frac{\partial u}{\partial x^1} - \rho\frac{\partial \rho}{\partial x^1}\right), \quad \Gamma^2_{00} = e^{4u-2\kappa}\frac{\partial u}{\partial x^2}, \qquad (2)\\
&\Gamma^2_{11} = \frac{\partial u}{\partial x^2} - \frac{\partial \kappa}{\partial x^2}, \quad \Gamma^2_{21} = \Gamma^2_{12} = \frac{\partial \kappa}{\partial x^1} - \frac{\partial u}{\partial x^1},\\
&\Gamma^2_{22} = \frac{\partial \kappa}{\partial x^2} - \frac{\partial u}{\partial x^2}, \quad \Gamma^2_{33} = e^{-2\kappa}\left(\rho^2\frac{\partial u}{\partial x^2} - \rho\frac{\partial \rho}{\partial x^2}\right),\\
&\Gamma^3_{31} = \Gamma^3_{13} = \rho^{-1}\frac{\partial \rho}{\partial x^1} - \frac{\partial u}{\partial x^1},\\
&\Gamma^3_{32} = \Gamma^3_{23} = \rho^{-1}\frac{\partial \rho}{\partial x^2} - \frac{\partial u}{\partial x^2}.
\end{aligned}$$

The nonzero components of the Ricci tensor are

$$\begin{aligned}
R_{00} =& -e^{4u-2\kappa}\left[\frac{\partial^2 u}{(\partial x^1)^2} + \frac{\partial^2 u}{(\partial x^2)^2} + \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1} + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right)\right],\\
R_{11} =& \frac{\partial^2 \kappa}{(\partial x^1)^2} + \frac{\partial^2 \kappa}{(\partial x^2)^2} - \frac{\partial^2 u}{(\partial x^1)^2} - \frac{\partial^2 u}{(\partial x^2)^2} + 2\left(\frac{\partial u}{\partial x^1}\right)^2 + \rho^{-1}\frac{\partial^2 \rho}{(\partial x^1)^2}\\
&- \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1} + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right) - \rho^{-1}\left(\frac{\partial \kappa}{\partial x^1}\frac{\partial \rho}{\partial x^1} - \frac{\partial \kappa}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right),\\
R_{12} =& 2\frac{\partial u}{\partial x^1}\frac{\partial u}{\partial x^2} - \rho^{-1}\left(\frac{\partial \kappa}{\partial x^1}\frac{\partial \rho}{\partial x^2} + \frac{\partial \kappa}{\partial x^2}\frac{\partial \rho}{\partial x^1}\right) + \rho^{-1}\frac{\partial^2 \rho}{\partial x^1 \partial x^2}, \qquad (3)\\
R_{22} =& \frac{\partial^2 \kappa}{(\partial x^1)^2} + \frac{\partial^2 \kappa}{(\partial x^2)^2} - \frac{\partial^2 u}{(\partial x^1)^2} - \frac{\partial^2 u}{(\partial x^2)^2} + 2\left(\frac{\partial u}{\partial x^2}\right)^2 + \rho^{-1}\frac{\partial^2 \rho}{(\partial x^2)^2}\\
&- \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1} + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right) + \rho^{-1}\left(\frac{\partial \kappa}{\partial x^1}\frac{\partial \rho}{\partial x^1} - \frac{\partial \kappa}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right),\\
R_{33} =& -\rho^2 e^{-2\kappa}\left[\frac{\partial^2 u}{(\partial x^1)^2} + \frac{\partial^2 u}{(\partial x^2)^2} + \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1} + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2} - \frac{\partial^2 \rho}{(\partial x^1)^2} - \frac{\partial^2 \rho}{(\partial x^2)^2}\right)\right].
\end{aligned}$$

II. EXPRESSIONS FOR THE ENERGY-MOMENTUM TENSOR $T_{\mu\nu}$ OF A STATIC ELECTRIC FIELD

According to

$$T_{\mu\nu} = F_{\mu\sigma}F_\nu{}^\sigma - \tfrac{1}{4}g_{\mu\nu}F_{\sigma\tau}F^{\sigma\tau}, \quad (4)$$

where $F_{\mu\nu}$ is the antisymmetry tensor of the electromagnetic field, i.e., $F_{\mu\nu} = \partial A_\nu/\partial x^\mu - \partial A_\mu/\partial x^\nu$, and A_μ is the four-dimensional potential of the electromagnetic field.

The axisymmetrical component of the axisymmetrical four-dimensional potential should be zero, i.e., $A_3 = A_\varphi = 0$. Because the electromagnetic potential has gauge freedom, i.e., $A'_\mu = A_\mu + \partial\Lambda/\partial x^\mu$, where $\Lambda = \Lambda(x^1,x^2)$, we can choose $\Lambda(x^1,x^2)$, such that

$$A'_1 = A_1 + \frac{\partial \Lambda}{\partial x^1} = 0, \quad A'_2 = A_2 + \frac{\partial \Lambda}{\partial x^2} = 0.$$

Then the nonzero component of the electromagnetic potential only has A_0, and the nonzero components of the electromagnetic field tensor are

 0022-2488/88/122618-04$02.50

$$F_{01} = -F_{10} = -\frac{\partial A_0}{\partial x^1}, \quad F_{02} = -F_{20} = -\frac{\partial A_0}{\partial x^2},$$

$$F^{01} = -F^{10} = e^{-2\kappa}\frac{\partial A_0}{\partial x^1}, \quad F^{02} = -F^{20} = e^{-2\kappa}\frac{\partial A_0}{\partial x^2}. \tag{5}$$

Combining the above equations, we obtain the nonzero components of the electromagnetic field energy-momentum tensor $T_{\mu\gamma}$:

$$\begin{aligned}
T_{00} &= \frac{1}{2}e^{2u-2\kappa}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 + \left(\frac{\partial A_0}{\partial x^2}\right)^2\right],\\
T_{11} &= -\frac{1}{2}e^{-2u}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 - \left(\frac{\partial A_0}{\partial x^2}\right)^2\right],\\
T_{22} &= \frac{1}{2}e^{-2u}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 - \left(\frac{\partial A_0}{\partial x^2}\right)^2\right],\\
T_{33} &= \frac{1}{2}\rho^2 e^{-2u-2\kappa}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 + \left(\frac{\partial A_0}{\partial x^2}\right)^2\right],\\
T_{12} &= -e^{-2u}\frac{\partial A_0}{\partial x^1}\frac{\partial A_0}{\partial x^2},\\
T &= g^{\mu\nu}\left(F_{\mu\sigma}F_\nu{}^\sigma - \tfrac{1}{4}g_{\mu\nu}F_{\sigma\tau}F^{\sigma\tau}\right) = 0.
\end{aligned} \tag{6}$$

III. EINSTEIN-MAXWELL FIELD EQUATIONS

Substituting the above expressions for $R_{\mu\nu}$, $T_{\mu\nu}$, and T into the Einstein gravitational equation, we have

$$\frac{\partial^2 u}{(\partial x^1)^2} + \frac{\partial^2 u}{(\partial x^2)^2} + \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1} + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right) = \frac{\kappa_0}{2}e^{-2u}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 + \left(\frac{\partial A_0}{\partial x^2}\right)^2\right], \tag{7}$$

$$\begin{aligned}
&\frac{\partial^2 \kappa}{(\partial x^1)^2} + \frac{\partial^2 \kappa}{(\partial x^2)^2} - \frac{\partial^2 u}{(\partial x^1)^2} - \frac{\partial^2 u}{(\partial x^2)^2} + 2\left(\frac{\partial u}{\partial x^1}\right)^2\\
&\quad + \rho^{-1}\frac{\partial^2 \rho}{(\partial x^1)^2} - \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1} + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right)\\
&\quad - \rho^{-1}\left(\frac{\partial \kappa}{\partial x^1}\frac{\partial \rho}{\partial x^1} - \frac{\partial \kappa}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right)\\
&\quad = \frac{\kappa_0}{2}e^{-2u}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 - \left(\frac{\partial A_0}{\partial x^2}\right)^2\right],
\end{aligned} \tag{8}$$

$$\begin{aligned}
&2\frac{\partial u}{\partial x^1}\frac{\partial u}{\partial x^2} - \rho^{-1}\left(\frac{\partial \kappa}{\partial x^1}\frac{\partial \rho}{\partial x^2} + \frac{\partial \kappa}{\partial x^2}\frac{\partial \rho}{\partial x^1}\right)\\
&\quad + \rho^{-1}\frac{\partial^2 \rho}{\partial x^1 \partial x^2} = \kappa_0 e^{-2u}\frac{\partial A_0}{\partial x^1}\frac{\partial A_0}{\partial x^2},
\end{aligned} \tag{9}$$

$$\begin{aligned}
&\frac{\partial^2 \kappa}{(\partial x^1)^2} + \frac{\partial^2 \kappa}{(\partial x^2)^2} - \frac{\partial^2 u}{(\partial x^1)^2} - \frac{\partial^1 u}{(\partial x^2)^2} + 2\left(\frac{\partial u}{\partial x^2}\right)^2\\
&\quad + \rho^{-1}\frac{\partial^2 \rho}{(\partial x^2)^2} - \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1} + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right)\\
&\quad + \rho^{-1}\left(\frac{\partial \kappa}{\partial x^1}\frac{\partial \rho}{\partial x^1} - \frac{\partial \kappa}{\partial x^2}\frac{\partial \rho}{\partial x^2}\right)\\
&\quad = -\frac{\kappa_0}{2}e^{-2\kappa}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 - \left(\frac{\partial A_0}{\partial x^2}\right)^2\right],
\end{aligned} \tag{10}$$

$$\begin{aligned}
&\frac{\partial^2 u}{(\partial x^1)^2} + \frac{\partial^2 u}{(\partial x^2)^2} + \rho^{-1}\left(\frac{\partial u}{\partial x^1}\frac{\partial \rho}{\partial x^1}\right.\\
&\quad \left. + \frac{\partial u}{\partial x^2}\frac{\partial \rho}{\partial x^2} - \frac{\partial^2 \rho}{(\partial x^1)^2} - \frac{\partial^2 \rho}{(\partial x^2)^2}\right)\\
&\quad = \frac{\kappa_0}{2}e^{-2u}\left[\left(\frac{\partial A_0}{\partial x^1}\right)^2 + \left(\frac{\partial A_0}{\partial x^2}\right)^2\right].
\end{aligned} \tag{11}$$

In order to simplify the gravitational equations (7)–(11), we introduce the classical coordinates

$$x^1 = \rho = r, \quad x^2 = z$$

Obviously, the classical coordinates are reasonable because they satisfy

$$\frac{\partial^2 \rho}{(\partial x^1)^2} + \frac{\partial^2 \rho}{(\partial x^2)^2} = 0, \tag{12}$$

which is obtained by (7) − (11). Therefore (7)–(11) can be simplified into

$$\frac{\partial^2 u}{\partial r^2} + \frac{\partial^2 u}{\partial z^2} + \frac{1}{r}\frac{\partial u}{\partial r} = \frac{\kappa_0}{2}\left[\left(\frac{\partial A_0}{\partial r}\right)^2 + \left(\frac{\partial A_0}{\partial z}\right)^2\right]e^{-2u}, \tag{13}$$

$$\frac{1}{r}\frac{\partial \kappa}{\partial r} + \left(\frac{\partial u}{\partial z}\right)^2 - \left(\frac{\partial u}{\partial r}\right)^2 = -\frac{\kappa_0}{2}\left[\left(\frac{\partial A_0}{\partial r}\right)^2 - \left(\frac{\partial A_0}{\partial z}\right)^2\right]e^{-2u}, \tag{14}$$

$$\frac{1}{r}\frac{\partial \kappa}{\partial z} - 2\left(\frac{\partial u}{\partial r}\frac{\partial u}{\partial z}\right) = -\kappa_0\left(\frac{\partial A_0}{\partial x}\right)\left(\frac{\partial A_0}{\partial z}\right)e^{-2u}, \tag{15}$$

$$\frac{\partial^2 \kappa}{\partial r^2} + \frac{\partial^2 \kappa}{\partial z^2} + \left(\frac{\partial u}{\partial r}\right)^2 + \left(\frac{\partial u}{\partial z}\right)^2 = \frac{\kappa_0}{2}\left[\left(\frac{\partial A_0}{\partial r}\right)^2 + \left(\frac{\partial A_0}{\partial z}\right)^2\right]e^{-2u}, \tag{16}$$

where $\kappa_0 = 8\pi G$.

Substituting the expressions for g in (1) and $F^{\mu\nu}$ into the Maxwell equation

$$\frac{1}{\sqrt{-g}}\frac{\partial(\sqrt{-g}F^{\mu\nu})}{\partial x^\mu} = 0,$$

we have

$$\frac{\partial(-re^{-2u}\,\partial A_0/\partial r)}{\partial r} + \frac{\partial(-re^{-2u}\,\partial A_0/\partial z)}{\partial z} = 0. \tag{17}$$

It can be seen that when

$$\frac{\partial A_0}{\partial r} = e^{2u}\frac{\partial V}{\partial r}, \quad \frac{\partial A_0}{\partial z} = e^{2u}\frac{\partial V}{\partial z}, \tag{18}$$

the Maxwell equation becomes

$$\frac{\partial^2 V}{\partial r^2} + \frac{\partial^2 V}{\partial z^2} + \frac{1}{r}\frac{\partial V}{\partial r} \equiv \Delta V = 0.$$

This is just the Laplace equation of a static electric potential without a gravitational field. Therefore V should be the static electric potential in Minkowski space, and $u \to 0$, $\partial A_0/\partial r \to \partial V/\partial r$, $\partial A_0/\partial z \to \partial V/\partial z$, when $r \to \infty$; thus (18) satisfies the principle of equivalence and (17) has the solution of (18).

From the Maccullagh formula, the approximate expression of a static electric potential of an axisymmetric charged body is[2]

$$V = \frac{kQ}{(r^2+z^2)^{1/2}} + \frac{kQ(I_3 - I_1)}{2(r^2+z^2)^{3/2}}\left(1 - \frac{3z^2}{r^2+z^2}\right), \tag{19}$$

where Q is the total electric charge of a celestial body, k is the rate constant of the electric field, and I_3 and I_1 are the moments of inertia about the symmetry axis (as the third central principal axis) and the first or second central principal axis divided by the total mass of the charged body, respectively. Also, we have

$$\begin{aligned}\frac{\partial V}{\partial r}&=-\frac{kQr}{(r^2+z^2)^{3/2}}-\frac{3kQ(I_3-I_1)r}{(r^2+z^2)^{5/2}}\\&\quad+\frac{15kQ(I_3-I_1)rz^2}{2(r^2+z^2)^{7/2}},\\\frac{\partial V}{\partial z}&=-\frac{kQz}{(r^2+z^2)^{3/2}}-\frac{9kQ(I_3-I_1)z}{(r^2+z^2)^{5/2}}\\&\quad+\frac{15kQ(I_3-I_1)z^3}{2(r^2+z^2)^{7/2}}.\end{aligned}\tag{20}$$

Substituting (18) into (13)–(16), we have

$$\frac{\partial^2 u}{\partial r^2}+\frac{\partial^2 u}{\partial z^2}+\frac{1}{r}\frac{\partial u}{\partial r}=\frac{\kappa_0}{2}\left[\left(\frac{\partial V}{\partial r}\right)^2+\left(\frac{\partial V}{\partial z}\right)^2\right]e^{2u},\tag{21}$$

$$\frac{1}{r}\frac{\partial\kappa}{\partial r}+\left(\frac{\partial u}{\partial z}\right)^2-\left(\frac{\partial u}{\partial r}\right)^2=-\frac{\kappa_0}{2}\left[\left(\frac{\partial V}{\partial r}\right)^2-\left(\frac{\partial V}{\partial z}\right)^2\right]e^{2u},\tag{22}$$

$$\frac{1}{r}\frac{\partial\kappa}{\partial z}-2\frac{\partial u}{\partial r}\frac{\partial u}{\partial z}=-\kappa_0\left(\frac{\partial V}{\partial r}\right)\left(\frac{\partial V}{\partial z}\right)e^{2u},\tag{23}$$

$$\begin{aligned}&\frac{\partial^2\kappa}{\partial r^2}+\frac{\partial^2\kappa}{\partial z^2}+\left(\frac{\partial u}{\partial r}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2\\&\quad=\frac{\kappa_0}{2}\left[\left(\frac{\partial V}{\partial r}\right)^2+\left(\frac{\partial V}{\partial z}\right)^2\right]e^{2u};\end{aligned}\tag{24}$$

(21)–(23) are called the Einstein–Maxwell equations of the charged axisymmetric celestial body.

IV. THE SOLUTIONS OF EINSTEIN–MAXWELL EQUATIONS AND THE METRIC TENSOR (REF. 3)

From (20), it can be seen that the right-hand side is the power series of $1/(r^2+z^2)$, therefore we can assume that (20) contains the solution of this power series. Since $r\to\infty$ or $z\to\infty$, $g_{00}=e^{-2u}=-1-2\Phi$, Φ is Newton's gravitational potential, i.e.,

$$\Phi=-\frac{GM}{(r^2+z^2)^{1/2}}+\frac{GM(I_3-I_1)}{2(r^2+z^2)^{3/2}}\left(\frac{3z^2}{r^2+z^2}-1\right)+\cdots,\tag{25}$$

where G is the gravitational constant, M is the total mass of celestial body, and $\Delta\Phi=0$, we obtain

$$u=u_0+\Phi.\tag{26}$$

Inserting (26) into (21), we have

$$\frac{\partial^2 u_0}{\partial r^2}+\frac{1}{r}\frac{\partial u_0}{\partial r}+\frac{\partial^2 u_0}{\partial z^2}=\frac{\kappa_0}{2}\left[\left(\frac{\partial V}{\partial r}\right)^2+\left(\frac{\partial V}{\partial z}\right)^2\right]e^{2(u_0+\Phi)}.\tag{27}$$

Suppose that the solution of (27) is

$$u_0=\frac{a}{r^2+z^2}+\frac{b}{(r^2+z^2)^{3/2}}+\frac{cr^2+dz^2}{(r^2+z^2)^3},\tag{28}$$

where a, b, c, and d are undetermined coefficients. Let

$$\alpha=(\kappa_0/2)(kQ)^2=4\pi G(kQ)^2,\tag{29}$$

$$\beta=I_3-I_1.\tag{30}$$

From (20), (25), and (28), we obtain that the right-hand side of (27) is

$$\begin{aligned}&\frac{\kappa_0}{2}\left[\left(\frac{\partial V}{\partial r}\right)^2+\left(\frac{\partial V}{\partial z}\right)^2\right]e^{2(u_0+\Phi)}\\&=\frac{\alpha}{(r^2+z^2)^2}-\frac{2GM\alpha}{(r^2+z^2)^{5/2}}\\&\quad+\frac{2\alpha a-6\alpha\beta+2\alpha(GM)^2}{(r^2+z^2)^3}+\frac{9\alpha\beta r^2}{(r^2+z^2)^4}.\end{aligned}$$

The above and following equations are taken approximately up to $1/r^6$ or $1/z^6$, then the left-hand side of (27) becomes

$$\begin{aligned}\frac{\partial^2 u_0}{\partial r^2}+\frac{1}{r}\frac{\partial u_0}{\partial r}+\frac{\partial^2 u_0}{\partial z^2}&=\frac{2a}{(r^2+z^2)^2}+\frac{6b}{(r^2+z^2)^{5/2}}\\&\quad+\frac{4c+8d}{(r^2+z^2)^3}+\frac{6(c-d)r^2}{(r^2+z^2)^4}.\end{aligned}$$

Comparing the coefficients of the two sides of (27), we have

$$\begin{aligned}&a=\alpha/2,\quad b=-\tfrac{1}{3}\alpha GM,\\&c=\alpha^2/12+\tfrac{1}{2}\alpha\beta+\tfrac{1}{6}\alpha(GM)^2,\\&d=\alpha^2/12-\alpha\beta+\tfrac{1}{6}\alpha(GM)^2.\end{aligned}\tag{31}$$

Next, we find the expression for κ. Since the right-hand side of (23) is

$$\begin{aligned}&-\kappa_0\left(\frac{\partial V}{\partial r}\right)\left(\frac{\partial V}{\partial z}\right)e^{2u}\\&\approx-\frac{2\alpha rz}{(r^2+z^2)^3}+\frac{4\alpha GMrz}{(r^2+z^2)^{7/2}}\\&\quad-\frac{[2\alpha^2+4\alpha(GM)^2-18\alpha\beta]rz}{(r^2+z^2)^4}-\frac{30\alpha\beta r^3z}{(r^2+z^2)^5},\end{aligned}$$

and

$$\begin{aligned}2\frac{\partial u}{\partial r}\frac{\partial u}{\partial z}&\approx\frac{2(GM)^2rz}{(r^2+z^2)^3}-\frac{4\alpha GMrz}{(r^2+z^2)^{7/2}}\\&\quad+\frac{[2\alpha^2+4\alpha(GM)^2-18\beta(GM)^2]rz}{(r^2+z^2)^4}\\&\quad+\frac{30\beta(GM)^2r^3z}{(r^2+z^2)^5}.\end{aligned}$$

Then Eq. (23) becomes

$$\begin{aligned}\frac{1}{r}\frac{\partial\kappa}{\partial z}&=\frac{2[(GM)^2-\alpha]rz}{(r^2+z^2)^3}+\frac{18\beta[\alpha-(GM)^2]rz}{(r^2+z^2)^2}\\&\quad+\frac{30\beta[(GM)^2-\alpha]r^3z}{(r^2+z^2)^5}.\end{aligned}$$

Integrating the above equation, we have

$$\kappa=\frac{Ar^2}{2(r^2+z^2)^2}-\frac{3\beta Ar^2}{(r^2+z^2)^3}+\frac{15\beta Ar^4}{4(r^2+z^2)^4}+f(r),\tag{32}$$

where $A=\alpha-(GM)^2$. Now, since (22) becomes

$$\begin{aligned}\frac{1}{r}\frac{\partial\kappa}{\partial r}&\simeq\frac{-A}{(r^2+z^2)^2}+\frac{2Az^2}{(r^2+z^2)^3}\\&\quad-\frac{3A\beta}{(r^2+z^2)^3}+\frac{27\beta Az^2}{(r^2+z^2)^4}-\frac{30\beta Az^4}{(r^2+z^2)^5}\end{aligned}$$

at the same time, its solution is

$$\kappa = \frac{Ar^2}{2(r^2+z^2)^2} - \frac{3\beta Ar^2}{(r^2+z^2)^3} + \frac{15\beta Ar^4}{4(r^2+z^2)^4} + g(z). \tag{33}$$

Since $r\to\infty$, $\kappa\to 0$, hence $f(r) = g(z) = 0$ and

$$\kappa = \frac{Ar^2}{2(r^2+z^2)^2} - \frac{3\beta Ar^1}{(r^2+z^2)^3} + \frac{15\beta Ar^4}{4(r^2+z^2)^4}. \tag{34}$$

Finally, we obtain the nonzero components of the metric tensor, i.e.,

$$g_{00} = -\exp\left[2\Phi + \frac{2a}{r^2+z^2} + \frac{2b}{(r^2+z^2)^{3/2}} + \frac{2(cr^2+dz^2)}{(r^2+z^2)^3}\right], \tag{35}$$

$$g_{11} = g_{22} = \exp\left\{-2\Phi - \frac{2a}{r^2+z^2} + \frac{Ar^2}{(r^2+z^2)^2} - \frac{2b}{(r^2+z^2)^{3/2}} + \frac{2[(c+3\beta A)r^2+dz^2]}{(r^2+z^2)^3} + \frac{15\beta Ar^4}{2(r^2+z^2)^4}\right\}, \tag{36}$$

$$g_{33} = r^2\exp\left\{-2\Phi - \frac{2a}{r^2+z^2} - \frac{2b}{(r^2+z^2)^{3/2}} - \frac{2(cr^2+dz^2)}{(r^2+z^2)^3}\right\}. \tag{37}$$

V. CONCLUSION

Equations (35)–(37) are the general formulas of the exterior metric tensor that are applicable to all kinds of charged axisymmetric celestial bodies. The constant β in c and d is different for different kinds of celestial bodies. For example, for a cylindrical body, β is $R^2/4 - l^2/3$ (where R and $2l$ are the radius and length of the cylinder, respectively). For a cone body, β is $3R^2/20 - 3h^2/80$ (where R and h are the radius of the bottom and the altitude of the cone, respectively). When $r\to\infty$ or $z\to\infty$ and/or $r = z\to\infty$, this metric tensor tends to the metric tensor of Minkowski space. And when $Q = 0$, i.e., there is no charge in the celestial body, then

$$g_{00} = -\exp(2\Phi),$$

$$g_{11} = g_{22} = \exp\left\{-2\Phi - \frac{(GM)^2r^2}{(r^2+z^2)^2} + \frac{6\beta(GM)^2r^2}{(r^2+z^2)^3} - \frac{15\beta(GM)^2r^4}{(r^2+z^2)^4}\right\},$$

$$g_{33} = r^2\exp(-2\Phi).$$

are coincident with the PPN.[4,5] The only difference in these methods is the difference in their deriving process; the former approximation is taken at the end of the deriving process and the latter is taken at the beginning.

[1] M. Carmeli, *Classical Fields, General Relativity and Gauge Theory* (Wiley, New York, 1982), Secs. 4.3 and 4.4.

[2] H. Goldstein, *Classical Mechanics* (Addison–Wesley, London, 1980), 2nd ed., pp. 225–228.

[3] D. Kramer, H. Stephani, and E. Herlt, *Exact Solution of Einstein's Field Equation* (Deutscher Verlag der Wissenschaften, 1979), p. 6.

[4] C. M. Will, *The Theoretical Tools of Experimental Gravitation* (Academic, London, 1974), Sec. 4.2.

[5] S. W. Hawking and W. Israel, *An Einstein Centenary Survey: General Relativity* (Cambridge U.P., London, 1979), pp. 42–45.

Collision of impulsive gravitational waves followed by dust clouds

A. H. Taub
Mathematics Department, University of California, Berkeley, California 94720

(Received 17 May 1988; accepted for publication 3 August 1988)

The evolution of a space-time containing two colliding plane impulsive gravitational waves each of whose leading edges is followed by a distribution of null dust is determined. The conditions on the Ricci tensor that ensure that the evolution of such a space-time is unambiguous are determined. These are the same as those that apply in the planar case. The equations of motion of the medium contained in the region of interaction of the dust clouds are determined. These equations determine the change in energy density of each dust cloud as the interaction proceeds and involve the functions whose specifications ensure that the evolution of the space-time is unambiguous.

I. INTRODUCTION

Chandrasekhar and Xanthopolous determined two different exact solutions of the Einstein field equations that describe two different space-times, each containing two colliding impulsive plane gravitational waves, each wave being followed by a distribution of null dust. In the first solution[1] it was assumed that the region of interaction of the dust clouds contained a perfect fluid with energy density ϵ equal to pressure p and in the second[2] it was assumed that this region is filled with a mixture of null dusts moving in opposite directions. The space-times involved are identical in the regions before the instant of collision of the impulsive waves. Thus different assumptions concerning the nature of the energy-momentum tensor in the region of interaction of the dust clouds lead to different exact solutions to the Einstein field equations, that is, to different solutions to the Cauchy initial value problem posed on a spacelike three-surface that intersects the wave fronts at a time earlier than the instant of collision of the impulsive wave fronts.

In a recent paper[3] the Einstein equations involving distribution valued curvature tensors for space-times admitting the three-dimensional group of motions of the Euclidean two-plane were discussed. It was shown that a similar phenomenon exists, namely, if such a space-time contains two colliding null hypersurfaces containing null dust, each of which is followed by a distribution of null dust, then the Cauchy data, on a spacelike three-surface that intersects the fronts of the dust distributions before the instant of collision, does not lead to a unique solution to the Einstein field equations unless additional conditions are imposed on the energy-momentum tensor in the region of interaction of the two dust clouds.

In this paper the evolution of a space-time containing two colliding plane impulsive gravitational waves whose leading edges may be followed by distributions of null dust is discussed. It will be shown that the generalized Einstein field equations when written in terms of the Ricci tensor are identical with those that occur in Ref. 3. The conditions on the Ricci tensor that were shown to be sufficient to make the solutions of the Cauchy problem unique in case two planar singular dust hypersurfaces collide are the same as those that are sufficient to make the solutions of the Cauchy problem unique when two impulsive plane gravitational waves, each followed by a dust cloud, collide.

In addition, the equations of motion of the medium in the region of interaction of the dust clouds will be determined. It will be shown that the conditions imposed on the Ricci tensor in order to ensure the uniqueness of the Cauchy problem are those that determine the interaction of the dust clouds with each other, namely, the change in energy density of each dust cloud as the interaction proceeds.

Space-times that admit two commuting spacelike Killing vectors admit coordinate systems in which the line element is given by

$$ds^2 = g_{ij}\, dx^i\, dx^j + g_{AB}\, dx^A\, dx^B, \tag{1.1}$$

where

$$\alpha, \beta = 0,1,2,3, \quad i,j = 0,3, \quad A,B = 1,2, \tag{1.2}$$

$$g_{\alpha\beta}\delta^\alpha_i \delta^\beta_j = g_{ij} = e^{\omega}\eta_{ij} = e^{\omega}(\delta_{ij} - 2\delta^3_i\delta^3_j), \tag{1.3}$$

$$g_{\alpha\beta}\delta^\alpha_A \delta^\beta_B = g_{AB} = -e^{\mu}\gamma_{AB}, \tag{1.4}$$

$$\gamma_{11} = \chi^{-1}, \quad \gamma_{12} = -q_2\chi^{-1}, \quad \gamma_{22} = \chi + q_2^2/\chi, \tag{1.5}$$

$$-{}^4g = -\det\|g_{\alpha\beta}\| = e^{2(\omega+\mu)}. \tag{1.6}$$

The quantities ω, μ, χ, q_2 are functions of x^0 and x^3 alone. They may also be considered as functions of the null coordinates

$$u = x^0 - x^3, \tag{1.7}$$

$$v = x^0 + x^3, \tag{1.8}$$

The null hypersurfaces $u = 0$ and $v = 0$ will be used to divide a region of space-time into four subregions:

region I, where $u > 0$ and $v > 0$;
region II, where $u > 0$ and $v < 0$;
region III, where $u < 0$ and $v > 0$;
region IV, where $u < 0$ and $v < 0$.

The hypersurface $u = 0$ ($v = 0$) will be interpreted as the wave front of a gravitational wave traveling in the x^3 ($-x^3$) direction. The variables

$$2x^0 = (v + u) \tag{1.9}$$

and

$$2x^3 = (v - u) \tag{1.10}$$

are measures of the time from $(u = 0, v = 0)$, the instant of collision, and the distance between the wave fronts, respectively.

The study of the collision and subsequent interaction of plane impulsive gravitational waves in a vacuum or when coupled with the motion of a medium such as a perfect fluid proceeds as follows: Exact solutions of the Einstein field equations with the source terms derived from the energy momentum tensor of the medium are sought. They are to be such that the metric tensor is of the form given by Eq. (1.1) and are to involve the variables u and v in region I. The resulting $g_{\alpha\beta}(u,v)$ is extended to region II (III) by taking the metric there to be $g_{\alpha\beta}(u,0)$ $[g_{\alpha\beta}(0,v)]$. It is further extended to region IV by taking the metric there to be $g_{\alpha\beta}(0,0)$ and thereby ensuring that region IV is flat.

This method of extending the solution of region I produces metrics that are continuous across the hypersurfaces $u = 0$ and $v = 0$ but may have discontinuous first derivatives across these hypersurfaces. If so, the curvature tensor derived from $g_{\alpha\beta}$ will be distribution valued, i.e., will contain delta functions with support on these null hypersurfaces.

In addition, space-times with metrics obtained as above are said to contain impulsive gravitational waves only if the components of the Einstein tensor (equivalently the Ricci tensor) do not contain such delta functions.

II. THE STRENGTH OF AN IMPULSIVE PLANE WAVE

The wave front of such a wave, the null hypersurface, is described by the equation

$$\phi(x) = 0, \tag{2.1}$$

where

$$\phi(x) = u \tag{2.2}$$

or

$$\phi(x) = v \tag{2.3}$$

if the coordinates of space-time are (u, v, x^1, x^2) and the line element is given by Eqs. (1.1)–(1.5). In case Eq. (2.2) [(2.3)] holds, the wave is traveling in the positive (negative) x^3 direction. We shall denote by

$$[f] = f^+ - f^-, \tag{2.4}$$

where

$$f^+(x) = \lim_{u\to 0^+} f(u,v) \quad \left[\lim_{v\to 0^+} f(u,v)\right], \tag{2.5a}$$

the value of the limit of the quantity $f(x)$ as the event in region I approaches the hypersurface $u = 0$ $(v = 0)$. Similarly $f^-(x)$ denotes the limit of $f(x)$ as the event in region III (II) approaches the hypersurface $u = 0$ $(v = 0)$. That is,

$$f^-(x) = \lim_{u\to 0^-} f(u,v) \quad \left[\lim_{v\to 0^-} f(u,v)\right]. \tag{2.5b}$$

It has been shown in Ref. 4 that the discontinuities of the first derivatives of the metric tensor across the hypersurface Σ are characterized by the tensor value functions of events on Σ, $b_{\mu\nu}$, defined by the equations

$$l_\sigma b_{\mu\nu} = [g_{\mu\nu,\sigma}], \tag{2.6}$$

where

$$l_\sigma = \frac{\partial\phi}{\partial x^\sigma} = \phi_{,\sigma}. \tag{2.7}$$

It has been further shown in Ref. 4 that the components of the Ricci tensor computed from the $g_{\alpha\beta}$ with first derivative discontinuous across Σ will have components free of delta functions if

$$l_\rho g^{\sigma\rho} b'_{\sigma\tau} = l_\rho g^{\sigma\rho}(b_{\sigma\tau} - \tfrac{1}{2} g_{\sigma\tau} b) = l_\rho b'^{\rho}{}_{\tau} = 0, \tag{2.8}$$

where

$$b = g^{\sigma\rho} b_{\sigma\rho} \tag{2.9}$$

and we have raised indices by use of the metric tensor on Σ.

If Σ is part of the null hypersurface $v = 0$ separating regions I and III and the line element is given by Eq. (1.1), then

$$b_{ij}(u) = g^+_{ij}(\omega_{,v})^+, \qquad b_{AB}(u) = (g_{AB,v})^+. \tag{2.10}$$

When Σ is part of the null hypersurface $v = 0$ separating regions III and IV, the $b_{\mu\nu}$ are constant given by setting $u = 0$ in the right-hand sides of Eqs. (2.10).

It follows from Eqs. (2.9) and (2.10) that

$$b(u) = 2((\omega_{,v})^+ + (\mu_{,v})^+) \tag{2.11}$$

and hence

$$g^{ik} b'_{kj} = b'^{i}_{j} = -(\mu_{,v})^+ \delta^i_j,$$
$$g^{AC} b'_{CB} = b'^{A}{}_{B} = (g^{AC} g_{CB,v})^+ - ((\omega_{,v})^+ + (\mu_{,v})^+)\delta^A_B. \tag{2.12}$$

Thus Eq. (2.8) obtains if

$$[\mu_{,v}] = (\mu_{,v})^+ = 0. \tag{2.13}$$

It has also been shown in Ref. 4 that if Eq. (2.8) holds and if

$$[R_{\mu\nu}] = 0, \tag{2.14}$$

then

$$(\tau l^\sigma)_{;\sigma} = 0, \tag{2.15}$$

where

$$\tau = b^{\mu\nu} b_{\mu\nu} - b^2/2 \tag{2.16}$$

and the semicolon in Eq. (2.15) denotes the covariant derivative in Σ as computed from the Christoffel symbols involving either $(g_{\alpha\beta,\gamma})^+$ or $(g_{\alpha\beta,\gamma})^-$. In view of this conservation theorem we may regard τ as measuring the strength of the impulsive gravitational wave.

When the space-time metric is given by Eqs. (1.1)–(1.6) we have when Σ is the hypersurface $v = 0$,

$$\tau = 2e^{-2\mu}([g_{12,v}]^2 - [g_{11,v}][g_{22,v}]), \tag{2.17}$$

as a consequence of Eq. (2.13). Thus

$$\tau = 2\left\{\left[\left(\frac{q_{2,v}}{\chi}\right)^+\right]^2 + \left[\left(\frac{\chi_{,v}}{\chi}\right)^+\right]^2\right\} \tag{2.18}$$

as follows from Eq. (1.5).

A similar discussion of the boundary separating regions I and III, namely Σ given by $u = 0$, leads to the equations

$$b_{ij}(v) = g_{ij}^+ (\omega_{,u})^+,$$
$$b_{AB}(v) = (g_{AB,u})^+, \tag{2.10'}$$

and

$$b(v) = 2((\omega_{,v})^+ + (\mu_{,v})^+), \tag{2.11'}$$

$$b'^i_j = -(\mu_{,u})^+ \delta^i_j,$$
$$b'^A{}_B = (g^{AC} g_{CB,u})^+ - ((\omega_{,u})^+ + (\mu_{,u})^+)\delta^A_B. \tag{2.12'}$$

Hence Eq. (2.8) holds if

$$[\mu_{,u}] = (\mu_{,u})^+ = 0, \tag{2.13'}$$

Equation (2.15) holds where now $l_\mu = u_{,\mu} = \delta^u_\mu$ and

$$\tau = 2\{[(q_{2,u}/\chi)^+]^2 + [(\chi_{,u}/\chi)^+]^2\}. \tag{2.18'}$$

III. THE RICCI TENSOR

It is a consequence of the fact that the coordinates of space-time may be chosen so that the vectors δ^μ_1 and δ^μ_2 are Killing vectors that the line element of space-time is given by Eq. (1.1), where the g_{ij} need not satisfy Eq. (1.3). It is a further consequence that the components of the Ricci tensor are such that

$$R^i{}_A = 0 \tag{3.1}$$

and

$$2R^C{}_A = (e^{-\mu}/\sqrt{-g})(e^\mu \sqrt{-g} g^{ij} g^{CB} g_{CA,i})_{,j}, \tag{3.2}$$

where

$$g = \det\|g_{ij}\|. \tag{3.3}$$

In case Eqs. (1.3) obtain

$$-g = e^\omega, \tag{3.4}$$

$$R_{uu} = \mu_{,uu} + \tfrac{1}{2}\mu^2_{,u} - \mu_{,u}\omega_{,u} - S_{uu}, \tag{3.5}$$

$$R_{uv} = \mu_{,uv} + \tfrac{1}{2}\mu_{,u}\,\mu_{,v} + \omega_{,uv} - S_{uv}, \tag{3.6}$$

$$R_{vv} = \mu_{,vv} + \tfrac{1}{2}\mu^2_{,v} - \mu_{,v}\omega_{,v} - S_{uv}, \tag{3.7}$$

where

$$S_{ij} = \tfrac{1}{4}\gamma^{AB}_{,i}\gamma_{AB,j} = -(1/2\chi^2)(\chi_{,i}\chi_{,j} + q_{2,i}q_{2,j}), \tag{3.8}$$

γ_{AB} is given by Eq. (1.5), γ^{AB} is such that

$$\gamma^{AB}\gamma_{BC} = \delta^A{}_C, \tag{3.9}$$

that is

$$\gamma^{11} = \chi + q_2^2/\chi, \quad \gamma^{12} = q_2/\chi, \quad \gamma^{22} = 1/\chi. \tag{3.10}$$

It follows from Eqs. (3.2) and (3.4) that

$$2R^C{}_C = e^{-(\mu+\omega)}(e^\mu \eta^{ij})_{,ij} = e^{-\omega}(\mu_{,ij} + \mu_{,i}\,\mu_{,j})\eta^{ij} \tag{3.11}$$

and that

$$2R^A{}_B = 2R^C{}_C\delta^A{}_B + e^{-(\mu+\omega)}(e^\mu \gamma^{CA}\gamma_{BC,i}\eta^{ij})_{,j}. \tag{3.12}$$

The equations $R_{AB} = 0$ then imply that

$$(e^\mu)_{,uv} = e^\mu(\mu_{,uv} + \mu_{,u}\,\mu_{,v}) = 0 \tag{3.13}$$

and

$$(e^\mu \gamma_{,u}\gamma^{-1})_{,v} + (e^\mu \gamma_{,v}\gamma^{-1})_{,u} = 0, \tag{3.14}$$

where

$$\gamma = \|\gamma_{AB}\|. \tag{3.15}$$

In view of Eqs. (1.5) and (3.10), Eq. (3.12) may be written as

$$(e^\mu q_{2,u}/\chi^2)_{,v} + (e^\mu q_{2,v}/\chi^2)_{,u} = 0 \tag{3.16}$$

and

$$(e^\mu \chi_{,u}/\chi)_{,v} + (e^\mu \chi_{,v}/\chi)_{,u} = -2(e^\mu/\chi^2)(q_{2,u}q_{2,v}). \tag{3.17}$$

In the coordinate system in which Eqs. (1.1) and (1.3) hold, the Einstein field equations supplemented by the condition $R_{AB} = 0$ consist of the system of Eqs. (3.5)–(3.7) supplemented by Eqs. (3.13), (3.16), and (3.17). In these equations the R_{ij} are assumed to be given as functions of u and v. These quantities must of course satisfy the integrability conditions for the determination of ω from Eqs. (3.5)–(3.7).

Note that if μ and the matrix γ are solutions of Eqs. (3.13) and (3.14) [equivalently (3.13), (3.16), and (3.18)] and ω^0 is a solution of Eqs. (3.5)–(3.7) with $R_{ij} = 0$, then

$$\omega = \Omega + \omega^0$$

satisfies the latter equations with $R_{ij} \neq 0$ if Ω is a solution of the equations

$$R_{uu} = -\mu_{,u}\Omega_{,u}, \tag{3.18}$$

$$R_{uv} = \mu_{,uv}, \tag{3.19}$$

$$R_{vv} = -\mu_{,v}\Omega_{,v}. \tag{3.20}$$

The integrability conditions of these equations are

$$((\mu_{,u})^{-1}R_{uu})_{,v} = ((\mu_{,v})^{-1}R_{vv})_{,u} = -R_{uv}. \tag{3.21}$$

These conditions restrict the possible sources of the gravitational fields described by $g_{\mu\eta}$.

It should be noted that in Refs. 1 and 2 Chandrasekhar and Xanthopolous use coordinates (y^0, y^3, x^1, x^2) given in terms of the coordinates (u, v, x^1, x^2) used above by the equations

$$y^0 = \cos((v-u)/2), \tag{3.22}$$

$$y^3 = \cos((v+u)/2), \tag{3.23}$$

so that the line element given by Eq. (1.11) becomes

$$dS^2 = e^\omega\left(\frac{(dy^0)^2}{1-(y^0)^2}\right) - \left(\frac{(dy^3)^2}{1-(y^3)^2}\right) - e^\mu(\gamma_{AB}\,dx^A\,dx^B). \tag{3.24}$$

In Refs. 1 and 2 the variables y^0 and y^3 are denoted by η and μ, respectively, whereas the quantity denoted by μ above is taken to be

$$2e^\mu = \cos u - \cos v. \tag{3.25}$$

The solutions of the transforms of Eqs. (3.16) and (3.17) are shown to be expressed in terms of the Ernst equation.

IV. THE VACUUM EQUATIONS

These equations are

$$R_{\alpha\beta} = 0.$$

That is, the R_{ij} vanish in Eqs. (3.18)–(3.20). Thus Ω is a constant that may be taken to be zero. In addition, Eqs.

(3.13) and (3.14), or equivalently, Eqs. (3.13), (3.16), and (3.17) hold as a consequence of $R_{AB}=0$. Either set of the latter equations may be used to determine $\gamma_{AB}(u,v)$, that is, $\chi(u,v)$ and $q_2(u,v)$. Finally ω^0 may be determined as solutions of the equations

$$S_{uu}=\mu_{,uv}+\tfrac{1}{2}\mu_{,u}^2-\mu_{,u}\omega^0_{,u}, \tag{4.1}$$

$$S_{uv}=\mu_{,uv}+\tfrac{1}{2}\mu_{,uv}+\tfrac{1}{2}\mu_{,u}\,\mu_{,v}+\omega^0_{,uv}, \tag{4.2}$$

$$S_{vv}=\mu_{,vv}+\tfrac{1}{2}\mu_{,v}^2-\mu_{,v}\omega^0_{,v}, \tag{4.3}$$

where μ satisfies Eq. (3.13) and the S_{ij} are defined by Eqs. (3.8). The integrability conditions of Eqs. (4.1)–(4.3) considered as equations for ω^0 given γ_{AB} and μ are satisfied as a consequence of Eqs. (3.13), (3.16), and (3.17) or equivalently of Eqs. (3.13) and (3.14).

Chandrasekhar and Ferrari[5] showed that solutions of the Ernst equation may be used to determine the $\gamma_{AB}(u,v)$ leading to the solution of Khan and Penrose[6] and of Nutku and Hallil[7] for colliding plane impulsive gravitational waves. Ferrari and Ibanez[8] have used the inverse scattering method of Belinski and Zakharov[9] to obtain the $\gamma_{AB}(uv)$.

One may determine a solution of the equations describing the collision of impulsive gravitational waves by solving Eqs. (4.1)–(4.3) after determining μ and γ_{AB} by either of the two methods mentioned above.

To determine the impulsive waves undergoing the collision, the solution in region I is extended to regions II, III, and IV as described in Sec. II. At the wave front $v=0$

$$\tau(u)=-(S_{vv})^+ \tag{4.4}$$

and on $u=0$

$$\tau(v)=-(S_{uu})^+, \tag{4.5}$$

where $\tau(u)$ and $\tau(v)$ are given by Eqs. (2.18) and (2.18′), respectively. They measure the strengths of the respective impulsive waves.

The situation is simpler when the plane impulsive gravitational waves are linearly polarized, that is, when

$$q_2=0, \tag{4.6}$$

and the R_{ij} form a 2×2 matrix of rank 1, that is,

$$R_{ij}=-\tau_i\tau_j. \tag{4.7}$$

As a consequence of Eq. (3.13) we have

$$e^{\mu}=1+U(u)+V(v), \tag{4.8}$$

where U and V are functions of u alone and v alone, respectively. It is no restriction to assume that

$$U(0)=V(0)=0. \tag{4.9}$$

Equations (2.13) and (2.13′) imply that

$$U'(0)=V'(0)=0, \tag{4.10}$$

where the prime denotes the derivatives of the functions U and V with respect to their arguments.

If we now set

$$\omega^0=\Omega^0-\tfrac{1}{2}\mu+\ln(U'(u)V'(v)), \tag{4.11}$$

Eqs. (4.1)–(4.3) become

$$(1+U+V)S_{uu}=-U'\Omega^0_{,u}, \tag{4.12}$$

$$S_{uv}=\Omega^0_{,uv}, \tag{4.13}$$

$$(1+u+v)S_{vv}=-V'\Omega^0_{,v}, \tag{4.14}$$

respectively. These equations are identical with Eqs. (3.18)–(3.20) if the S_{ij} and R_{ij} are identified and μ is given by Eq. (4.8).

Further, the former equations and Eq. (4.8) are the same as the equations derived in Ref. 3 for the determination of the line element of a plane symmetric space-time for which the line element is given by Eq. (1.1) with Eqs. (1.2) and (1.3) holding and Eq. (1.4) replaced by

$$g_{AB}=-e^{\mu}\delta_{AB}.$$

It is a consequence of Eqs. (4.6), (3.16), and (3.17) that if

$$\sqrt{2}\sigma=\ln\chi, \tag{4.15}$$

then

$$(e^{\mu}\sigma_{,u})_{,v}+(e^{\mu}\sigma_{,v})_{,u}=0 \tag{4.16}$$

and

$$S_{ij}=-\sigma_{,i}\sigma_{,j}. \tag{4.17}$$

It follows from Eqs. (4.7) and (3.21) that there exists a function λ such that

$$\tau_i=\lambda_{,i} \tag{4.18}$$

and λ satisfies Eqs. (4.16).

When $R_{AB}=0$ and Eqs. (4.7) and (4.18) hold, the energy momentum tensor of space-time may be interpreted as either one due to a scalar field or as one due to a perfect fluid with energy density equal to pressure.

The general solution of Eqs. (4.6) and (4.16) has been given by Szekeres[10] and by Tabensky and Taub.[11] In Ref. 10 a class of specific solutions of these equations is listed. This class includes the solution given earlier by Szekeres[12] and the one given by Kahn and Penrose.[6]

V. $R_{AB}=0$ and $R_{ij}\neq 0$

Suppose that a space-time is such that in region I its line element is given by Eq. (1.1) where $\mu(u,v)$, $\gamma_{AB}(u,v)$, and $\omega(u,v)$ are such that μ satisfies Eq. (4.8), γ_{AB} satisfies Eq. (3.14) [equivalently (3.16) and (3.17)] and

$$\omega=\Omega+\Omega^0-\tfrac{1}{2}\mu+\ln(U'(u)V'(v)), \tag{5.1}$$

where Ω^0 satisfies Eqs. (4.12)–(4.14) and Ω is related to the R_{ij} by means of Eqs. (3.18)–(3.20). Then if this line element is extended to regions II–IV, as described in Sec. I, one finds that the latter equations reduce to

$$R^{\rm II}uu=-\mu_{,u}\Omega,u, \tag{5.2}$$

$$R^{\rm II}uv=R^{\rm II}_{vv}=0, \tag{5.3}$$

$$\mu^{\rm II}=1+U(u), \tag{5.4}$$

in region II.

In region III one finds

$$R^{\rm III}uu=R^{\rm III}uv=0, \tag{5.5}$$

$$R^{\rm III}vv=-\mu_{,v}\Omega_{,v}, \tag{5.6}$$

with

$$\mu^{\rm III}=1+V(v). \tag{5.7}$$

In region IV one finds

$$R^{\mathrm{IV}}{}_{ij}=0 \tag{5.8}$$

and

$$\mu^{\mathrm{II}}=1. \tag{5.9}$$

Thus in region II the Einstein energy-momentum tensor $T^{\mu\nu}$ given by

$$-\kappa T^{\mu\nu}=G^{\mu\nu} \tag{5.10}$$

is that of a null dust cloud moving in the x^3 direction behind the wave front $v=0$. Similarly in region III the energy-momentum tensor is that of a null dust cloud moving in the $-x^3$ direction behind the wave front $u=0$. Region IV has a vanishing energy-momentum tensor and is flat.

Region I of such a space-time is said to be the region of interaction of the two null dust clouds moving in regions II and III. If we are given R^{II}_{ij} and R^{II}_{ij} then $\Omega^{\mathrm{I}}(u,0)$ and $\Omega^{\mathrm{I}}(0,v)$ are determined but $\Omega^{\mathrm{I}}(u,v)$ is not for $u>0$ and $v>0$. This is the situation that obtains when two impulsive null hypersurfaces $u=0$ and $v=0$ each followed by distributions of null dust collide at $u=0$ and $v=0$ (cf. Ref. 3). Thus the evolution of a plane symmetric space-time in which two plane symmetric impulsive gravitational waves, each followed by a distribution of null dust, are not uniquely determined by the Einstein equations in the region of interaction of the two dust clouds. That is, the solutions of the generalized Einstein field equations are not unique if only the initial values of the metric tensor and its derivatives are prescribed on a spacelike hypersurface that intersects the wave fronts before the instant of collision. In other words, in such a case the Cauchy problem does not have an unique solution.

Chandrasekhar and Xanthopolous in Refs. 1 and 2 have selected two different choices of $\Omega^{\mathrm{I}}(u,v)$ for the same values of $\Omega^{\mathrm{I}}(0,v)$ and $\Omega^{\mathrm{I}}(u,0)$ by imposing two different requirements on the energy momentum tensor in region I. In the next section we show how the nature of the interaction of the dust cloud in region II with that in region III determines the energy-momentum tensor in region I and hence the function $\Omega^{\mathrm{I}}(u,v)$.

VI. THE EQUATIONS $T^{\mu\nu}{}_{;\nu}=0$

That the above description of a collision of two dust clouds each fronted by an impulsive gravitational wave does not lead to a unique outcome is due to the fact that no account has been taken of the nature of the interaction of these dust clouds. The situation is similar to that which obtains in the theory of self-gravitating perfect fluids. In that theory the energy-momentum tensor is not completely specified until either a relation between energy density and pressure is prescribed or a caloric equation of state is given and conservation of particle number is postulated. Such additional assumptions enable one to deal with the equations of motion of the fluid and the determination of the gravitational field. Without such additional assumptions the physical and mathematical problems are not completely specified.

In this section we shall determine the energy momentum tensor in region I and determine the equations of motion of the medium contained therein from its divergence.

In the coordinate system in which Eqs. (1.1) and (1.7)–(1.9) hold we have

$$\frac{\partial u}{\partial x^\mu}=u_{,\mu}=\delta^0_\mu-\delta^3_\mu, \tag{6.1a}$$

$$\frac{\partial v}{\partial x^\mu}=v_{,\mu}=\delta^0_\mu+\delta^3_\mu, \tag{6.1b}$$

and

$$\frac{\partial x^\mu}{\partial u}=\frac{1}{2}(\delta^\mu_0-\delta^\mu_3), \tag{6.2a}$$

$$\frac{\partial x^\mu}{\partial v}=\frac{1}{2}(\delta^\mu_0+\delta^\mu_3). \tag{6.2b}$$

Let

$$u^\mu=g^{\mu\nu}u_{,\nu}=e^{-\omega}(\delta^\mu_0+\delta^\mu_3), \tag{6.3a}$$

$$v^\mu=g^{\mu\nu}v_{,\nu}=e^{-\omega}(\delta^\mu_0-\delta^\mu_3), \tag{6.3b}$$

then

$$u^\mu u_{,\mu}=u^i v_{,i}=v^\mu v_{,\mu}=v^i v_{,i}=0, \tag{6.4a}$$

$$u^\mu v_{,\mu}=v^\mu u_{,\mu}=v^i u_{,i}=u^i v_{,i}=2e^{-\omega}, \tag{6.4b}$$

and

$$u_{,i}v_{,j}+u_{,j}v_{,i}-g_{ij}(u^i v_{,j})=0. \tag{6.5}$$

It follows that

$$-R_{ij}=\epsilon_1 u_{,i}u_{,j}+\epsilon_2 v_{,i}v_{,j}+C(u_{,i}v_{,j}+u_{,j}v_{,i}), \tag{6.6}$$

where

$$\epsilon_1=-R_{ij}\frac{\partial x^i}{\partial u}\frac{\partial x^j}{\partial v}=-R_{uu}, \tag{6.7a}$$

$$\epsilon_2=-R_{ij}\frac{\partial x^i}{\partial v}\frac{\partial x^j}{\partial v}=-R_{vv}, \tag{6.7b}$$

$$C=-R_{ij}\frac{\partial x^i}{\partial^u}\frac{\partial x^j}{\partial^v}=-R_{uv}=-R_{vu}. \tag{6.7c}$$

In addition since $R_{AB}=0$, the scalar curvature obeys

$$-R=2C(u_{,i}v^i)=4Ce^{-\omega}. \tag{6.8}$$

Hence

$$-G_{ij}=-(R_{ij}-\tfrac{1}{2}Rg_{ij})=\epsilon_1 u_{,i}v_{ij}+\epsilon_2 v_{,i}v_{,i} \tag{6.9a}$$

and

$$-G_{AB}=\tfrac{1}{2}Rg_{AB}=-2Ce^{-\omega}g_{AB}. \tag{6.9b}$$

That is,

$$-G_{\mu\nu}=\epsilon_1 u_{,\mu}u_{,\nu}+\epsilon_2 v_{,\mu}v_{,\nu}-2Ce^{-\omega}g_{AB}\delta^A_\mu\delta^B_\nu. \tag{6.9c}$$

The energy momentum tensor $T_{\mu\nu}$ of the medium in region I is given by

$$-\kappa T_{\mu\nu}=G_{\mu\nu}, \tag{6.10}$$

where κ is the Einstein gravitational constant and $G_{\mu\nu}$ is given by Eq. (6.9c). This energy-momentum tensor is that of a medium consisting of two null fluids one with null four-velocity $u_{,\mu}$ and energy density ϵ_1 and the other with null four-velocity $v_{,\mu}$ and energy density ϵ_2. They are acted upon by transverse stresses described by the last term in Eq. (6.9c).

The nature of this interaction may be determined from the equations of motion satisfied by the null fields, namely the equations

$$-\kappa T^{\mu\nu}{}_{;\nu} = G^{\mu\nu}{}_{;\nu} = 0.$$

These equations are

$$G^{\mu\nu}{}_{;\nu} = e^{-(\omega+\mu)}(e^{(\omega+\mu)}G^{\mu\nu})_{,\nu} + G^{\rho\sigma}\Gamma^{\mu}{}_{\rho\sigma} = 0.$$

Thus

$$e^{-(\omega+\mu)}(e^{\omega+\mu}G^{ij})_{j} + G^{kl}\Gamma^{i}_{kl} + G^{AB}\Gamma^{i}_{AB} = 0, \quad (6.11)$$

where Γ^i_{kl} are the Christoffel symbols formed from the g_{ij} and

$$\Gamma^{i}_{AB} = -\tfrac{1}{2}g_{AB,j}g^{ij}, \quad (6.12)$$

and

$$G^{A\nu}{}_{;\nu} \equiv 0. \quad (6.13)$$

It is a consequence of Eqs. (6.9) and the fact that

$$(u^{i}{}_{j} + u^{k}\Gamma^{i}_{kj})u^{j} = (v^{i}{}_{j} + v^{k}\Gamma^{i}_{kj})v^{j} = 0,$$

that

$$u_i(\epsilon_1 u^{\nu})_{;\nu} + v_i(\epsilon_2 v^{\nu})_{;\nu} - (R/2)\mu_{,i} = 0. \quad (6.14)$$

Therefore

$$2(\epsilon_1 u^{v})_{;v} - R\mu_{,u} = 0 \quad (6.15a)$$

and

$$2(\epsilon_2 u^{v})_{;v} - R\mu_{,v} = 0. \quad (6.15b)$$

These equations imply that the energy densities of the two null fluids are not conserved in region I unless $R = 0$, i.e., $C = R_{uv}$ vanishes. Equations (6.15) describe the interaction between the two fluids. They may be rewritten as

$$e^{-\mu}(e^{\mu}\epsilon_1\eta^{ij}u_j)_{,i} + 2C\mu_{,u} = 0 \quad (6.16a)$$

and

$$e^{-\mu}(e^{\mu}\epsilon_2\eta^{ij}v_j)_{,i} + 2C\mu_{,v} = 0. \quad (6.16b)$$

These equations in turn may be written as

$$(\epsilon_1(\mu_{,u})^{-1})_{,v} + C = 0 \quad (6.17a)$$

and

$$(\epsilon_2(\mu_{,v})^{-1})_{,u} + C = 0 \quad (6.17b)$$

as a consequence of Eq. (3.13).

In view of Eqs. (6.7), Eqs. (6.17) may be written as Eqs. (3.21). Since

$$((\mu_{,u})^{-1}R_{uu})_{,v} - ((\mu_{,v})^{-1}R_{vv})_{,u} = 0$$

there exists a function Ω such that Eqs. (3.18) and (3.20) hold.

As has been pointed out above, when $R_{uv} = C = 0$, the two null dust clouds described by the energy momentum tensor $T_{\mu\nu}$ via Eqs. (6.9) do not interact. This is the case described in Ref. 2. When

$$C^2 = \epsilon_1\epsilon_2, \quad (6.18)$$

it follows from Eq. (6.6) that

$$-R_{ij} = V_i V_j = V_\mu V_\nu \delta^{\mu}_{i}\delta^{\nu}_{j}, \quad (6.19)$$

where

$$V_i = (\epsilon_1)^{1/2}u_{,i} + (\epsilon_2)^{1/2}v_{,i}. \quad (6.20)$$

Equations (3.13) and (3.21) imply that the vector field V_μ is the gradient of a scalar that is the energy momentum tensor in region I that is due to a scalar field or equivalently that of a perfect fluid with pressure equal to energy density. This is the case described in Ref. 1.

VII. THE TWO FLUID INTERPRETATION

The energy momentum tensor $T_{\mu\nu}$ determined by Eqs. (6.9) and (6.10) may be expressed as

$$\kappa T_{\mu\nu} = -G_{\mu\nu} = (w+p)W_\mu W_\nu - pg_{\mu\nu} + nu_{,\mu}u_{,\nu}, \quad (7.1)$$

that is, as the sum of that of a perfect fluid and a null fluid. It may be shown that

$$W_\mu = W_u u_{,\mu} + W_v v_{,\nu}$$

and that as a consequence of $R_{AB} = 0$, one has

$$(w+p)W_\mu W^\mu - 2p = 0.$$

Hence if

$$W_\mu W^\mu = 0,$$

then

$$p = 0$$

and $T_{\mu\nu}$ is the sum of the energy momentum tensors of two null dusts.

If

$$W_\mu W^u = 1,$$

then

$$p = w,$$

and the first two terms of the right-hand side of Eq. (7.1) form the energy-momentum tensor of a perfect fluid with energy density equal to the pressure. It follows from the equating of $G_{\mu\nu}$ in Eqs. (6.9c) and (7.1) that

$$2pW_u^2 + n = \epsilon_1,$$

$$2pW_v^2 = \epsilon_2,$$

and

$$2pW_u W_v = C.$$

Note that when $n = 0$,

$$C^2 = \epsilon_1\epsilon_2.$$

That is, the medium in region I is that treated in Ref. 1.

[1] S. Chandrasekhar and B. C. Xanthopolous, Proc. R. Soc. London Ser. A **402**, 37 (1985).
[2] S. Chandrasekhar and B. C. Xanthopolous, Proc. R. Soc. London Ser. A **403**, 189 (1986).
[3] A. H. Taub, J. Math. Phys. **29**, 690 (1988).
[4] A. H. Taub, J. Math. Phys. **21**, 1423 (1980).
[5] S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London Ser. A **396**, 55 (1984); **398**, 429 (1985).
[6] K. Kahn and R. Penrose, Nature **229**, 185 (1971).
[7] Y. Nutku and M. Halil, Phys. Rev. Lett. **39**, 1379 (1977).
[8] V. Ferrari and J. Ibañez, Gen. Relativ. Gravit. **19**, 405 (1987).
[9] V. Belinskii and V. Zakharov, Sov. Phys. JETP **48**, 985 (1978); **50**, 1 (1979).
[10] P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[11] R. Tabensky and A. H. Taub, Commun. Math. Phys. **53**, 472 (1951).
[12] P. Szekerres, Nature **228**, 1183 (1970).

The bi-Hamiltonian formulation of the Landau–Lifshiftz equation

E. Barouch, A. S. Fokas,[a] and V. G. Papageorgiou
Department of Mathematics and Computer Science, Clarkson University, Potsdam, New York 13676

(Received 18 May 1988; accepted for publication 3 August 1988)

The Landau–Lifshitz (LL) equation is a universal model for integrable magnetic systems. It contains the sine–Gordon (SG), nonlinear Schrödinger (NLS), and the Heisenberg model (HM) equations as particular or limiting cases. It is well known that the NLS, SG, and HM equations possess *recursion operators*. A recursion operator of an equation in Hamiltonian form generates (a) a hierarchy of integrable equations, and (b) a second Hamiltonian operator and more generally a hierarchy of Poisson structures. Here the recursion operator of the LL equation is obtained algorithmically, and hence its bi-Hamiltonian formulation is established.

I. INTRODUCTION

The Landau–Lifshitz (LL) equation describes nonlinear spin waves in an anisotropic ferromagnet. It is given by

$$\mathbf{S}_t = \mathbf{S}\wedge\mathbf{S}_{xx} + \mathbf{S}\wedge J\mathbf{S}, \qquad (1.1\text{a})$$

where

$$J=\operatorname{diag}(J_1,J_2,J_3),\quad \mathbf{S}=(S_1,S_2,S_3),\quad |\mathbf{S}|^2=\mathbf{S}\cdot\mathbf{S}=1. \qquad (1.1\text{b})$$

In the above the diagonal matrix J is a measure of the anisotropy, $J_1<J_2<J_3$, $\mathbf{S}$ is an x- and t-dependent vector of unit norm in $\mathbb{R}^3$, and $\cdot$ and $\wedge$ denote the usual scalar and vector products.

The partially anisotropic Heisenberg model (HM) and the HM equations correspond to $J_1=J_2<J_3$ and $J_1=J_2=J_3$, respectively. It was pointed out in Ref. 1 that the LL is the most general magnet model admitting an r-matrix formulation. Furthermore, both the sine–Gordon (SG) and nonlinear Schrödinger (NLS) equations are limiting cases of the LL equation. The analysis of the LL is technically more complicated than that of HM, SG, and NLS. This is because the isospectral linear eigenvalue problem associated with LL involves elliptic functions,[2]

$$U_x(x,t,\lambda) = -i\Big(\sum_{j=1}^{3} S_j(x,t)\,W_j(\lambda)\sigma_j\Big)U(x,t,\lambda) \doteq -iLU, \qquad (1.2\text{a})$$

where the Pauli spin matrices are given by

$$\sigma_1\doteq\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2\doteq\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3\doteq\begin{pmatrix}1&0\\0&-1\end{pmatrix}, \qquad (1.2\text{b})$$

and

$$W_1(\lambda)\doteq\rho[1/\operatorname{sn}(\lambda,k)],\quad W_2(\lambda)\doteq\rho[\operatorname{dn}(\lambda,k)/\operatorname{sn}(\lambda,k)],\quad W_3(\lambda)\doteq\rho[\operatorname{cn}(\lambda,k)/\operatorname{sn}(\lambda,k)], \qquad (1.3\text{a})$$

with

$$k\doteq((J_2-J_1)/(J_3-J_1))^{1/2},\quad 0<k<1,\quad \rho\doteq\sqrt{J_3-J_1}. \qquad (1.3\text{b})$$

In the isospectral problems associated with the HM, SG, and NLS equations the spectral parameter λ ranges over the complex plane $\mathbb{C}$; however, the natural range of λ in (1.2) is an elliptic curve: The torus $E=\mathbb{C}/\Gamma$, where Γ is the lattice generated by $4K$ and $4iK'$, where K and K' are the complete elliptic integrals of moduli k and $k'=\sqrt{1-k^2}$.

The Lax pair of the LL was found by Sklyanin[2] (see also Ref. 3), who also obtained the action-angle variables (for rapidly approaching a fixed unit vector boundary conditions) by introducing the notion of the classical r matrix. The initial value problem for similar data was studied by Mikhailov[4] (see also Ref. 5) using a Riemann–Hilbert problem on an elliptic curve. A general description of finite-gap solutions was given in Ref. 6 and explicit formulas were obtained in Refs. 7 and 8 in terms of Prym theta functions.

Algebraic properties of the LL were studied in Ref. 7 where also the next member of its hierarchy was explicitly given. Fuchssteiner[9] presented hierarchies of time-independent symmetries, time-dependent symmetries, and conserved quantities using the notion of a master symmetry introduced in Ref. 10. However, the recursion operator could not be found and hence its bi-Hamiltonian formulation could not be established. This is a serious disadvantage since the bi-Hamiltonian property appears to be a fundamental property underlying integrability.[11–15] Indeed, the bi-Hamiltonian formulation of NLS and SG is well established. Also the recursion operator and the hierarchy of Hamiltonian operators associated with the HM have been found in Ref. 16 using the gauge equivalence of the HM to the NLS.[17,18]

There exist various approaches in the literature for constructing recursion operators.[19] We favor the one that uses the associated isospectral problem. Indeed, this approach has also been successful for obtaining recursion operators in lattices[20] and in multidimensions.[21] Also, it has the advantage to yield hereditary recursion operators.[22] In Sec. II we illustrate our method by deriving the recursion operator of the HM equation; this operator coincides with the one given in Ref. 16. In Sec. III we derive the recursion operator of the LL equation and establish its bi-Hamiltonian factorization.[23]

The method of deriving the recursion operator from an isospectral problem makes crucial use of a certain expansion in powers of the spectral parameter λ. The main difficulty we encountered in applying this method to LL stemmed from

[a] Present address: Department of Mathematics, Stanford University, Stanford, California 94304.

 0022-2488/88/122628-06$02.50

the fact that λ moves on an elliptic curve. This problem was bypassed by using the parametrization

$$\nu \doteq \tfrac{1}{2} W_1 W_2 W_3, \quad \mu \doteq W_3^2, \tag{1.4}$$

$$\nu^2 = \tfrac{1}{4}\mu(\mu+\alpha)(\mu+\beta); \quad \alpha \doteq -\tfrac{1}{4}(J_1 - J_3), \quad \beta \doteq -\tfrac{1}{4}(J_2 - J_3). \tag{1.5}$$

This paper is organized as follows. In Sec. I A we review the basic notions of symmetries, gradients of conserved quantities, recursion operators, and Hamiltonian operators. In Sec. I B we establish the connection between these results and those of Fuchssteiner[9] by showing how the recursion operator derived in this paper algorithmically implies the master symmetry found in Ref. 9. In Secs. II and III we derive the factorizable recursion operators of the HM and LL equations, respectively.

A. Basic notions

We consider the evolution equation (1.1) in the abstract form

$$\mathbf{S}_t = K(\mathbf{S}). \tag{1.6}$$

Let E denote the vector space of C^∞ *maps* from $\mathbb{R}$ into $\mathbb{R}^3$ and let TE denote the space of suitable C^∞ vector fields on E. The manifold on which the flow (1.6) takes place is denoted by M and the space of its smooth vector fields by TM. Clearly, M is a subspace of E such that $\mathbf{S}\in E$ satisfies $\mathbf{S}\cdot\mathbf{S} = 1$. Similarly TM is a subspace of TE such that $V(\mathbf{S})\in T_{\mathbf{S}}E$ satisfies $V(\mathbf{S})\cdot\mathbf{S} = 0$, i.e., $V(\mathbf{S}(x))$ belongs to the tangent plane of the unit sphere at $\mathbf{S}(x)$.

In TM we define the usual Lie bracket by

$$[K,G]_L \doteq K'[G] - G'[K], \tag{1.7a}$$

where $K'[G]$ denotes the Fréchet derivative of K in the direction G, i.e.,

$$K'[G] \doteq \frac{\partial}{\partial\epsilon} K(S+\epsilon G)|_{\epsilon=0}. \tag{1.7b}$$

Let T^*M be the dual of TM with respect to the bilinear form

$$(\gamma,\sigma) \doteq \int_{\mathbb{R}} dx\, \gamma\cdot\sigma, \quad \gamma\in T^*M, \quad \sigma\in TM. \tag{1.8}$$

Let I: $M\to\mathbb{R}$ be a functional; then its gradient, ∇I, is defined by

$$I'[v] \doteq (\nabla I, v), \quad v\in TM. \tag{1.9}$$

It is well known that a function f is a gradient iff $f' = (f')^+$, where the adjoint L^+ of an operator L is defined by $(L^+\gamma,\sigma) = (\gamma, L\sigma)$. In order to make the gradient unique we consider its projection onto the tangent plane of the unit sphere in $\mathbb{R}^3$ at the point $\mathbf{S}(x)$; i.e., $\gamma\cdot\mathbf{S} = 0$.

The conserved quantities of the LL equation take the form

$$I = \int_{\mathbb{R}} dx(\Gamma(\mathbf{S}) - \Gamma(e)), \quad e \doteq (0,0,1)^+, \tag{1.10}$$

where we have assumed that $\mathbf{S}\to e$ as $x\to\pm\infty$. As an example consider

$$H_0 = \int_{\mathbb{R}} dx(\Gamma_0(\mathbf{S}) - \Gamma_0(e)), \quad \Gamma_0 \doteq \frac{1}{2}(\mathbf{S}\cdot J\mathbf{S} - \mathbf{S}_x\cdot\mathbf{S}_x). \tag{1.11a}$$

Then

$$H_0'[\mathbf{v}] = \int_{\mathbb{R}} dx(\mathbf{v}\cdot J\mathbf{S} - \mathbf{v}_x\cdot\mathbf{S}_x) = \int_{\mathbb{R}} dx\, \mathbf{v}\cdot(J\mathbf{S} + \mathbf{S}_{xx}).$$

Thus

$$\nabla H_0 = \pi(\mathbf{S}_{xx} + J\mathbf{S}), \quad \pi\mathbf{a} \doteq -\mathbf{S}\wedge(\mathbf{S}\wedge\mathbf{a}) = \mathbf{a} - (\mathbf{a}\cdot\mathbf{S})\mathbf{S}. \tag{1.11b}$$

(i) The hierarchy of the LL equation consists of all flows that commute with (1.1); i.e., it consists of all time-independent symmetries σ. We recall that σ is a symmetry of (1.1) iff

$$\frac{\partial\sigma}{\partial t} + [\sigma,K]_L = 0, \quad \sigma\in TM. \tag{1.12}$$

(ii) Equation (1.6) is a Hamiltonian system iff it can be written in the form

$$\mathbf{S}_t = \Theta\nabla H, \tag{1.13a}$$

where Θ is a Hamiltonian operator, i.e., Θ is skew symmetric with respect to (1.8) and it satisfies, also, the Jacobi identity,

$$(\nabla I_1, \Theta'[\nabla I_2]\nabla I_3) + \text{cyclic permutations} = 0, \quad \nabla I_i\in T^*M, \quad i = 1,2,3, \tag{1.13b}$$

and H is a functional. The Hamiltonian operator Θ induces the following Poisson bracket:

$$\{I_1,I_2\} \doteq (\nabla I_1, \Theta\nabla I_2). \tag{1.14}$$

(iii) A functional I is a conserved quantity of (1.6) iff $I'[K] = 0$, or [cf. (1.13a)]

$$I'[K] = (\nabla I, \Theta\nabla H) = \{I,H\} = 0.$$

It turns out that it is more convenient to work with gradients of conserved quantities; these conserved gradients satisfy

$$\frac{\partial\gamma}{\partial t} + \gamma'[K] + (K')^+[\gamma] = 0, \quad \gamma \doteq \nabla I. \tag{1.15}$$

For Hamiltonian systems there is an isomorphism between the Lie commutator (1.7a) and the Poisson bracket (1.14),[10-12]

$$[\Theta\nabla I_1, \Theta\nabla I_2]_L = \Theta\nabla(\{I_1,I_2\}). \tag{1.16}$$

This isomorphism implies that, for a Hamiltonian system, symmetries and gradients of conserved quantities are related by

$$\sigma = \Theta\nabla I, \quad \sigma\in TM, \quad \nabla I\in T^*M. \tag{1.17}$$

It is well known that the LL equation is a Hamiltonian system. Indeed, it can be written in the form:

$$\mathbf{S}_t = \mathbf{S}\wedge\nabla H_0, \tag{1.18}$$

where ∇H_0 is defined by (1.11) and $\Theta = \mathbf{S}\wedge$ is a Hamiltonian operator (Θ is obviously skew symmetric and it is a straightforward exercise to show that it satisfies the Jacobi identity).

Fundamental role in the characterization of the algebraic properties of integrable evolution equations is played by hereditary (Nijenhuis) recursion operators.

If Φ is a hereditary (Nijenhuis) operator then

$$[\Phi^n K,\Phi^m K]_L=0,\quad ((\Phi^+)^n\nabla H,\Theta(\Phi^+)^m\nabla H)=0, \tag{1.19}$$

and $\Phi^n\Theta$ are Hamiltonian operators compatible with Θ, for all $n,m\in\mathbb{N}$. (Two Hamiltonian operators are compatible if their sum is a Hamiltonian operator.)

In Secs. II and III we derive hereditary recursion operators for HM and LL equations. Then $\Phi^n K$, $(\Phi^+)^n\nabla H_0$, $\Phi^n(\mathbf{S}\wedge\cdot)$ define hierarchies of commuting symmetries, conserved gradients in involution, and Hamiltonian operators, respectively.

B. Master symmetries

The general theory associated with master symmetries of evolution equations in one spatial and one temporal dimension is well established.[21,24,25] Here we only note that given a time-dependent symmetry σ of the form

$$\sigma=\sigma_0+t\sigma_1, \tag{1.20a}$$

and a recursion operator Φ, then

$$\tau=\Phi\sigma_0 \tag{1.20b}$$

is a master symmetry. Alternatively, if

$$\gamma=\gamma_0+t\gamma_1 \tag{1.21a}$$

is a time-dependent conserved gradient, and $\Psi=\Phi^+$, then

$$T=\Theta\Psi\gamma_0 \tag{1.21b}$$

is a master symmetry.

It turns out that

$$\tau=\mathbf{S}\wedge\Psi_{\rm LL}(x\mathbf{S}), \tag{1.22}$$

where $\Psi_{\rm LL}$ is the adjoint of the recursion operator of the LL [see Eq. (3.1)], is a master symmetry of the LL equation. This coincides with the one given by Fuchssteiner.[9]

II. THE HEISENBERG MODEL (HM)

The HM equation is given by

$$\mathbf{S}_t=\mathbf{S}\wedge\mathbf{S}_{xx},\quad \mathbf{S}\cdot\mathbf{S}=1. \tag{2.1}$$

Its associated isospectral eigenvalue problem is given by

$$U_x=\frac{i}{\lambda}\sum_{j=1}^{3}S_j\sigma_j U, \tag{2.2}$$

where λ is the spectral parameter and the Pauli matrices σ_j are defined in (1.2b).

Proposition 2.1: (a) The isospectral eigenvalue problem (2.2) yields the recursion operator $\Phi_{\rm HM}$ defined by

$$\Phi_{\rm HM}\doteq-\tfrac{1}{2}[\mathbf{S}\wedge D-\{D^{-1}(\mathbf{S}\wedge\mathbf{S}_x\cdot\cdot)\}\mathbf{S}_x]. \tag{2.3}$$

(b) The adjoint of $\Phi_{\rm HM}$ with respect to the bilinear form (1.8),

$$\Psi_{\rm HM}\doteq\Phi_{\rm HM}^+=-\tfrac{1}{2}(\mathbf{S}\wedge D-\{D^{-1}(\mathbf{S}\cdot D\cdot)\}\mathbf{S}\wedge\mathbf{S}_x) \tag{2.4}$$

satisfies

$$\mathbf{S}\wedge(\Psi_{\rm HM}\cdot)=\Phi_{\rm HM}(\mathbf{S}\wedge\cdot). \tag{2.5}$$

(c) The isospectral problem (2.2) is associated with the hierarchy of integrable evolution equations,

$$\mathbf{S}_t=\mathbf{S}\wedge\Psi_{\rm HM}^n(\mathbf{S}\wedge\mathbf{S}_x)=\Phi_{\rm HM}^n(-\mathbf{S}_x),\quad n=0,1,2,3,\ldots. \tag{2.6}$$

The HM equation corresponds to $n=1$.

(d) The hierarchy $\mathbf{S}\wedge\Psi_{\rm HM}^n$, $n=0,1,2,\ldots$, is a hierarchy of Hamiltonian operators. In particular the second Hamiltonian operator of the HM is given by $\Omega_{\rm HM}\doteq\mathbf{S}\wedge\Psi_{\rm HM}$; thus the HM is a bi-Hamiltonian system with compatible Hamiltonian operators $\mathbf{S}\wedge$ and $\Omega_{\rm HM}$.

Proof: Given (2.2) we look for compatible flows in the form

$$U_t=-i\sum_{l=1}^{3}V_l\sigma_l U. \tag{2.7}$$

The compatibility condition $U_{tx}=U_{xt}$ of Eqs. (2.2) and (2.7) implies

$$\mathbf{S}_t=\lambda\mathbf{V}_x-2\mathbf{S}\wedge\mathbf{V},\quad \mathbf{V}=(V_1,V_2,V_3). \tag{2.8}$$

We seek solutions $\mathbf{V}$ in the form

$$\mathbf{V}=\sum_{k=1}^{n}\mathbf{V}^{(k)}\lambda^{-k}. \tag{2.9}$$

Then (2.8) yields

$$\mathbf{S}_t=\mathbf{V}_x^{(1)}, \tag{2.10}$$

$$\mathbf{V}_x^{(j+1)}=2\mathbf{S}\wedge\mathbf{V}^{(j)},\quad j=1,\ldots,n-1, \tag{2.11}$$

$$\mathbf{S}\wedge\mathbf{V}^{(n)}=0. \tag{2.12}$$

Since $\mathbf{V}_x^{(j)}\cdot\mathbf{S}=0$, we define $\mathbf{v}^{(j)}$ as follows:

$$\mathbf{v}^{(j)}\doteq-\mathbf{S}\wedge\mathbf{V}_x^{(j)}, \tag{2.13}$$

with

$$\mathbf{v}^{(j)}\cdot\mathbf{S}=0. \tag{2.14}$$

Then Eqs. (2.10)–(2.12) are transformed into

$$\mathbf{S}\wedge\mathbf{S}_t=-\mathbf{v}^{(1)}, \tag{2.15}$$

$$\mathbf{v}^{(j+1)}=-2[\mathbf{S}\wedge(\mathbf{S}\wedge(D^{-1}\{\mathbf{S}\wedge\mathbf{v}^{(j)}\}))], \tag{2.16}$$

$$\mathbf{S}\wedge D^{-1}(\mathbf{S}\wedge\mathbf{v}^{(n)})=0. \tag{2.17}$$

We solve Eq. (2.16) for $\mathbf{v}^{(j)}$ as follows. Equation (2.16) is equivalent to

$$\mathbf{v}^{(j+1)}=2D^{-1}\{\mathbf{S}\wedge\mathbf{v}^{(j)}\}-2(\mathbf{S}\cdot D^{-1}\{\mathbf{S}\wedge\mathbf{v}^{(j)}\})\mathbf{S}.$$

Hence

$$\mathbf{v}_x^{(j+1)}=2\mathbf{S}\wedge\mathbf{v}^{(j)}-2(\mathbf{S}\cdot D^{-1}(\mathbf{S}\wedge\mathbf{v}^{(j)}))\mathbf{S}_x-2(\mathbf{S}\cdot D^{-1}\{\mathbf{S}\wedge\mathbf{v}^{(j)}\})_x\mathbf{S}. \tag{2.18}$$

From Eq. (2.18), taking $\mathbf{S}\wedge$ and $\mathbf{S}\cdot$ of both sides we obtain

$$\mathbf{S}\wedge\mathbf{v}_x^{(j+1)}=-2\mathbf{v}^{(j)}-2(\mathbf{S}\cdot D^{-1}\{\mathbf{S}\wedge\mathbf{v}^{(j)}\})\mathbf{S}\wedge\mathbf{S}_x \tag{2.19}$$

and

$$\mathbf{S}\cdot\mathbf{v}_x^{(j+1)}=-2(\mathbf{S}\cdot D^{-1}\{\mathbf{S}\wedge\mathbf{v}^{(j)}\})_x,$$

i.e.,

$$2\mathbf{S}\cdot D^{-1}\{\mathbf{S}\wedge\mathbf{v}^{(j)}\}=-D^{-1}(\mathbf{S}\cdot\mathbf{v}_x^{(j+1)}). \tag{2.20}$$

Substituting in (2.19), we get

$$\mathbf{v}^{(j)} = -\tfrac{1}{2}(\mathbf{S}\wedge\mathbf{v}_x^{(j+1)} - \{D^{-1}(\mathbf{S}\cdot\mathbf{v}_x^{(j+1)})\}\mathbf{S}\wedge\mathbf{S}_x), \tag{2.21}$$

i.e. [cf. (2.4)],

$$\mathbf{v}^{(j)} = \Psi\mathbf{v}^{(j+1)}.$$

So,

$$\mathbf{v}^{(1)} = \Psi^{n-1}\mathbf{v}^{(n)},$$

and solving (2.15) and (2.17) we get

$$\mathbf{S}_t = \mathbf{S}\wedge\Psi^{n-1}(\mathbf{S}\wedge\mathbf{S}_x). \tag{2.22}$$

In the Appendix, we show that $\mathbf{S}\wedge$ and Ω_{HM} are compatible Hamiltonian operators, thus establishing the bi-Hamiltonian structure of the HM.

Remarks 2.1: (i) Equation (1.15) is derived by differentiating $(\gamma,K)=0$ in the arbitrary direction $\mathbf{v}$, where $\mathbf{v}\cdot\mathbf{S}=0$. Thus one can extend the definition of a conserved gradient by allowing functions γ that are not of the form $\pi\tilde{\gamma}$, provided that

$$\mathbf{S}\wedge\left(\frac{\partial\tilde{\gamma}}{\partial t} + \tilde{\gamma}'[K] + (K')^+[\tilde{\gamma}]\right) = 0, \tag{2.23}$$

$$([\tilde{\gamma}' - (\tilde{\gamma}')^+]\mathbf{a},\mathbf{b}) = 0,\ \mathbf{a},\mathbf{b} \text{ orthogonal to } \mathbf{S}. \tag{2.24}$$

Indeed the starting γ of the HM hierarchy satisfies

$$\tilde{\gamma} \doteq \mathbf{S}\wedge\mathbf{S}_x,\quad \tilde{\gamma}'[K] + (K')^+[\tilde{\gamma}] = -\tfrac{3}{2}(\mathbf{S}_x\cdot\mathbf{S}_x)_x\mathbf{S}, \tag{2.25}$$

$$([\tilde{\gamma}' - (\tilde{\gamma}')^+]\mathbf{a},\mathbf{b}) = (\mathbf{S}_x\wedge\mathbf{a},\mathbf{b}) = 0. \tag{2.26}$$

(ii) $\Psi_{\text{HM}}(\mathbf{S}\wedge\mathbf{S}_x) = \pi\mathbf{S}_{xx} = \nabla H_0$, where

$$H_0 \doteq \int_{-\infty}^{\infty} dx(\Gamma_0(\mathbf{S}) - \Gamma_0(e)),\quad \Gamma_0 = -\frac{1}{2}\mathbf{S}_x\cdot\mathbf{S}_x. \tag{2.27}$$

(iii) $\gamma^{(1)} \doteq x\mathbf{S}\wedge\mathbf{S}_x - 2t\,\mathbf{S}_{xx}$ is a conserved gradient of the HM. Hence

$$\tau \doteq \Theta\Psi_{\text{HM}}(x\mathbf{S}\wedge\mathbf{S}_x) = x\mathbf{S}\wedge\nabla H_0 + \mathbf{S}\wedge\mathbf{S}_x \tag{2.28}$$

is a master symmetry of HM. This coincides with Eq. (12) of Ref. 9 if $J=0$.

(iv) It is shown in the Appendix that the operator $\Omega_{\text{HM}} \doteq \mathbf{S}\wedge\Psi_{\text{HM}}$ satisfies the Jacobi identity. Since $\mathbf{S}\cdot\mathbf{a}=0$, Ω_{HM} is equivalent to $\tilde{\Omega} = \tfrac{1}{2}(D + D\{\mathbf{S}D^{-1}(\mathbf{S}_x\cdot\cdot)\})$. However, in order to prove the Jacobi identity for $\tilde{\Omega}$ we have to take into account that $\tilde{\Omega}\mathbf{a}\cdot\mathbf{b} = \tilde{\Omega}\mathbf{b}\cdot\mathbf{c} = \tilde{\Omega}\mathbf{c}\cdot\mathbf{a} = 0$, which are Fréchet-derivative consequences of the equations $\mathbf{S}\cdot\mathbf{a} = \mathbf{S}\cdot\mathbf{b} = \mathbf{S}\cdot\mathbf{c} = 0$.

III. THE LANDAU–LIFSHITZ (LL) EQUATION

Proposition 3.1: (a) The isospectral eigenvalue problem (1.2) yields the recursion operator Φ_{LL} defined by

$$\begin{aligned}\Phi_{\text{LL}} \doteq \Phi_{\text{HM}}^2 &- \tfrac{1}{4}\pi((4A\,\mathbf{S})\wedge(\mathbf{S}\wedge\cdot)\\ &- (D^{-1}\{S\cdot 4A\,\mathbf{S}\wedge(\mathbf{S}\wedge\cdot)\})\mathbf{S}_x\\ &- (D^{-1}\{\mathbf{S}\cdot(\mathbf{S}\wedge\cdot)_x\})(4A\,\mathbf{S}\wedge\mathbf{S})).\end{aligned} \tag{3.1a}$$

(b) The adjoint of Φ_{LL} with respect to the bilinear form (1.8) is

$$\begin{aligned}\Psi_{\text{LL}} \doteq \Psi_{\text{HM}}^2 &+ \tfrac{1}{4}\mathbf{S}\wedge((4A\,\mathbf{S})\wedge\cdot - (D^{-1}\{\mathbf{S}\cdot 4A\,\mathbf{S}\wedge\cdot\})\mathbf{S}_x\\ &- (D^{-1}\{\mathbf{S}\cdot D\cdot\})4A\,\mathbf{S}\wedge\mathbf{S}),\end{aligned} \tag{3.1b}$$

and satisfies

$$\mathbf{S}\wedge(\Psi_{\text{LL}}\cdot) = \Phi_{\text{LL}}(\mathbf{S}\wedge\cdot) = \Omega_{\text{LL}}. \tag{3.1c}$$

(c) The associated hierarchy of integrable evolution equations is given by

$$\mathbf{S}_t = \mathbf{S}\wedge\Psi_{\text{LL}}^n(\alpha\mathbf{S}\wedge\mathbf{S}_x),\quad n=0,1,2,3,\ldots,\quad \alpha=\text{const}, \tag{3.2a}$$

$$\mathbf{S}_t = \mathbf{S}\wedge\Psi_{\text{LL}}^n(0),\quad n=0,1,2,3,\ldots. \tag{3.2b}$$

The LL equation corresponds to (3.2b), $n=1$. Note that in (3.2b) $D^{-1}(0)$ is understood as a constant.

(d) The hierarchy $\mathbf{S}\wedge\Psi_{\text{LL}}^n$, $n=0,1,2,\ldots$, is a hierarchy of Hamiltonian operators. In particular, the second Hamiltonian operator of the LL equation is given by $\Omega_{\text{LL}} \doteq \mathbf{S}\wedge\Psi_{\text{LL}}$, thus the LL is a bi-Hamiltonian system with compatible Hamiltonian operators $\mathbf{S}\wedge$ and Ω_{LL}.

Proof: Given (1.2), we seek compatible flows in the form

$$U_t = -i\left\{\sum_{j=1}^{3} W_j V_j \sigma_j\right\} U. \tag{3.3}$$

The compatibility condition $U_{tx} = U_{xt}$ of Eqs. (1.2) and (3.3) implies

$$\sum_{j=1}^{3} S_{j,t} W_j\sigma_j - \sum_{j=1}^{3} V_{j,x} W_j\sigma_j - i\left[\sum_{j=1}^{3} S_j W_j\sigma_j,\ \sum_{l=1}^{3} V_l W_l\sigma_l\right] = 0. \tag{3.4}$$

Equating coefficients of σ_j, for $j=1,2,3$, one obtains

$$S_{1,t} = (2W_2W_3/W_1)(S_3V_2 - S_2V_3) + V_{1,x}, \tag{3.5}$$

and cyclic permutations.

In terms of the parameters μ,ν [cf. (1.4) and (1.5)], we get

$$S_{1,t} = [\mu(\mu+\beta)/\nu](S_3V_2 - S_2V_3) + V_{1,x}, \tag{3.6a}$$

$$S_{2,t} = [(\mu+\alpha)\mu/\nu](S_1V_3 - S_3V_1) + V_{2,x}, \tag{3.6b}$$

$$S_{3,t} = [(\mu+\beta)(\mu+\alpha)/\nu](S_2V_1 - S_1V_2) + V_{3,x}. \tag{3.6c}$$

We seek solutions V_j, $j=1,2,3$, in the form

$$V_1 = \frac{\mu(\mu+\beta)}{\nu}\sum_{j=0}^{n}\mu^{n-j}a_1^{(j)} + \sum_{j=0}^{n}\mu^{n-j}b_1^{(j)}, \tag{3.7a}$$

$$V_2 = \frac{(\mu+\alpha)\mu}{\nu}\sum_{j=0}^{n}\mu^{n-j}a_2^{(j)} + \sum_{j=0}^{n}\mu^{n-j}b_2^{(j)}, \tag{3.7b}$$

$$V_3 = \frac{(\mu+\beta)(\mu+\alpha)}{\nu}\sum_{j=0}^{n}\mu^{n-j}a_3^{(j)} + \sum_{j=0}^{n}\mu^{n-j}b_3^{(j)}. \tag{3.7c}$$

Upon substitution of (3.7) in (3.6) one obtains

$$S_{1,t}=\frac{\mu(\mu+\beta)}{\nu}\sum_{j=0}^{n}\mu^{n-j}a_{1,x}^{(j)}+\sum_{j=0}^{n}\mu^{n-j}b_{1,x}^{(j)}-\frac{\mu(\mu+\beta)}{\nu}\Big[S_2\Big(\frac{(\mu+\alpha)(\mu+\beta)}{\nu}\sum_{j=0}^{n}\mu^{n-j}a_3^{(j)}+\sum_{j=0}^{n}\mu^{n-j}b_3^{(j)}\Big)$$
$$-S_3\Big(\frac{\mu(\mu+\alpha)}{\nu}\sum_{j=0}^{n}\mu^{n-j}a_2^{(j)}+\sum_{j=0}^{n}\mu^{n-j}b_2^{(j)}\Big)\Big], \quad (3.8)$$

i.e.,

$$S_{1,t}=\frac{\mu(\mu+\beta)}{\nu}\sum_{j=0}^{n}\mu^{n-j}(a_{1,x}^{(j)}-S_2b_3^{(j)}+S_3b_2^{(j)})+\sum_{j=0}^{n}\mu^{n-j}(b_{1,x}^{(j)}-4\beta S_2a_3^{(j)})-4\sum_{j=-1}^{n-1}\mu^{n-j}(S_2a_3^{(j+1)}-S_3a_2^{(j+1)}), \quad (3.9)$$

and similarly for the other two equations.

Equating coefficients of μ^j and $\nu^{-1}\mu^j$ independently, one obtains

$$\mathbf{S}\wedge\mathbf{a}^{(0)}=0, \quad (3.10)$$

$$\mathbf{S}\wedge\mathbf{b}^{(j)}=\mathbf{a}_x^{(j)}, \quad (3.11)$$

$$\mathbf{S}\wedge\mathbf{a}^{(j+1)}=\tfrac{1}{4}\{\mathbf{b}_x^{(j)}-(4A\,\mathbf{S})\wedge\mathbf{a}^{(j)}\}, \quad (3.12)$$

$$\mathbf{S}_t=\mathbf{b}_x^{(n)}-(4A\,\mathbf{S})\wedge\mathbf{a}^{(n)}. \quad (3.13)$$

We define

$$q^{(j)}=-\mathbf{S}\wedge\{b_x^{(j)}-(4A\,\mathbf{S})\wedge\mathbf{a}^{(j)}\}. \quad (3.14)$$

Then (3.12) yields

$$\tfrac{1}{4}\mathbf{q}^{(j)}=\mathbf{a}^{(j+1)}-(\mathbf{S}\cdot\mathbf{a}^{(j+1)})\mathbf{S}. \quad (3.15)$$

Since $\mathbf{a}_x^{(j+1)}\cdot\mathbf{S}=0$ [cf. (3.11)],

$$\mathbf{a}^{(j+1)}=\tfrac{1}{4}(\mathbf{q}^{(j)}-\{D^{-1}(\mathbf{S}\cdot\mathbf{q}_x^{(j)})\}\mathbf{S}). \quad (3.16)$$

Applying the operators $(4A\,\mathbf{S})\wedge$ and $D(\mathbf{S}\wedge)D$ on (3.16) we obtain

$$(4A\,\mathbf{S})\wedge\mathbf{a}^{(j+1)}=(A\,\mathbf{S})\wedge\mathbf{q}^{(j)}-\{D^{-1}(\mathbf{S}\cdot\mathbf{q}_x^{(j)})\}(A\,\mathbf{S})\wedge\mathbf{S}, \quad (3.17)$$

and

$$-\mathbf{b}_x^{(j+1)}+(\mathbf{S}\cdot\mathbf{b}^{(j+1)})_x\mathbf{S}+(\mathbf{S}\cdot\mathbf{b}^{(j+1)})\mathbf{S}_x=\tfrac{1}{4}D\{\mathbf{S}\wedge\mathbf{q}_x^{(j)}-[D^{-1}(\mathbf{S}\cdot\mathbf{q}_x^{(j)})]\mathbf{S}\wedge\mathbf{S}_x\}, \quad (3.18)$$

because of (3.11).

Taking $\mathbf{S}\cdot$ of (3.18), (3.12), and (3.17) we get

$$-(\mathbf{S}\cdot\mathbf{b}_x^{(j+1)})+(\mathbf{S}\cdot\mathbf{b}^{(j+1)})_x=\tfrac{1}{4}\mathbf{S}\cdot D\{\mathbf{S}\wedge\mathbf{q}_x^{(j)}\}, \quad (3.19)$$

$$\mathbf{S}\cdot\mathbf{b}_x^{(j+1)}=\mathbf{S}\cdot(4A\,\mathbf{S})\wedge\mathbf{a}^{(j+1)}, \quad (3.20)$$

and

$$\mathbf{S}\cdot(4A\,\mathbf{S})\wedge\mathbf{a}^{(j+1)}=\tfrac{1}{4}\mathbf{S}\cdot(4A\,\mathbf{S})\wedge\mathbf{q}^{(j)}. \quad (3.21)$$

Therefore

$$\mathbf{S}\cdot\mathbf{b}^{(j+1)}=\tfrac{1}{4}D^{-1}\{\mathbf{S}\cdot[D\{\mathbf{S}\wedge\mathbf{q}_x^{(j)}\}+(4A\,\mathbf{S})\wedge\mathbf{q}^{(j)}]\}. \quad (3.22)$$

From (3.14), (3.17), (3.18), and (3.22) we get [cf. (2.4) also]

$$\mathbf{q}^{(j+1)}=\Psi_{\rm HM}^2\mathbf{q}^{(j)}+\tfrac{1}{4}\mathbf{S}\wedge((4A\,\mathbf{S})\wedge\mathbf{q}^{(j)}-(D^{-1}\{\mathbf{S}\cdot 4A\,\mathbf{S}\wedge\mathbf{q}^{(j)}\})\mathbf{S}_x-(D^{-1}\{\mathbf{S}\cdot\mathbf{q}_x^{(j)}\})(4A\,\mathbf{S})\wedge\mathbf{S}), \quad (3.23)$$

therefore establishing (3.1b).

Remarks 3.1: (i) $\gamma_0=x\mathbf{S}$ is a conserved gradient for the LL equation not, however, in T^*M. It turns out that

$$\tau=\mathbf{S}\wedge\Psi_{\rm LL}(x\mathbf{S})=x(\mathbf{S}\wedge\mathbf{S}_{xx}+\mathbf{S}\wedge J\mathbf{S})+\mathbf{S}\wedge\mathbf{S}_x, \quad (3.24)$$

is a master symmetry of the LL equation.

(ii) In the isotropic limit $(A\to{\rm diag}(0,0,0))$, $\Phi_{\rm LL}\to\Phi_{\rm HM}^2$.

(iii) There exist several equivalent forms of the recursion operator $\Phi_{\rm LL}$ and of the second Hamiltonian operator $\Omega_{\rm LL}$. One may verify the Jacobi identity of these equivalent forms by using the approach of Remark 2.1 (iv).

ACKNOWLEDGMENTS

One of the authors (A.S.F.) is grateful to J. Keller and his group at the Department of Mathematics, for their hospitality during the author's sabbatical leave at Stanford University.

This work was supported in part by the Air Force Office of Scientific Research under Grant No. 87-0310 and the National Science Foundation under Grant No. DMS-8501325.

APPENDIX: SKEW SYMMETRY AND JACOBI IDENTITY FOR THE OPERATORS Θ AND $\Omega_{\rm HM}$

In this appendix, we prove that the operator $\Omega_{\rm HM}$ given by the formula

$$\Omega_{\rm HM}\mathbf{a}=\mathbf{S}\wedge(\Psi_{\rm HM}\mathbf{a})=\tfrac{1}{2}(\mathbf{a}_x-D\{\mathbf{S}D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\}) \quad (A1)$$

is a Hamiltonian operator compatible with $\Theta=\mathbf{S}\wedge$.

In the following "$\equiv$" will denote equality up to perfect derivatives.

(i) $\Omega_{\rm HM}$ is skew symmetric. Consider **a**,**b** in T^*M; then

$$2(\Omega_{\rm HM}\mathbf{a})\cdot\mathbf{b}=\mathbf{a}_x\cdot\mathbf{b}-\mathbf{b}\cdot D\{\mathbf{S}D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\}$$
$$\equiv-\mathbf{a}\cdot\mathbf{b}_x+(\mathbf{b}_x\cdot\mathbf{S})D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)$$
$$\equiv-\mathbf{a}\cdot\mathbf{b}_x-(\mathbf{S}\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)$$
$$=-\mathbf{a}\cdot\mathbf{b}_x+(\mathbf{a}\cdot\mathbf{S}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)$$
$$=-2\Omega_{\rm HM}\mathbf{b}\cdot\mathbf{a},$$

therefore

$$(\Omega_{\rm HM}\mathbf{a},\mathbf{b})=-(\mathbf{a},\Omega_{\rm HM}\mathbf{b}). \quad (A2)$$

(ii) $\Omega_{\rm HM}$ satisfies the Jacobi identity. Consider **a**,**b**,**c** in T^*M; then

$$4(\Omega'_{HM}[\Omega_{HM}\mathbf{b}]\mathbf{a})\cdot\mathbf{c}\equiv\{\mathbf{b}_x\cdot\mathbf{c}_x-(\mathbf{S}\cdot\mathbf{c}_x)(\mathbf{S}\cdot\mathbf{b}_x)-(\mathbf{S}_x\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)$$
$$-\{\mathbf{b}_x\cdot\mathbf{a}_x-(\mathbf{S}\cdot\mathbf{a}_x)(\mathbf{S}\cdot\mathbf{b}_x)-(\mathbf{S}_x\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{c}_x). \quad (A3)$$

Therefore,

$$\begin{aligned}&4(\Omega'_{HM}[\Omega_{HM}\mathbf{b}]\mathbf{a})\cdot\mathbf{c}+(\text{cyclic permutations of }\mathbf{a,b,c})\\&\equiv\{\mathbf{b}_x\cdot\mathbf{c}_x-(\mathbf{S}\cdot\mathbf{c}_x)(\mathbf{S}\cdot\mathbf{b}_x)-(\mathbf{S}_x\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\\&+\{-\mathbf{b}_x\cdot\mathbf{a}_x+(\mathbf{S}\cdot\mathbf{a}_x)(\mathbf{S}\cdot\mathbf{b}_x)+(\mathbf{S}_x\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)\\&+\{\mathbf{c}_x\cdot\mathbf{a}_x-(\mathbf{S}\cdot\mathbf{a}_x)(\mathbf{S}\cdot\mathbf{c}_x)-(\mathbf{S}_x\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\\&+\{-\mathbf{c}_x\cdot\mathbf{b}_x+(\mathbf{S}\cdot\mathbf{b}_x)(\mathbf{S}\cdot\mathbf{c}_x)+(\mathbf{S}_x\cdot\mathbf{b}_x)D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\\&+\{\mathbf{a}_x\cdot\mathbf{b}_x-(\mathbf{S}\cdot\mathbf{b}_x)(\mathbf{S}\cdot\mathbf{a}_x)-(\mathbf{S}_x\cdot\mathbf{b}_x)D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)\\&+\{-\mathbf{a}_x\cdot\mathbf{c}_x+(\mathbf{S}\cdot\mathbf{c}_x)(\mathbf{S}\cdot\mathbf{a}_x)+(\mathbf{S}_x\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\}D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\equiv0.\end{aligned}\quad (A4)$$

(iii) The Hamiltonian operators Ω_{HM} and Θ are compatible, i.e., their sum is a Hamiltonian operator. Since Ω_{HM} and Θ are Hamiltonian operators, it is sufficient to prove that

$$(\{\Omega'_{HM}[\Theta\mathbf{b}]\mathbf{a}+\Theta'[\Omega_{HM}\mathbf{b}]\mathbf{a}\},\mathbf{c})+\text{cyclic permutations}=0, \quad (A5)$$

for any **a**,**b**,**c** in T^*M.

Indeed

$$\begin{aligned}&-2(\Omega'_{HM}[\Theta\mathbf{b}]\mathbf{a}+\Theta'[\Omega_{HM}\mathbf{b}]\mathbf{a})\cdot\mathbf{c}\\&=(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{c})(\mathbf{S}\cdot\mathbf{a}_x)+[(\mathbf{S}\wedge\mathbf{b})_x\cdot\mathbf{c}](\mathbf{S}\cdot\mathbf{a}_x)+(\mathbf{c}\cdot\mathbf{S}_x)D^{-1}(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{a}_x)-\mathbf{b}_x\wedge\mathbf{a}\cdot\mathbf{c}+(\mathbf{S}\wedge\mathbf{a}\cdot\mathbf{c})\mathbf{S}\cdot\mathbf{b}_x\\&\equiv-[(\mathbf{S}\wedge\mathbf{b})_x\cdot\mathbf{c}]D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)-(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)+[(\mathbf{S}\wedge\mathbf{b})_x\cdot\mathbf{c}]D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\\&+D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{a}_x)-(\mathbf{b}_x\wedge\mathbf{a}\cdot\mathbf{c})-[(\mathbf{S}\wedge\mathbf{a})_x\cdot\mathbf{c}]D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)-(\mathbf{S}\wedge\mathbf{a}\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\\&\equiv-(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)+(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)-(\mathbf{b}_x\wedge\mathbf{a}\cdot\mathbf{c})-(\mathbf{S}\wedge\mathbf{a}_x\cdot\mathbf{c})D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)-\mathbf{S}\wedge\mathbf{a}\cdot\mathbf{c}_xD^{-1}(\mathbf{S}\cdot\mathbf{b}_x).\end{aligned}\quad (A6)$$

So

$$\begin{aligned}&2(\Omega'_{HM}[\Theta\mathbf{b}]\mathbf{a}+\Theta'[\Omega\mathbf{b}]\mathbf{a})\cdot\mathbf{c}+\text{cyclic permutations of }\mathbf{a,b,c}\\&\equiv\mathbf{b}_x\wedge\mathbf{a}\cdot\mathbf{c}+(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)-(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)+(\mathbf{S}\wedge\mathbf{a}_x\cdot\mathbf{c})D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)+(\mathbf{S}\wedge\mathbf{a}\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)\\&+\mathbf{c}_x\wedge\mathbf{b}\cdot\mathbf{a}+(\mathbf{S}\wedge\mathbf{c}\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)-(\mathbf{S}\wedge\mathbf{c}\cdot\mathbf{b}_x)D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)+(\mathbf{S}\wedge\mathbf{b}_x\cdot\mathbf{a})D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)+(\mathbf{S}\wedge\mathbf{b}\cdot\mathbf{a}_x)D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)\\&+\mathbf{a}_x\wedge\mathbf{c}\cdot\mathbf{b}+(\mathbf{S}\wedge\mathbf{a}\cdot\mathbf{b}_x)D^{-1}(\mathbf{S}\cdot\mathbf{c}_x)-(\mathbf{S}\wedge\mathbf{a}\cdot\mathbf{c}_x)D^{-1}(\mathbf{S}\cdot\mathbf{b}_x)+(\mathbf{S}\wedge\mathbf{c}_x\cdot\mathbf{b})D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)+(\mathbf{S}\wedge\mathbf{c}\cdot\mathbf{b}_x)D^{-1}(\mathbf{S}\cdot\mathbf{a}_x)\\&=(\mathbf{b}\wedge\mathbf{a}\cdot\mathbf{c})_x\equiv0.\end{aligned}$$

[1] L. D. Fadeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, Berlin, 1987).

[2] E. K. Sklyanin, "On complete integrability of the Landau–Lifshitz equation," Preprint LOMI, E-3-79, Leningrad, 1979.

[3] A. E. Borovik and V. N. Robuk, "Linear pseudodifferentials and conservation laws for the Landau–Lifshitz equation describing the nonlinear dynamics of a ferromagnet with uniaxial anisotropy," Theor. Math. Phys. **46**, 242 (1981).

[4] A. V. Mikhailov, "The Landau–Lifshitz equation and the Riemann boundary problem on a torus," Phys. Lett. A **92**, 51 (1982).

[5] Yu. L. Rodin, "The Riemann boundary problem on Riemann surfaces and the inverse scattering problem for the Landau–Lifshitz equation," Physica D **11**, 90 (1984).

[6] I. V. Cherednik, "Integrable differential equations and coverings of elliptic curves," Math USSR–Izv. **22**, 357 (1984).

[7] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, "Landau–Lifshitz equation: Solitons; quasi-periodic solutions and infinitely dimensional Lie algebras," J. Phys. A **16**, 221 (1983).

[8] A. I. Bobenko, "Real algebraic-geometric solutions of the Landau–Lifshitz equation in terms of Prym theta-functions," Funct. Anal. Appl. **19**, 6 (1985) (Russian).

[9] B. Fuchssteiner, "On the hierarchy of the Landau–Lifshitz equation," Physica D **13**, 387 (1984); E. Barouch and B. Fuchssteiner, "Master symmetries and similarity equations of the *XYh* model," Stud. Appl. Math. **73**, 221 (1985).

[10] A. S. Fokas and B. Fuchssteiner, Phys. Lett. A **86**, 341 (1983).

[11] F. Magri, J. Math. Phys. **18**, 1405 (1977); *A Geometrical Approach to the Nonlinear Solvable Equations*, in *Lecture Notes in Physics*, Vol. 120, edited by M. Boiti, F. Pempinelli, and G. Soliani (Springer, Berlin, 1980).

[12] A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 249 (1980); B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).

[13] I. M. Gel'fand and I. Y. Dorfman, Funct. Anal. Appl. **13**, 248 (1979); **14**, 71 (1970).

[14] P. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).

[15] I. Gel'fand (private communication).

[16] V. S. Gerdjikov and A. B. Yanovski, Phys. Lett. A **103**, 232 (1984); Commun. Math. Phys. **103**, 549 (1986).

[17] M. Lakshmanan, Phys. Lett. A **61**, 53 (1977).

[18] V. E. Zakharov and Y. A. Takhtajan, Theor. Math. Phys. **38**, 17 (1974).

[19] B. G. Konopelchenko, *Nonlinear Integrable Equations, Lecture Notes in Physics*, Vol. 270 (Springer, Berlin, 1987).

[20] E. Barouch, A. S. Fokas, and V. G. Papageorgiou, RIMS Publ. **650**, 179 (1988).

[21] A. S. Fokas and P. M. Santini, Stud. Appl. Math. **75**, 179 (1986); P. M. Santini and A. S. Fokas, Commun. Math. Phys. **115**, 375 (1988); A. S. Fokas and P. M. Santini, *ibid.* **116**, 449 (1988).

[22] A. S. Fokas and R. L. Anderson, J. Math. Phys. **23**, 1066 (1982).

[23] V. Papageorgiou, Ph.D. thesis, Clarkson University, 1988.

[24] W. Oevel, "A geometrical approach to integrable systems admitting scaling symmetries," to appear.

[25] A. S. Fokas, "Symmetries and integrability," Stud. Appl. Math. **77**, 253 (1987).

Algebraic construction of partition functions for $c<1$ minimal conformal field theories

P. Suranyi[a)]
Blackett Laboratory, Imperial College, London SW7 2BZ, United Kingdom

(Received 3 May 1988; accepted for publication 3 August 1988)

Partition functions with periodic and twisted boundary conditions are constructed for $c<1$ minimal conformal field theories using the modular transformation properties of the characters of the Virasoro algebra alone. The construction helps to clarify the connection between twisted partition functions of $c<1$ and periodic Gaussian conformal field theories.

I. INTRODUCTION

Following the work of Cardy,[1] Cappelli *et al.*,[2] and Gepner[3] we were able to determine the operator content of $c<1$ minimal conformal field theories (CFT's) using modular invariance and a relation between modular properties of the characters of the Virasoro algebra and those of affine algebras. The classification of modular invariant bilinear forms of affine characters then allows one to find all modular invariant partition functions in which the coefficients of bilinear forms of characters of the Virasoro algebra are integer. The task of finding all the partition functions with positive coefficients,

$$Z=\sum N_{\lambda,\bar\lambda}\chi_\lambda(\tau)\chi_{\bar\lambda}(\tau), \tag{1}$$

where $N_{\lambda,\bar\lambda}\in Z^+$ and τ is the modular ratio, has been performed for up to $m'=100$. This m' is defined through the value of the central charge c,

$$c=1-6(m-m')^2/mm', \tag{2}$$

where m and m' are relatively prime integers $[(m,m')=1]$.

Besides two infinite series of solutions, the diagonal series for $m'=4,5,6,\ldots$ and the so-called AD series with $m'=6,8,10,\ldots$, they found exceptional solutions with $m'=12$, 18, and 30. One of the advantages of the group theoretic construction of Refs. 2 and 3 is that it allows finding Z_n symmetries present in these models; the Z_n groups correspond to symmetries of Dynkin diagrams of groups associated with individual solutions.

Nevertheless, it is clear that the construction of a complete set of modular invariants in a given CFT, with known character functions of known modular properties, should be possible without making use of their relation to the characters of the affine algebra. The purpose of the present paper is to demonstrate one possible construction of that kind of purely algebriac nature. Such a general construction may become useful for other CFT's, which cannot be readily related to affine theories.

There are some other advantages of the algebraic method discussed in the paper. Among others, partition functions with twisted boundary conditions, when they are appropriate, are obtained simultaneously. Such partition functions were obtained previously by Cardy[4] and Zuber.[5] Furthermore, it was pointed out by Di Francesco *et al.*[6] that known positive definite partition functions with periodic boundary conditions can be constructed from combinations of Gaussian (Coulombic) partition functions. These Gaussian forms and their extension to twisted boundary conditions[7] can be obtained directly from our algebriac construction as well.

In Sec. II, we introduce some definitions and set up the algebraic program of constructing all partition functions.In Sec. III, we deal with the actual construction, valid for all possible Z_n groups and all possible twists. Partition functions for periodic boundary conditions are also discussed. In Sec. IV we determine which of the Z_n symmetries can be realized in a given CFT. In Sec. V, a summary of our results is given. In the two appendices, algebraic problems arising in the process of the construction are dealt with.

II. MODULAR PROPERTIES OF PARTITION FUNCTIONS WITH TWISTED BOUNDARY CONDITIONS

For the sake of establishing notation we discuss briefly the modular properties of partition functions with twisted boundary conditions. Operators T and S acting on characters $\chi_\lambda(\tau)$, where λ is related to the dimension of the primary field $\Delta=(\lambda^2-1)/2N$ and $N=2mm'$, are defined as follows[2]:

$$T\chi_\lambda(\tau)=\chi_\lambda(\tau+1), \tag{3}$$

$$S\chi_\lambda(\tau)=\chi_\lambda(-1/\tau). \tag{4}$$

The symmetry properties of the characters are

$$\chi_\lambda(\tau)=\chi_{-\lambda}(\tau)=\chi_{\lambda+N}(\tau)=-\chi_{\lambda_0\lambda}(\tau), \tag{5}$$

where λ_0 satisfies

$$\lambda_0\lambda\equiv\lambda_0(rm'-sm)=rm'+sm \mod 2N. \tag{6}$$

Bilinear combinations of character functions transform under the following representation of the modular group[2]:

$$S\chi_\lambda(\tau)\chi_{\bar\lambda}(\tau)=\frac{1}{N}\sum e^{2\pi i(\lambda\lambda'+\bar\lambda\bar\lambda')/2N}\chi_{\lambda'}(\tau)\chi_{\bar\lambda'}(\tau), \tag{7}$$

$$T\chi_\lambda(\tau)\chi_{\bar\lambda}(\tau)=e^{2\pi i(\lambda^2-\bar\lambda^2)/2N}\chi_\lambda(\tau)\chi_{\bar\lambda}(\tau). \tag{8}$$

The range of summation in Eq. (7) is $0\leqslant\lambda,\bar\lambda\leqslant N-1$.

If an operator $\widetilde{Q}$, having eigenvalues $Q=0,1,2,\ldots,n-1$, with periodicity n, exists and is conserved, $[L_0,\widetilde{Q}]=[\bar L_0,\widetilde{Q}]=0$, then partition functions with twisted boundary conditions can be defined on a complex torus,[4,5] using the operator

[a)] Permanent address: Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221-0011.

 0022-2488/88/122634-07$02.50

$$\Sigma^k = \exp\{2\pi i \widetilde{Q}k/n\}, \quad k = 0,1,2,...,n-1. \tag{9}$$

Partition functions with twists in two different complex directions satisfy the following modular transformation rules[8]:

$$T\widetilde{Z}^{k_1,k_2}(\tau) = \widetilde{Z}^{k_1+k_2,k_2}(\tau), \tag{10}$$

$$S\widetilde{Z}^{k_1,k_2}(\tau) = \widetilde{Z}^{-k_2,k_1}(\tau). \tag{11}$$

The existence of a charge conjugation operator C, satisfying $[L_0,C] = [\overline{L}_0,C] = 0$ and $\widetilde{Q}C + C\widetilde{Q} = 0$, implies that

$$\widetilde{Z}^{k_1,k_2}(\tau) = \widetilde{Z}^{-k_1,-k_2}(\tau). \tag{12}$$

The partition function Z can be expanded in charge states, as follows:

$$\widetilde{Z}^{k_1,k_2}(\tau) = \sum_{Q=0}^{n-1} \exp\left\{2\pi i \frac{Qk_1}{n}\right\} Z^{Q,k_2}(\tau), \tag{13}$$

where $Z^{Q,k}(\tau)$ is a *physical partition function* defined on a cylinder and as such it has a character expansion with integer coefficients

$$Z^{Q,k}_{\lambda,\bar\lambda}(\tau) = \sum N^{Q,k}_{\lambda,\bar\lambda}\chi_\lambda(\tau)\chi_{\bar\lambda}(\bar\tau), \tag{14}$$

where $N^{Q,k}_{\lambda,\bar\lambda}\in Z^+$.

The integer expansion coefficients satisfy the angular momentum constraint

$$2N(\Delta - \Delta') = \lambda^2 - \bar\lambda^2 = 2NQk/n \mod 2N. \tag{15}$$

Furthermore Eqs. (7) and (11) imply that they also satisfy the following consistency condition:

$$N^{Q,k}_{\lambda,\bar\lambda} = \frac{1}{nN} R \sum_{Q',k'} \sum_{\lambda',\bar\lambda'} \cos\left(\frac{Qk' + Q'k}{n} 2\pi\right) \times \exp\left\{2\pi i \frac{\bar\lambda\lambda' + \lambda\bar\lambda'}{N}\right\} N^{Q',k'}_{\lambda',\bar\lambda'}, \tag{16}$$

where the operator R imposes symmetry properties (5) on a function of λ and $\bar\lambda$. Equation (16) is just the generalization of Cardy's self-consistency equation[3] for twisted boundary conditions.

Let us define now the symmetric partition function $\overline{Z}$:

$$\overline{Z}^k(\tau) = \tfrac{1}{2}(\widetilde{Z}^{k,k}(\tau) + \widetilde{Z}^{k,-k}(\tau)). \tag{17}$$

The function $\overline{Z}$ is obviously invariant for the exchange of toroidal boundaries (S). The coefficients of its character expansion satisfy the equation

$$\overline{Z}^k_{\lambda,\bar\lambda} = S\overline{Z}^k_{\lambda,\bar\lambda} \equiv \frac{1}{N}\sum_{\lambda'\bar\lambda'} \exp\left\{2\pi i \frac{\bar\lambda'\lambda + \lambda'\bar\lambda}{N}\right\} \overline{Z}^k_{\lambda',\bar\lambda'}. \tag{18}$$

Also, using Eqs. (13) and (15) we obtain

$$\overline{Z}^k_{\lambda,\bar\lambda} = \cos\left(\pi\frac{\lambda^2 - \bar\lambda^2}{N}\right)\sum_Q N^{Q,k}_{\lambda,\bar\lambda} \equiv \sum_Q \cos\left(2\pi\frac{Qk}{n}\right) N^{Q,k}_{\lambda,\bar\lambda}. \tag{19}$$

In other words, the symmetric partition functions with twisted boundary conditions must be linear combinations of eigenvectors of S of eigenvalue 1, of the form

$$\Psi_{\lambda,\bar\lambda} = \cos(\pi(\lambda^2 - \bar\lambda^2)/N) N_{\lambda,\bar\lambda}, \tag{20}$$

where $N_{\lambda,\bar\lambda}\in Z^+$. After finding the appropriate symmetric partition functions, Eq. (19) allows one to decompose these coefficients into components $N^{Q,k}$.

For the sake of simplicity, from this point on we shall only consider Z_n groups with prime n. This is sufficient for $c<1$ theories. The generalization to nonprime values of n is straightforward, though tedious. It follows from (15) that, for prime n, $\lambda - \bar\lambda$ and $\lambda + \bar\lambda$ are even. Therefore the notation

$$x_1 = \frac{\lambda - \bar\lambda}{2}, \quad x_2 = \frac{\lambda + \bar\lambda}{2}, \quad M = \frac{N}{2} \tag{21}$$

will be introduced. Then we can rewrite (18) and (19) as

$$\overline{Z}^k_{x_1,x_2} = S\overline{Z}^k_{x_1,x_2} = \frac{1}{2M} R \sum_{x_1'=0}^{2M-1} \sum_{x_2'=0}^{M-1} \exp\left\{2\pi i \frac{x_1x_2' + x_2x_1'}{M}\right\} \overline{Z}^k_{x_1',x_2'}, \tag{18'}$$

$$\overline{Z}^k_{x_1,x_2} = \cos\left(2\pi\frac{x_1x_2}{M}\right)\sum_Q N^{Q,k}_{x_1,x_2}. \tag{19'}$$

The symmetries of the characters are reflected by the following symmetry properties of the coefficients N (the superscripts Q and k are suppressed):

$$N_{x_1,x_2} = N_{x_2,x_1} = N_{-x_1,-x_2} = N_{x_1+2M,x_2} = N_{x_1+M,x_2+M}. \tag{22}$$

In the first part of this discussion the antisymmetrization of the expressions in $\lambda\to\lambda_0\lambda$ and $\bar\lambda\to\lambda_0\bar\lambda$ will not be enforced, consequently the positiveness of the integer coefficients should not be enforced either.

III. CONSTRUCTION OF PARTITION FUNCTIONS

Operator S of Eq. (18) has M^2 eigenfunctions of nonzero eigenvalue. Labeling these eigenfunctions with $0 \leqslant a,b \leqslant M-1$, they can be written as

$$\Psi = \exp\left\{2\pi i\frac{ax_1}{M}\right\}\delta^M_{x_2,b} + i^s\exp\left\{2\pi i\frac{bx_1}{M}\right\}\delta^M_{x_2,-a} + i^{2s}\exp\left\{-2\pi i\frac{ax_1}{M}\right\}\delta^M_{x_2,-b} + i^{3s}\exp\left\{-2\pi i\frac{bx_1}{M}\right\}\delta^M_{x_2,a}, \tag{23}$$

where $s=0,1,2,3$ and δ^M is a periodic Kronecker delta of period M. These eigenvectors belong to eigenvalue i^s. We only need eigenvectors of eigenvalue $+1$ ($s=0$). Their number is $M^2/4+2$ and they automatically satisfy the symmetry requirement for the exchange $(x_1,x_2)\to(-x_1,-x_2)$,

$$\Psi_{ab} = \exp\left\{2\pi i\frac{ax_1}{M}\right\}\delta^M_{x_2,b} + \exp\left\{2\pi i\frac{bx_1}{M}\right\}\delta^M_{x_2,-a} + (x_1,x_2)\to(-x_1,-x_2). \tag{24}$$

A general eigenvector of S of eigenvalue $+1$ can now be written as

$$\Psi(x_1,x_2)=\sum_{a,b} c_{a,b}\Psi_{a,b}(x_1,x_2). \tag{25}$$

We call an eigenvector admissible if it is of the form (20),

$$\Psi(x_1,x_2)=\cos\left(2\pi\frac{x_1x_2}{M}\right)N(x_1,x_2), \tag{20'}$$

where $N(x_1,x_2)\in Z$ and it satisfies symmetry properties (22).

The most general admissible eigenvector of S, as is proven in Appendix A by considering constraints on coefficients $c_{a,b}$, has the form

$$\Psi(x_1,x_2)=\cos\left(2\pi\frac{x_1x_2}{M}\right)\sum_K N^K_{r_1,r_2}M^K_{r_1,r_2}(x_1,x_2), \tag{26}$$

where $K=1,2,\ldots,M$ runs over the divisors of M, N^K is an arbitrary integer for all r_1 and r_2, and "structure constants" $M^K(x_1,x_2)$ are also integers, having the form

$$M^K_{r_1,r_2}(x_1,x_2)\sim\sum_{t_1=0}^{2M/K-1}\sum_{t_2=0}^{K-1}(\delta^{2M}_{x_1,Kt_1+r_1}\delta^{M}_{x_2,t_2M/K+r_2} + (r_1,r_2)\to(-r_1,-r_2))+(x_1-x_2), \tag{27}$$

where $r_1=0,1,\ldots,K-1$, $r_2=0,1,\ldots,M/k-1$. The equivalence sign $\sim$ implies equality up to an overall integer multiplier. The set of admissible eigenfunctions (27) defined above is usually overcomplete.

Partition functions for periodic boundary conditions have only contributions corresponding to integer spins, implying $x_1x_2=0 \bmod M$, which is satisfied only if $r_1=r_2=0$; consequently the partition function (expanded in characters) should be a linear combination of the form (26) with $r_1=r_2=0$ antisymmetrized for the exchanges $\lambda\leftrightarrow\lambda_0\lambda$ and $\bar\lambda\leftrightarrow\lambda_0\bar\lambda$. Of course, such an antisymmetrization generates negative terms even in the fundamental domain $1\leqslant r\leqslant m-1$, $1\leqslant s\leqslant m'-1$ [see Eq. (6)] except for the diagonal invariant, $K=1$. The negative terms should be eliminated by taking appropriate linear combinations of terms of different values of K.

The known positive definite modular invariant partition functions[1-3] can be expressed by our invariants $M^K(x_1,x_2)$ as follows: The diagonal invariants[1] are

$$Z^{(1)}(x_1,x_2)\sim M^1(x_1,x_2), \tag{28}$$

while those of the second infinite series of Cappelli *et al.*[2] can be written as a combination of $K=1$ and $K=2$ terms,

$$Z^{(2)}(x_1,x_2)\sim M^1(x_1,x_2)-RM^2(x_1,x_2), \tag{29}$$

where R antisymmetrizes in $x_1+x_2=\lambda\to\lambda_0\lambda$ and $x_2-x_1=\bar\lambda\to\lambda_0\bar\lambda$. The exceptional solutions of Cappelli *et al.*[2] and of Gepner[3] for $m'=12$ and 18 (E^6,E^7) are given as

$$Z^{(6,7)}\sim M^1-RM^2-RM^3,$$

where the arguments x_1 and x_2 of M were dropped together with the subscripts $r_1=r_2=0$. Finally the exceptional model E^8, $m'=30$, is given by

$$Z^{(8)}\sim M^1-RM^2-RM^3-RM^5.$$

One can show (see Appendix B) that, for the case where K is a divisor of m' (not only of M),

$$RM^K=-RM^{Km}, \tag{30}$$

where again $M=m'm$. Consequently, in the fundamental domain one can write the previous (expanded in characters) partition functions as follows:

$$Z^{(1)}\sim M^1-M^m, \tag{31}$$

$$Z^{(2)}\sim M^1-M^m-M^2+M^{2m}, \tag{32}$$

$$Z^{(6,7)}\sim M^1-M^m-M^2+M^{2m}-M^3+M^{3m}, \tag{33}$$

$$Z^{(8)}\sim M^1-M^m-M^2+M^{2m}-M^3+M^{3m}-M^5+M^{5m}. \tag{34}$$

Using the explicit form of characters[9] one can easily write down the complete partition functions (after multiplication by bilinear combinations of characters) in terms of the Coulombic (periodic Gaussian) partition functions of Di Francesco *et al.*[6] defined as a function of the coupling parameter g,

$$Z_C(g)=\frac{1}{\eta(q)\eta(\bar q)}\sum_{k_1,k_2\in Z} q^{\Delta(k_1,k_2)}\,\bar q^{\bar\Delta(k_1,k_2)}, \tag{35}$$

where $q=e^{2\pi i\tau}$, the dimensions Δ and $\bar\Delta$ are given by

$$\left.\begin{matrix}\Delta\\ \bar\Delta\end{matrix}\right\}(k_1,k_2)=(k_1\pm gk_2)^2/(4g), \tag{36}$$

and $\eta(q)$ is the Dedekind function (including the conformal prefactor). The partition functions are

$$Z^{(1)}(\tau)\sim Z_C(m'm)-Z_C(m'm), \tag{37}$$

$$\begin{aligned} &Z^{(2)}(\tau)\sim Z^{(1)}(\tau)-Z_C(m'm/4)+Z_C(m'/4m),\\ &Z^{(6,7)}(\tau)\sim Z^{(2)}(\tau)-Z_C(m'm/9)+Z_C(m'/9m),\\ &Z^{(8)}(\tau)\sim Z^{(6)}(\tau)-Z_C(m'm/25)+Z_C(m'/25m).\end{aligned} \tag{38}$$

IV. IMPOSITION OF Z_n SYMMETRY

Suppose now that a minimal $c<1$ CFT has Z_n symmetry as well, where n is a prime number (obviously if the CFT had Z_n symmetry with n not prime, then it would also have Z_p symmetry, where p is any of the prime factors of n).

Equations (15) and (21) show that n has to be a prime factor of M. Write $M=K_1K_2n$ and use $K\to K_1n$ in Eq. (27). Then one can see from the condition $x_1x_2=Mt/n \bmod M$, where t is an integer that only invariants with $r_2=0$ and $r_1=K_1r$ can contribute, where $r=0,1,\ldots,n-1$. Writing $k_2\to nk_2+s$, we get the general form of admissible eigenfunctions of S, contributing to $\bar Z^k(x_1,x_2)$:

$$\Psi^{K_1}_r(x_1,x_2)=R\sum_{s=0}^{n-1}\sum_{k_1=0}^{2K_2-1}\sum_{k_2=0}^{K_1-1}(\delta^{2M}_{x_1,K_1(nk_1+r)}\times\delta^{M}_{x_2,K_2(nk_2+s)}+(r\to-r))\cos\left(2\pi\frac{rs}{n}\right). \tag{39}$$

It is tempting to identify r with k and Ψ_r with $\bar Z^k$, based on the similarity of the phase factors in Eqs. (19) and (39). Then, of course, (19) forces the identification $Q=s$ on us. In fact, the function

$$\tilde N^{Q,k}(x_1,x_2)=\sum_{k_1,k_2}\delta^{2M}_{x_1,K_1(nk_1+k)}\,\delta^{M}_{x_2,K_2(nk_2+Q)} \tag{40}$$

has correct modular transformation properties (16). This is, of course, not a surprise, since the extension of the periodic

Gaussian partition functions to twisted boundary conditions satisfying (16) is obtained[7] if the dimensions Δ and $\overline{\Delta}$ of (36) are substituted by

$$\left.\begin{matrix}\Delta^{Q,k}\\ \overline{\Delta}^{Q,k}\end{matrix}\right\} = \frac{\{k_1 n + Q \pm g(k_2 n + k)\}^2}{4ng}. \tag{41}$$

The substitutions $g \to gn$, $t_1 \to t_1 + n/k$, and $t_2 \to t_2 n + Q$ lead directly from expressions (27)–(29) to (40).

It is easy to see that identification (40) with the twisted partition function cannot be correct. Observe only that the ground state, $\lambda = \overline{\lambda} = 1$, corresponding to $x_1 = 0$, $x_2 = 1$ appears in the $Q = 1$, $k = 0$ and $Q = n - 1$, $k = 0$ sectors of the $K_2 = 1$ term (present in all of the modular invariant models), instead of the $Q = 0$, $k = 0$ sector. This is, of course, forbidden, since the energy momentum tensor is in the Verma module of the ground state and it cannot have nonzero quantum numbers in the $\widetilde{Q}$ invariant theory. This circumstance is related to the Feigin–Fuchs construction,[10] which implies that because of the presence of charges at infinity the $Q = 0$ sectors of the Gaussian and $c < 1$ theories are different.

One has to use the freedom left in identifying sectors of given Q and k for the solution of this apparent contradiction. Instead of accepting (40) (using appropriate combinations of $K_2 = 1, 2, 3$, and 5 terms) as the twisted partition function we can introduce a matrix X of rational (integer) elements such that

$$N^{Q,k}(x_1,x_2) = \sum_{Q',k'} X^{Q,k}_{Q',k'} \widetilde{N}^{Q',k'}(x_1,x_2), \tag{42}$$

where the matrix X satisfies

$$\sum_{Q',k'} X^{Q,k}_{Q',k'} \cos\left(2\pi \frac{Q''k' + Q'k''}{n}\right) = \sum_{Q',k'} \cos\left(2\pi \frac{Qk' + Q'k}{n}\right) X^{Q',k'}_{Q'',k''}, \tag{43}$$

so that $N^{Q,k}$ would have correct modular transformation properties (16).

Another constraint on matrix X is (with appropriate normalization)

$$X^{Q,k}_{1,0} = X^{Q,k}_{n-1,0} = m\delta^n_{Q,0}\delta^n_{k,0} \tag{44}$$

(where m is an arbitrary integer), ensuring that the ground state appears in the correct sector. Furthermore, summing over Q and multiplying by the correct phase factor should lead to twisted partition functions $\overline{Z}^k(x_1,x_2)$, which should be combinations of admissible eigenfunctions (39) of S. Consequently, the matrix X should also satisfy the equation

$$\sum_Q X^{Q,k}_{Q',k'} = C^k_{k'}, \tag{45}$$

where $C^k_{k'}$ is independent of Q'. Finally, the phase condition requires that nonzero components of X should satisfy the constraint

$$Qk = Q'k' \mod n. \tag{46}$$

Combining (43)–(45) one gets the following sum rule for $C^k_{k'}$:

$$\sum_{k'=0}^{n-1} C^k_{k'} \cos\left(2\pi \frac{k'}{n}\right) = m, \tag{47}$$

where m is an integer. Introducing

$$D^k_{k'} = c^k_{k'} + c^k_{-k'}, \quad \text{for } k' > 0,$$
$$d^k_0 = c^k_0,$$

the sum rule for $d^k_{k'}$ reads as

$$\sum_{k'=0}^{(n-1)/2} d^k_{k'} \cos\left(2\pi \frac{k'}{n}\right) = m. \tag{48}$$

Equation (48) should have $(n + 1)/2$ independent "solutions," for $k = 0, 1, ..., (n - 1)/2$, so that the $(n + 1)/2$ independent twisted partition functions $\overline{Z}^k$ could be constructed. The trivial solution $d^0_{k'} = m\delta_{k',0}$ must correspond to periodic boundary conditions since it involves integer spins only ($Qk = 0 \mod n$).

Notice that one can express

$$\cos\left(2\pi \frac{k}{n}\right) = \sum_{j=0}^{k} r^k_j z^j, \tag{49}$$

where $z = \cos(2\pi/n)$ and $r^k_j \in Z$; consequently (48) is an $[(n - 1)/2]$th-order algebraic equation for z. For $n > 3$, z is an irrational but algebraic number of order $(n - 1)/2$, its unique (up to an integer) irreducible polynomial is of order $(n - 1)/2$.[11] As a result, one can find only one twisted partition function besides the trivial (periodic boundary condition) for all n, implying $(n - 1)/2 \leqslant 1$, $n \leqslant 3$. In other words, one can see on a purely algebraic basis that no Z_n symmetry with $n > 3$ can be realized in a $c < 1$ CFT. Note that this result does not rely on the form of known positive definite partition functions; consequently it is valid for hitherto undiscovered $c < 1$ CFT's as well.

Now take one of the CFT's of (28) and (29) with $M/3 \in Z$. Not all of these theories can support Z_3 symmetry. This can be seen, e.g., by considering the $Q = 1$, $k = 1$ sector, in which, as a result of constraint (46), the matrix X is proportional to the identity matrix (except for a trivial mixing with the charge conjugated, $Q = 2$, $k = 2$ sector). Let us write $m' = 6t$ and $m = 3\mu + \nu$, where $t, u \in Z$, $\nu = 1,2$. Taking $n = 3$, the appropriate combinations of the $Q = 1$, $k = 1$ component of the partition functions (40) are as follows:

$$\left[\sum (-1)^{v+\bar{v}} \delta^{12t}_{s,3\alpha - 2\nu + 2t} \delta^{12t}_{\bar{s},-3\alpha + 2\nu + 2t} + s \leftrightarrow \bar{s}\right] \delta_{r,\bar{r}}, \tag{50}$$

for the $K_2 = 1$ contribution, and

$$\left[\sum (-1)^{v+\bar{v}} \delta^{12t}_{s,6\alpha - 4\nu + t} \delta^{12t}_{\bar{s},-6\alpha + 4\nu + t} + \sum (-1)^{v+\bar{v}} \delta_{s,6\alpha - 4\nu + 4t} \delta_{\bar{s},-6\alpha + 4\nu + 4t} + s \leftrightarrow \bar{s}\right] \delta_{r,\bar{r}}, \tag{51}$$

for the $K_2 = 2$ contribution. Here s and r are defined as in (6). The range of summation over α is set by the conditions $4t - 1 \geqslant \alpha \geqslant 0$ in (50), $2t - 1 \geqslant \alpha \geqslant 0$ in (51). The signs before individual terms (controlled by $v, \bar{v} \in Z$) are related to the reduction of s and $\bar{s}$ to the fundamental domain ($1 \leqslant s, \bar{s} \leqslant 6t - 1$) by transformations $s \to 12tu + (-1)^v s$, $\bar{s} \to 12t\bar{u} + (-1)^{\bar{v}}\bar{s}$, where $u, \bar{u} \in Z$. After reducing s and $\bar{s}$ to the fundamental domain, individual terms of sum (50) satisfy one of the constraints $s + \bar{s} = 4t$, $s - \bar{s} = \pm 4t$, or

$s+\bar{s}=8t$. Similarly, individual terms of the sums of (51) satisfy one of the constraints $s+\bar{s}=2t$, $s-\bar{s}=\pm 2t$, $s+\bar{s}=10t$, $s+\bar{s}=8t$, $s-\bar{s}=\pm 4t$, or $s+\bar{s}=4t$. The terms in which $s-\bar{s}=\pm 2t$ or $\pm 4t$ are negative. The $K_2=1$ term is not positive definite for $t>1$, consequently the partition function (28) cannot be extended to a Z_3 invariant model. When the difference of the $K_2=1$ and $K_2=2$ terms is taken, corresponding to the partition function (29), then for $t\geqslant 3$ the term of (51) having $s+\bar{s}=10t$ has a nonvanishing contribution, which cannot be compensated for by any of the other terms in the fundamental domain. On the other hand, terms of the $K_2=1$ contribution (50) satisfying $s+\bar{s}=8t$ have opposite sign. There are half as many $s+\bar{s}=8t$ terms in the $K_2=2$ contribution as well, still leaving uncompensated $s+\bar{s}=8t$ contributions. They are of opposite sign to the $s+\bar{s}=10t$ contribution, giving an indefinite partition function.

Finally in the $t=2$ case ($m'=12$) the projected $k=1$, $Q=1$ sector is

$$2\delta_{s,2}\delta_{\bar{s},6}+\delta_{s,5}\delta_{\bar{s},3}-\delta_{s,11}\delta_{\bar{s},3}-\delta_{s,10}\delta_{\bar{s},6},\quad \text{for } \nu=2,$$

$$2\delta_{s,6}\delta_{\bar{s},2}+\delta_{s,3}\delta_{\bar{s},5}-\delta_{s,9}\delta_{\bar{s},1}-\delta_{s,6}\delta_{\bar{s},10},\quad \text{for } \nu=1.$$

These expressions are obviously indefinite.

In the $t=1$ ($m'=6$) case the matrix X can be uniquely determined from the constraints (43)–(46) and it gives a partition function in complete agreement with the results of Cardy[4] and Zuber[5]:

$$\delta_{r,\bar{r}}(\delta_{s,1}\delta_{\bar{s},1}+\delta_{s,1}\delta_{\bar{s},5}),\quad \text{for the } (k=0,\,Q=0),$$

$$\tfrac{1}{2}\delta_{r,\bar{r}}\delta_{s,3}\delta_{\bar{s},3},$$

for the $(Q=1,k=0)$, $(Q=2,\,k=0)$, $(Q=0,k=1)$, $(Q=0,\,k=2)$,

$$\delta_{r,\bar{r}}\delta_{s,3}\delta_{\bar{s},1},$$

for $(Q=1,\,k=1)$, $(Q=2,\,k=2)$,

$$\delta_{r,\bar{r}}\delta_{s,1}\delta_{\bar{s},3},$$

for $(Q=1,\,k=2)$, $(Q=2,\,k=1)$ states.

The last two of these assignments are valid for $m=1 \bmod 3$. If $m=2 \bmod 3$ they should be switched.

V. CONCLUSION

The classification of a $c<1$ CFT with periodic and twisted boundary conditions has been performed using modular properties of the characters of the Virasoro algebra alone. The most general eigenfunctions of operator S, exchanging toroidal boundaries, are of the form

$$\Psi_{\Delta,\bar{\Delta}}\sim\cos(2s\pi)N_{\Delta,\bar{\Delta}},$$

where $s=\Delta-\bar{\Delta}$ is the spin, Δ and $\bar{\Delta}$ are the chiral dimensions of primary fields, and $N_{\Delta,\bar{\Delta}}$ is an integer. All partition functions with periodic or twisted boundary conditions have been shown to be related to these eigenfunctions of S.

All the partition functions found by Cardy,[1,4] Cappelli *et al.*,[2] Gepner,[3] and Zuber[5] were recovered. Furthermore, using algebraic methods, it has been proved that no $c<1$ minimal CFT may have Z_n symmetry groups with n larger than 3. Imposing the constraint of positiveness on partition functions restricts the set of models having Z_3 symmetry to those having $m'=6$.

Our methods also lead to a generalization of the construction of Di Francesco *et al.*[6] of a $c<1$ CFT with periodic boundary conditions from periodic Gaussian partition functions. It is shown that in the case of the existence of a Z_n symmetry, the known generalization of Gaussian partition functions to twisted boundary conditions[7] directly leads to the twisted partition functions in $c<1$ theories.

ACKNOWLEDGMENTS

The author is indebted to I. Halliday and V. Rittenberg for valuable discussions. The author thanks the Blackett Laboratory of the Imperial College of Science and Technology for its kind hospitality.

The support of the U.K. Science Research Council and of the U.S. Department of Energy under Contract No. FG02-84-ER40153 is gratefully acknowledged.

APPENDIX A: EIGENVECTORS OF OPERATOR S

It follows from (24) that the coefficients $c_{a,b}$ of (25) satisfy the symmetry requirements

$$c_{a,b}=R_1c_{a,b}=c_{-a,-b},\quad c_{a,b}=R_2c_{a,b}=c_{a+2b,-b-a}. \tag{A1}$$

For such $c_{a,b}$ we could combine (24) and (25) to give

$$\Psi=\sum_{a,b}c_{a,b}\exp\left\{2\pi i\frac{ax_1}{M}\right\}\delta^M_{x_2,b}. \tag{A2}$$

Then (20′) implies, after Fourier transforming in x_1,

$$c_{x_1,x_2}\sim(1+R_1)(1+R_2)\sum_{x_1'}\exp\left\{-2\pi i\frac{x_1x_1'}{M}\right\}$$
$$\times\cos\left(2\pi\frac{x_1'x_2}{M}\right)N(x_1',x_2). \tag{A3}$$

In other words, although $c_{a,b}$ is not necessarily integer, it can be written as

$$c_{a,b}=\sum_{\gamma=0}^{M-1}m^{\gamma}_{a,b}\exp\left\{2\pi i\frac{\gamma}{M}\right\}, \tag{A4}$$

where $m^{\gamma}_{a,b}\in Z$.

For convenience we relabel eigenfunctions (24) as

$$\Psi_{a,b}\to\Psi_{a+b,b}=\exp\left\{2\pi i\frac{x_1x_2}{M}\right\}\exp\left\{2\pi i\frac{ax_1}{M}\right\}\delta^M_{b,x_2}$$
$$+\exp\left\{-2\pi i\frac{x_1x_2}{M}\right\}\exp\left\{2\pi i\frac{ax_1}{M}\right\}\delta^M_{a+b,x_2}$$
$$+(x_1,x_2)\to(-x_1,-x_2). \tag{A5}$$

Now define the space of admissible eigenfunctions as θ. Also define the space of (a,b,γ) triples, $0\leqslant a,b,\gamma\leqslant M-1$ as Ω. Then the support of a $\Psi\in\theta$ is the set $S\subset\Omega$ of triples (a,b,γ), such that $m^{\gamma}_{a,b}\neq 0$.

The support S satisfies the constraints that if $(a,b,\gamma)\in S$ then $R_1(a,b,\gamma)=(-a,-b,\gamma)\in S$, $R_2(a,b,\gamma)=(a+2b,-b-a,\gamma)\in S$, and $R_3(a,b,\gamma)=(-a-2b,n,-\gamma)\in S$. The last of these statements follow from $\mathrm{Im}(\Psi)=0$.

Scanning all possible initial contributions $(a_0,b_0,\gamma_0)\in S$

and adding to them other terms $(a,b,\gamma)\in S$ to build a $\Psi\in\theta$ will comprise our procedure of building all possible $\Psi\in\theta$.

It will be sufficient to consider *minimal extensions* of contributions (a_0,b_0,γ_0) to obtain a $\Psi\in\theta$. Such a $\Psi\in\theta$ extension will be called minimal if it cannot be written as

$$\Psi\sim n^{(1)}\Psi^{(1)}+n^{(2)}\Psi^{(2)}, \tag{A6}$$

where $n^{(1)},n^{(2)}\in Z$, $\Psi^{(1)},\Psi^{(2)}\in\theta$, and where $\Psi^{(1)}$ and $\Psi^{(2)}$ are linearly independent and their supports $S^{(1)}$ and $S^{(2)}$ satisfy

$$S^{(1)}\subset S,\quad S^{(2)}\subset S. \tag{A7}$$

It is clear that if an extension is nonminimal then one can find linear combinations of $\Psi^{(1)}$ and $\Psi^{(2)}$ such that

$$\Psi\sim\bar{n}^{(1)}\Psi^{(1)}+\Psi^{(3)},$$

where $\Psi^{(3)}\in\theta$, $S^{(3)}\subset S$, and $(a_0,b_0,\gamma_0)\in S^{(3)}$. Then, of course, $\Psi^{(3)}$ would be obtained as the minimal extension of another contribution $(a_1,b_1,\gamma_1)\in S$ and Ψ need not be considered.

We order all initial contributions (a_0,b_0,γ_0) according to the divisors of M contained in a_0.

(i) Take first $a_0=0 \bmod M$. There is no need to include other values of a in the extension of this contribution, because at arbitrary fixed b we obtain

$$\Psi\sim\cos\left(2\pi\frac{x_1x_2}{M}\right)\sum_\gamma e^{i\gamma}m^\gamma_{0,b}\delta_{b,x_2}$$

$$\sim\cos\left(2\pi\frac{x_1x_2}{M}\right)\delta_{b,x_2} \tag{A8}$$

(where γ had to be chosen to make the overall coefficient integer) an admissible eigenfunction.

(ii) Next take all $a_0=0 \bmod(M/p)$, where p is a prime divisor of M. Then write $a=(M/p)k$, $x_1=pk_1+r_1$, and $x_2=(M/p)k_2+r_2$. Substituting these expressions into the general form of Ψ [for simplicity we drop terms symmetrizing in $(x_1,x_2)\to(-x_1,-x_2)$] we obtain the following equation, expressing the fact that $\Psi\in\theta$:

$$\Psi\sim e^{i\phi}\sum_{k,\gamma}m^\gamma_{kM/p,k_2M/p+r_2}e^{i\chi}+e^{-i\phi}\sum_{k,\gamma}m^\gamma_{kM/p,(k+k_2)M/p+r_2}e^{i\chi}=2\cos\phi\,N(x_1,x_2), \tag{A9}$$

where

$$\phi=2\pi\left(\frac{r_1r_2}{M}+\frac{pk_1r_2}{M}+\frac{r_2k_1}{p}\right) \tag{A10}$$

and

$$\chi=2\pi\left(\frac{r_1k}{p}-\frac{\gamma}{M}\right). \tag{A11}$$

Since on the left-hand side of (A9) k_1 appears only in the phase factors ϕ, (A9) requires that

$$\sum_{k,\gamma}m^\gamma_{kM/p,k_2M/p+r_2}e^{i\chi}=\sum_{k,\gamma}m_{kM/p,(k+k_2)M/p+r_2}e^{i\chi}=N(x_1,x_2).$$

In other words, in both sums of the above equation the constraint

$$Mr_1k/p-\gamma=0 \bmod M \tag{A12}$$

must be satisfied, requiring $\gamma=0 \bmod(Mk/p)$. Then we can write $\gamma=Mkr/p$ and (A12) reads as

$$k(r_1-r)=0 \bmod p. \tag{A13}$$

Equation (A13) requires either $k=0 \bmod p$ or $r_1-r=0 \bmod p$. The first of these possibilities leads to eigenfunctions (A8) (i.e., no new $\Psi\in\theta$), while the second requires that $m^\gamma_{u,v}$ be independent of both u and v. Performing the appropriate summations gives us the invariants

$$\Psi\sim\cos\left(2\pi\frac{x_1x_2}{M}\right)\sum_{k_1=0}^{M/K-1}\sum_{k_2=0}^{2K-1}(\delta^M_{x_1,k_1K+r_1}\delta^M_{x_2,k_2M/K+r_2}+(r_1,r_2)\to(-r_1,-r_2))+(x_1\leftrightarrow x_2), \tag{A14}$$

where we substituted $p\to K$ for reasons that will become clear below.

(iii) In the next step we admit $a_0=0 \bmod(M/K)$, where $K=p_1p_2$, p_1 and p_2 are two (possibly identical) prime divisors of M, and we seek minimal extensions of these contributions. Following our previous procedure we write $a=(M/K)k$, $x_1=Kk_1+r_1$, and $x_2=(M/K)k_2+r_2$, and, after a series of arguments similar to those in (ii), we obtain the condition for the extension Ψ to satisfy $\Psi\in\theta$:

$$Mr_1k/K-\gamma=0 \bmod M.$$

Then, of course, $\gamma=0 \bmod (Mk/K)$, so writing $\gamma=rMk/K$ we obtain

$$k(r-r_1)=0 \bmod p_1p_2. \tag{A15}$$

Choosing $k=0 \bmod M$, $k=0 \bmod(M/p_1)$, or $k=0 \bmod(M/p_2)$, a is restricted to the subsets discussed in (i) and (ii), i.e., no new invariants are obtained. Then we are left with the choice of $r-r_1=0 \bmod p_1p_2$. Following the arguments of (ii) we find that the new invariants are of the form (A14), where $K=p_1p_2$.

(iv) Continuing our procedure in the spirit of (i)–(iii), taking more and more prime factors of M, we arrive at the conclusion that the general form of invariants is (A14), where K is an arbitrary divisor of M.

APPENDIX B: PROOF OF EQ. (30)

We prove that if K is a divisor of m' then

$$RM^K(x_1,x_2)=-RM^{mK}(x_1,x_2), \tag{B1}$$

where R imposes antisymmetry with respect to the exchanges $\lambda\to\lambda_0\lambda$ and $\bar\lambda\to\lambda_0\bar\lambda$. Here $M^K(x_1,x_2)$ (the subscripts $r_1=r_2=0$ were dropped) has been defined in (27).

Since $\lambda=x_1+x_2$, $\bar\lambda=x_2-x_1$, we have, on the left-hand side of (B1),

$$\lambda=m(m'/K)k_1+Kk_2,\quad \bar\lambda=m(m'/K)k_1-Kk_2, \tag{B2}$$

whereas, on the right-hand side, we have

$$\lambda=(m'/K)t_1+mKt_2,\quad \bar\lambda=(m'/K)t_1-mKt_2, \tag{B3}$$

where the ranges of variables k_1, k_2, t_1, and t_2 are from 0 to $2K-1$, $M/K-1$, $2mK-1$, *and* $m'/K-1$, respectively. We need to show that when k_1, k_2, t_1, and t_2 run over their range, the congruences

$$\lambda_0((m'/K)t_1+mKt_2)=-(m(m'/K)k_1+Kk_2) \bmod 2M, \tag{B4}$$

$$(m'/K)t_1 - mKt_2 = m(m'/K)k_1 - Kk_2 \mod 2M$$

are satisfied.

Let us write

$$t_1 = mk'_1 + Kk_3. \tag{B5}$$

Since m and K are relatively prime, t_1 runs over its complete range $\mod 2Km$, when the range of k_1 and k_3 is $0 \leqslant k'_1 \leqslant 2K-1$ and $0 \leqslant k_3 \leqslant m-1$, respectively. Substituting (B5) into (B4) we obtain

$$-m(m'/K)k'_1 + m'k_3 - mKt_2 = -m(m'/K)k_1 - Kk_2, \tag{B6}$$

$$m(m'/K)k'_1 + m'k_3 - mKt_2 = m(m'/K)k_1 - Kk_2.$$

Equation (B6) implies $k_1 = k'_1$ and

$$mt_2 - (m'/K)k_3 = k_2 \mod M/K. \tag{B7}$$

Equation (B7) is satisfied, because when t_2 and k_3 run over their respective ranges the left-hand side runs over all possible numbers mod M/K, just like the right side, since m and m'/K are relatively prime.

[1] J. L. Cardy, Nucl. Phys. B **270** [FS16], 186 (1986).
[2] A. Cappelli, C. Itzykson, and J.-B. Zuber, Nucl. Phys. B **280** [FS18], 415 (1987); J. Math. Phys. **113**, 1 (1987).
[3] D. Gepner, Nucl. Phys. B **287**, 111 (1987).
[4] J. L. Cardy, Nucl. Phys. B **275** [FS17], 200 (1986).
[5] J.-B. Zuber, Phys. Lett. B **176**, 127 (1986).
[6] P. Di Francesco, H. Saleur, and J.-B. Zuber, Nucl. Phys. B **285** [FS19], 454 (1987).
[7] P. Suranyi, J. Phys. A **21**, 401 (1988).
[8] L. Dixon, J. Harvey, C. Vafa, and E. Witten, Nucl. Phys. B **261**, 678 (1985); **274**, 285 (1986).
[9] A. Rocha-Caridi, in *Vertex Operators in Mathematics and Physics*, edited by J. Lepowski, S. Mandelstam, and I. Singer (Springer, New York, 1985).
[10] V. S. Dotsenko and V. A. Fateev, Nucl. Phys. B **240**, 312 (1984); **251**, 691 (1985).
[11] G. Birkhoff and S. MacLane, *A Survey of Modern Algebra* (MacMillan, New York, 1953).

Some global properties and invariance of bundle metrics in the Kaluza–Klein scheme

C. H. Oh and K. Singh
Department of Physics, National University of Singapore, Kent Ridge, Singapore 0511, Republic of Singapore

C. H. Lai[a]
International Centre of Theoretical Physics, Trieste, Italy

(Received 18 May 1988; accepted for publication 3 August 1988)

It is shown that the transition functions that give the global structure of the fiber bundle play an important role in the construction of the metric. The invariance properties of this metric under general gauge transformations are discussed and it is found that the usual requirement of a gauge-invariant metric leads to severe constraints on the gauge fields. To avoid them, it is shown that the metric should instead be covariant with respect to these transformations. Moreover the existence of global actions that are essential in the context of the consistency problem is also discussed. The presence of such actions is studied in both the principal and their associated bundles. In the case of a homogeneous bundle with G/H as the typical fiber, it is shown that a "spliced" bundle with $G\times N(H)/H$ as the structure group has to be used. The unified space is then taken as the bundle space of its associated bundle.

I. INTRODUCTION

Kaluza–Klein theory, when formulated in the framework of fiber bundles, provides an interesting and elegant way of geometrizing gravity with the other gauge interactions.[1–6] Indeed, when the bundle space is given a Riemannian structure this approach offers not only a more transparent interpretation of the Kaluza–Klein *Ansätze* but also allows one to "derive" the metric from more basic assumptions. The theory, which is Einstein's theory in $(4+n)$ dimensions, assumes that the extra n-dimensional space compactifies by some dynamical mechanism to a size ($\sim 10^{-33}$ cm) that is unresolvable at current available energies. Unlike the approach of Refs. 7 and 8, the split between the four-dimensional space and the compact n-dimensional internal space is not required to be global. This generalizes the earlier notion of the $(4+n)$-dimensional space being a global cross product.

The purpose of this paper is to construct a metric that is compatible with the bundle structure and study the restrictions that arise under various assumptions. In the following section, we will examine some of the assumptions that are made in constructing the bundle metric. In particular, it will be shown that the transition functions that define the global structure of the fiber bundle play an important role in its construction. Furthermore, we find that they allow the metric on the vertical subspaces to be dependent on the base manifold, thus introducing Brans–Dicke-like scalars. In fact, in our construction, these scalar fields can be directly attributed to the nontriviality of the transition functions. In Sec. III we study the properties of these metrics under general gauge transformations. It is shown that the usual requirement of a gauge-invariant metric[5] is not only unnecessary but also leads to severe constraints on the gauge potentials. In Sec. IV we turn to the consistency problem.[9–12] Here we provide a geometrical basis to the claim that a G-invariant scheme is consistent. The scheme that essentially requires the presence of a global action is applied to a homogeneous bundle with typical fiber G/H. Unlike the principal bundle, (P,M,π,G), on which the right action is canonically given, there is no corresponding action on the associated bundle $(E,M,\pi_E,G/H,G,P)$. Here the bundle automorphisms that are well defined can be used instead. However, if these actions are assumed to be isometries, then in light of Sec. III, they imply vanishing gauge potentials. In order to resolve this, we construct a bundle associated to a "spliced" bundle that admits the full isometry group of G/H, i.e., $G\times N(H)/H$, as the structure group. This is discussed in Sec. IV. Subsequently by placing a suitable restriction on the metric of the typical fiber, we obtain a scheme that is consistent.

II. BUNDLE METRICS

To begin with, let us recall some requirements needed when constructing a metric on a bundle space. In Ref. 2, these are summarized as follows.

(i) The horizontal subspaces of the tangent space to the bundle must be orthogonal in this metric to the vertical subspaces.

(ii) The projection of the metric onto the horizontal space must be isomorphic with the Riemannian metric of the base manifold.

(iii) The vertical part of the metric must be isomorphic to some metric of the space tangent to the fiber; i.e., to some metric on the Lie algebra of the structure group.

Condition (i), essentially, expresses the notion of compatibility between the Riemannian structure and the gauge structure so that the horizontal subspaces defined in both cases are consistent. Furthermore, it also implies that the

[a] On leave from Physics Department, National University of Singapore, Republic of Singapore.

 0022-2488/88/122641-12$02.50

fibers are Riemannian submanifolds. This results from the following. If the manifold P is endowed with a metric tensor $\mathfrak{g}_P$ and if Q is a submanifold of P then it is also a Riemannian submanifold if $j^*\mathfrak{g}_P$ is the metric tensor on Q. (Here j: $Q\subset P$ is the inclusion map.) Now, the tensor $\mathfrak{g}_Q$ on Q derived in this way is a metric tensor if and only if the tangent space $T_p(Q)$ is nondegenerate in $T_p(P)$ with respect to $\mathfrak{g}_P$ for each $p\in Q\subset P$. This can only hold if $T_p(P)$ is the direct sum of $T_p(Q)$ and $T_p(Q)^\perp$, where $T_p(Q)^\perp$ denotes the subspace of $T_p(P)$ that is orthogonal to $T_p(Q)$ relative to $\mathfrak{g}_P$.[13] When this is taken together with condition (ii), the map π: $P\to M$, where P is the bundle space and M the base space of a principal fiber bundle (P,M,π,G) becomes a Riemannian submersion (as defined in Ref. 13). In general, this assumption alone does not guarantee the existence of local triviality and, in some sense, is more general than a fiber bundle when the latter is also endowed with a metric. Later we will look at a sufficient condition that will ensure a bundle structure when starting with a Riemannian submersion. Here, however, we will regard π: $P\to M$ as a fiber bundle projection and admit condition (iii) as an additional assumption.

Using the three conditions above, a metric on $\pi^{-1}(U)\subset P$ in a trivialization ψ: $\pi^{-1}(U)\to U\times G$, with $\psi(p)=(\pi(p),\varphi(p))$, can be written as

$$\hat{\mathfrak{g}}_{\pi^{-1}(U)}(V,W)$$
$$=\pi^*\mathfrak{g}_U(V,W)+\varphi^*\mathfrak{h}(V-\hat{V},W-\hat{W}), \qquad (1)$$

for all vector fields V, W defined on $\pi^{-1}(U)$. Here $\mathfrak{g}_U$ and $\mathfrak{h}$ are the metric tensors defined on $U\subset M$ and G, respectively, and $\hat{V}$ denotes the horizontal component of V. It should also be stressed that the metric on the vertical subspaces is independent of the base manifold since $\mathfrak{h}$ is characterized by the coordinates of G alone.

In Ref. 5, the metric (1) is generalized by first assigning a x-dependent metric on G and then pulling back via φ_x, where φ_x: $\pi^{-1}(x)\to G$. Essentially, this allows the vertical part of the full metric to vary from fiber to fiber which, in turn, introduces Brans–Dicke-like scalars. In the following it will be shown that this x dependence arises more naturally when the metric is considered globally.

First, let us consider an open covering $\{U_i\}_{i\in I}$ of M. Then, with $\pi^{-1}(U_i)$ as open subsets of P, $U_{i\in I}\pi^{-1}(U_i)$ can be regarded as an opening covering of P by identifying points in the overlap regions $\pi^{-1}(U_i)\cap\pi^{-1}(U_j)$. If both M and P are assumed to be paracompact then their respective coverings are locally finite.[14,15] By this we mean that for each point of the space, there exists a neighborhood that has a nonempty intersection with only a finite number of U_i's [respectively, $\pi^{-1}(U_i)$'s]. Here we would like to emphasize that the condition of paracompactness is not an additional requirement since it is implied when one regards a manifold as a Riemannian space. Indeed, a manifold can be given a proper Riemannian structure if and only if it is paracompact.[15,16] Hence with these conditions, it is possible to find a partition of unity $\{f_i(p)\}$ subordinate to the covering $\{\pi^{-1}(U_i)\}$. Since we have already constructed a metric (1) on each open neighborhood $\pi^{-1}(U_i)$, we can extend this to P by taking the finite sum for each $p\in P$,

$$\hat{\mathfrak{g}}_p(V,W)=\sum_i f_i(p)\mathfrak{g}_p^{(i)}(V,W)$$
$$=\sum_i\big(f_i(p)\pi^*\mathfrak{g}^{(i)}_{\pi(p)}(V,W)$$
$$+f_i(p)\varphi_i^*\mathfrak{h}_{\varphi_i(p)}(V-\hat{V},W-\hat{W})\big), \qquad (2)$$

where we have used the superscript (i) to label the neighborhood on which the metric is being considered. Now if condition (ii) above is to be satisfied, then we must use the partition functions that depend on x alone. In other words we need only to consider a partition of unity $\{k_i(x)\}$ subordinate to the open covering $\{U_i\}_{i\in I}$ on M since

$$\sum_i k_i(x)\pi^*\mathfrak{g}_x^{(i)}(V,W)=\pi^*\sum_i k_i(x)\mathfrak{g}_x^{(i)}(V,W)$$
$$=\pi^*\mathfrak{g}_x(V,W). \qquad (3)$$

Then if we fix a particular φ_j, Eq. (2) can be rewritten as

$$\hat{\mathfrak{g}}_p(V,W)=\pi^*\mathfrak{g}_{\pi(p)}(V,W)+\sum_i k_i(x)$$
$$\times(L_{\phi_{ij}(x)}\varphi_j)^*\mathfrak{h}_{\phi_{ij}(x)\varphi_j(p)}(V-\hat{V},W-\hat{W}), \qquad (4a)$$

or more generally,

$$\hat{\mathfrak{g}}(V,W)=\pi^*\mathfrak{g}(V,W)+\sum_i k_i(x)$$
$$\times(L_{\phi_{ij}(x)}\circ\varphi_j)^*\mathfrak{h}(V-\hat{V},W-\hat{W}), \qquad (4b)$$

where ϕ_{ij}: $U_i\cap U_j\to G$ are the transition functions. From the above equation it can be observed that if we want to retain the x dependence in the metric on the vertical subspaces then $\mathfrak{h}$ should not be left invariant. This is because for a left-invariant $\mathfrak{h}$, we have

$$\sum_i k_i(x)(L_{\phi_{ij}(x)}\circ\varphi_j)^*\mathfrak{h}(V-\hat{V},W-\hat{W})$$
$$=\sum_i k_i(x)\varphi_j^*L^*_{\phi_{ij}(x)}\mathfrak{h}(V-\hat{V},W-\hat{W})$$
$$=\sum_i k_i(x)\varphi_j^*\mathfrak{h}(V-\hat{V},W-\hat{W})$$
$$=\varphi_j^*\mathfrak{h}(V-\hat{V},W-\hat{W}), \qquad (5)$$

where the last step follows from the fact that $\{k_i(x)\}$ are the partition functions and they satisfy $\Sigma_i k_i(x)=1$, for all $x\in M$. It is clear from (5) that it is independent of x. Furthermore, if $\mathfrak{h}$ is not left-invariant then this x dependence can vanish only when the bundle is trivial, since, for trivial bundles one can always construct a global section with $\phi_{ij}(x)=\mathrm{Id}$, $\forall x\in M$. But this does not mean that one cannot retain this x dependence for trivial bundles, since a trivialization with a global section is not the only possible choice.

Now let us consider the case in which the typical fiber is the homogeneous space G/H instead of G. It is well known that such a fiber bundle $(E,M,\pi_E,G/H,G,P)$, which is associated to (P,M,π,G), can be constructed by taking the quotient of $P\times G/H$ by the right action of G which is defined by

$$(p,y)g=(pg,g^{-1}y), \quad \forall p\in P, \quad y\in G/H, \quad g\in G. \qquad (6)$$

It is worth noting that the bundle space E that is characterized by the points $w = p\cdot y$ can also be regarded as the base space of the principal fiber bundle (P,E,μ,H), where μ: $P\to E$ is the bundle projection.[14]

In analogy with the metric (4) on P, the metric on E can be written as

$$\hat{\mathfrak{g}}_E(V,W) = \pi_E^*\mathfrak{g}_M(V,W) + \sum_i k_i(x)\overline{\varphi}_i^*\mathfrak{h}_{G/H}(V-\hat{V},W-\hat{W}), \tag{7}$$

where $\overline{\varphi}$: $\pi_E^{-1}(U)\to G/H$ is the trivialization satisfying

$$\overline{\varphi}\circ\mu = \vartheta\circ\varphi \tag{8}$$

and $\mathfrak{h}_{G/H}$ is a metric tensor on G/H. Here φ is the trivialization on P and ϑ defines the projection ϑ: $G\to G/H$. Note that a connection on E can be obtained by projecting the horizontal subspaces of P by the differential map μ_*.

Since E can be regarded as the base space of a principal bundle (P,E,μ,H), it is perhaps interesting to ask whether the map μ: $P\to E$ is a Riemannian submersion, when the metrics on the manifolds P and E are given by (3) and (7), respectively. The answer is in the affirmative, provided the map ϑ: $G\to G/H$ is also a Riemannian submersion. This can be shown as follows. Since μ is a bundle projection, it is obvious that it is a submersion in the ordinary sense. Now, if ϑ is a Riemannian submersion then the metric tensor on G, which is the bundle space of the principal bundle $(G,G/H,\vartheta,H)$, can be written as

$$\mathfrak{h}_G(X,Y) = \vartheta^*\mathfrak{h}_{G/H}(X,Y) + \sum_j l_j(y)\eta_j^*\mathfrak{h}_H(\mathrm{ver}(X),\mathrm{ver}(Y)), \tag{9}$$

where η_j: $\vartheta^{-1}(W)\to H$ is the trivialization; $\mathrm{ver}(X)$ denotes the vertical component of X and $\{l_j(y)\}$ are the partition functions on G/H. [Here we would like to remark that the assumption that $(G,G/H,\vartheta,H)$ is a principal fiber bundle is not an additional one. This is because in showing the existence of local triviality in (P,E,μ,H) one has already made use of this.[14]] Then by noting that $\pi = \pi_E\circ\mu$ and using Eq. (8) we have

$$\begin{aligned}
\hat{\mathfrak{g}}_P(V,W) &= \pi^*\mathfrak{g}_M(V,W) + \sum_i k_i(x)\varphi_i^*\mathfrak{h}_G(V-\hat{V},W-\hat{W})\\
&= (\pi_E\circ\mu)^*\mathfrak{g}_M(V,W) + \sum_i k_i(x)(\vartheta\circ\varphi_i)^*\mathfrak{h}_{G/H}(V-\hat{V},W-\hat{W})\\
&\quad + \sum_{ij} k_i(x)l_j(y)\eta_j^*\mathfrak{h}_H(\mathrm{ver}\,\varphi_{i*}(V-\hat{V}),\mathrm{ver}\,\varphi_{i*}(W-\hat{W}))\\
&= \mu^*\Big(\pi_E^*\mathfrak{g}_M(V,W) + \sum_i k_i(x)\overline{\varphi}_i^*\mathfrak{h}_{G/H}(V-\hat{V},W-\hat{W})\Big)\\
&\quad + \sum_{ij} k_i(x)l_j(y)\eta_j^*\mathfrak{h}_H(\mathrm{ver}\,\varphi_{i*}(V-\hat{V}),\mathrm{ver}\,\varphi_{i*}(W-\hat{W}))\\
&= \mu^*\hat{\mathfrak{g}}_E(V,W) + \sum_{ij} k_i(x)l_j(y)\eta_j^*\mathfrak{h}_H(\mathrm{ver}\,\varphi_{i*}(V-\hat{V}),\mathrm{ver}\,\varphi_{i*}(W-\hat{W})),
\end{aligned}\tag{10}$$

which shows that μ is really a Riemannian submersion. It should be noted that for ϑ to be a Riemannian submersion, the metric on G/H must be G invariant.[13]

Now on the homogeneous space G/H we can define a left action $\overline{L}_g$: $G/H\to G/H$, which is given by $(g,y)\to g\cdot y$, $\forall g\in G$, $y\in G/H$. Then under the projection ϑ: $G\to G/H$, which sends each $g\in G$ to the coset gH, we have

$$\vartheta\circ L_g = \overline{L}_g\circ\vartheta, \quad \forall g\in G, \tag{11}$$

where L_g: $G\to G$ is a left action on G. From Eqs. (8) and (11) it is easy to verify that the trivializations on $\pi_E^{-1}(U_i)$ and $\pi_E^{-1}(U_j)$ are related by

$$\overline{\varphi}_i(w) = \overline{L}_{\phi_{ij}(x)}\circ\overline{\varphi}_j(w), \quad w\in\pi_E^{-1}(U_i)\cap\pi_E^{-1}(U_j), \tag{12}$$

where ϕ_{ij}: $U_i\cap U_j\to G$ are the transition functions on P. By replacing $\overline{\varphi}_i$ in Eq. (7) by the above expression it can be observed that the metric on the vertical subspaces loses its x dependence when $\mathfrak{h}_{G/H}$ is G invariant. This is analogous to a left-invariant metric on G in the case of a principal fiber bundle. Later we will show that the left invariance in $\mathfrak{h}_G$ (or G invariance in $\mathfrak{h}_{G/H}$), which implies a gauge invariant metric, will also lead to constraints on the gauge fields. However, before turning to the question of invariance, we will study the metric on P under more general assumptions.

So far we have constructed a metric on a manifold that has an underlying fiber bundle structure. As was mentioned earlier, the first two of the three conditions used in constructing the bundle metric imply that the map π: $P\to M$ is a Riemannian submersion. Only the third condition makes an implicit assumption of local triviality and hence a fiber bundle. It is also instructive to consider the metric by reversing the assumptions; that is, instead of assuming a fiber bundle structure from the start, we regard P as a Riemannian space with a projection π: $P\to M$ that is a Riemannian submersion and study the restrictions one must make so that π: $P\to M$

becomes a fiber bundle projection. A sufficient condition is given by the following theorem of Hermann.[17]

Theorem: If P is complete as a Riemannian space, so is M. Here π is then a locally trivial fiber space. If, in addition, the fibers of π are totally geodesic submanifolds of P, π is a fiber bundle with the structure group the Lie group of isometries of the fiber.

Essentially the proof centers on defining a principal bundle with G, which is a group of isometries of the fiber F [defined by $\pi^{-1}(x_0)$ for some $x_0 \in M$], as the structure group. In order to do this one must first show the existence of local sections and this is provided by the following proposition (see Proposition 3.3 of Ref. 17).

If all the fibers of π are totally geodesic submanifolds of P, then for each path γ: $[0,1] \to M$, the diffeomorphism h_γ: $\pi^{-1}(\gamma(0)) \to \pi^{-1}(\gamma(1))$ obtained by mapping each $w_0 \in \pi^{-1}(\gamma(0))$ into the end point of the horizontal lift of γ is an isometry of $\pi^{-1}(\gamma(0))$ onto $\pi^{-1}(\gamma(1))$.

For simplicity, we will regard the fiber F to be a group manifold G which makes π: $P \to M$ a principal fiber bundle. To see what restriction the above assumption imposes on the bundle metric, we must first introduce a connection on P by choosing it to be the horizontal subspace as defined by the metric. This will ensure compatibility between the two structures. Now, one also encounters mappings between fibers in relation to connections and these are termed as parallel displacement of fibers [α: $\pi^{-1}(\gamma(0)) \to \pi^{-1}(\gamma(1))$].[14] In fact, under the compatibility requirement, we can identify h_γ with α and since the latter commute with the right action by G, it is easy to show the fundamental vector fields on P are h_γ related:

$$h_{\gamma *} \xi^*_{w_0} = h_{\gamma *} \frac{d}{dt} w_0 \exp t\xi \bigg|_{t=0} = \frac{d}{dt} h_\gamma(w_0) \exp t\xi \bigg|_{t=0} = \xi^*_{h_\gamma(w_0)} . \quad (13)$$

Now, if the map h_γ is an isometry between the fibers $\pi^{-1}(\gamma(0))$ and $\pi^{-1}(\gamma(1))$, then it follows that for any two fundamental vector fields ξ^*, η^*,

$$\hat{\mathfrak{g}}_{w_0}(\xi^*,\eta^*) = h^*_\gamma \hat{\mathfrak{g}}_{h_\gamma(w_0)}(\xi^*,\eta^*) = \hat{\mathfrak{g}}_{h_\gamma(w_0)}(\xi^*,\eta^*) . \quad (14)$$

It is immediately apparent from (14) that the metric on the vertical subspaces, in the basis of fundamental vector fields, is independent of the position of the fiber. In other words it is x independent. Let us relate this result to the metric (4), which we constructed earlier. First, note that π: $P \to M$ is a principal bundle whose structure group is the group of all isometries of the typical fiber $F = G$. Since G acts on the left of F, then it is obvious that the metric on the typical fiber must be left-invariant. Now, we have already seen from our earlier discussion that the metric on the fiber loses its x dependence if the metric on the typical fiber is left-invariant. On this consideration, we can say that both approaches are compatible. However, the present case is more restrictive as an additional constraint arises from the requirement of totally geodesic fibers. Indeed, the shape tensor II (Ref. 13) must vanish as a result, i.e., $\hat{\mathfrak{g}}(\nabla_V W, X) = 0$ for all vertical fields V, W, and horizontal field X. Then by using the Koszul equation,[13] we have

$$X\hat{\mathfrak{g}}(V,W) - \hat{\mathfrak{g}}([X,V],W) - \hat{\mathfrak{g}}(V,[X,W]) = 0 , \quad (15)$$

which resembles the Killing equation $\mathscr{L}_X \hat{\mathfrak{g}} = 0$. However, since Eq. (15) holds only for vertical fields V, W and not all vector fields on P, X is not necessarily a Killing vector field. To make Eq. (15) more explicit, we can choose X to be the horizontal lift of ∂_μ and the fundamental vector fields ξ^*_a, ξ^*_b for the vertical fields V and W, respectively. Thus with $X = e_\mu = \tilde{\partial}_\mu - A^c_\mu(x) \mathrm{Ad}(g^{-1})^d_c \xi^*_d$ in a given trivialization, we have

$$e_\mu \mathfrak{g}(\xi^*_a, \xi^*_b) = \tilde{\partial}_\mu g_{ab} - A^c_\mu(x) \mathrm{Ad}(g^{-1})^d_c \xi^*_d(g_{ab}) = 0 , \quad (16)$$

which implies that the components of the metric $g_{ab} = \hat{\mathfrak{g}}(\xi_a, \xi_b)$, besides being x dependent, must also be constant along the fibers.

Before concluding this section, we would like to remark that the above condition is only a sufficient condition, which presupposes a left-invariant metric $\mathfrak{h}$. This left invariance not only excludes a x-dependent fiber metric but also, as we shall see in the following section, constrains the gauge fields. The nature and severity of these constraints will force us to reexamine the issue of a gauge-invariant metric and also question its necessity.

III. GAUGE-INVARIANT BUNDLE METRICS

It is well known from classical field theory that all the fundamental equations of classical physics can be obtained from one mathematical construct called the action. This functional not only yields the field equations but also characterizes the system through its symmetries. When one talks about symmetries of the system, it is usually with respect to this action. Now, when a theory is formulated on a fiber bundle, it is natural to expect that it be independent of the gauge choice. In more precise terms, this is tantamount to choosing an action that is invariant under gauge transformations. In the context of the Kaluza–Klein theory, this gauge symmetry can be incorporated into the action by requiring the metric to be gauge invariant, since it is regarded as the basic dynamical variable. Here we would like to remark that this is a sufficient rather than a necessary condition. (In Ref. 5 this is taken as an additional assumption.) This requirement, as we will show below, is not only unnecessary but also disastrous. However, before we do so, let us recall some facts about gauge transformations.

For practical calculations, one usually works locally by fixing a gauge or choosing a trivialization $\{U_i, \psi_i\}$. This amounts to choosing a coordinate bundle from the class of coordinate bundles that are equivalent in the sense of Steenrod.[18] A change in the local structure that transforms a coordinate bundle to an equivalent one is generally called a gauge transformation and can be viewed as a change in the atlas of the fiber bundle.[19] In particular, if we transform from one coordinate bundle to another bundle, then we can always find a gauge function a: $U \to G$, such that

$$\varphi \to \varphi' = a(x)^{-1} \varphi , \quad (17a)$$

where both φ and φ' are defined on $\pi^{-1}(U)$. (Here we have kept the same covering for both coordinate bundles.) Furthermore, since

$$p = \sigma(x)\varphi(p) = \sigma'(x)\varphi'(p) = \sigma'(x)a(x)^{-1}\varphi(p) ,$$

the local section σ transforms as

$$\sigma(x) \to \sigma'(x) = \sigma(x)a(x) . \tag{17b}$$

Now transformations (17) are usually termed as local gauge transformations and they constitute what is known as the passive viewpoint. One can also regard gauge transformations as global automorphisms (active viewpoint) or more precisely vertical automorphisms of a principal fiber bundle that are defined as follows.[20,21]

A vertical automorphism of a principal fiber bundle π: $P \to M$ is a diffeomorphism $\mathscr{F}: P \to P$ satisfying the following conditions:

$$\text{(i)} \quad \mathscr{F}(pg) = \mathscr{F}(p)g, \quad \forall g \in G, \quad p \in P, \tag{18a}$$

$$\text{(ii)} \quad \pi(\mathscr{F}(p)) = \pi(p), \quad \forall p \in P. \tag{18b}$$

These transformations form a group $GA(P)$ that is called the group of gauge transformations and they can be realized by defining maps τ: $P \to G$, such that

$$\mathscr{F}(p) = p\tau(p), \quad \forall p \in P. \tag{19a}$$

In order to satisfy condition (18a), we must require that

$$\tau(pg) = g^{-1}\tau(p)g, \quad \forall g \in G, \quad p \in P. \tag{19b}$$

It is easy to verify that the group of gauge transformations $GA(P)$ is isomorphic to the space of all maps τ: $P \to G$ denoted by $C(P,G)$.[20]

Although the two viewpoints seem notably different, we can establish the equivalence in the following manner: If two trivializations ψ_i, ψ_i' are given over the same covering $\cup_i U_i = M$, then the diffeomorphisms $\varphi_{i,x}$ and $\varphi'_{i,x}$ are well defined. [Here $\varphi_{i,x}$ is the restriction of φ_i to the fiber $\pi^{-1}(x)$.] Since $\varphi_{i,x}$ (respectively, $\varphi'_{i,x}$) is a diffeomorphism then both $\varphi_{i,x}$ (respectively, $\varphi'_{i,x}$) and its inverse are continuously differentiable of at least class C^1. We can then define a composite function given by

$$\mathscr{F} \equiv \varphi_{i,x}^{-1} \circ \varphi'_{i,x} , \tag{20a}$$

which is also differentiable and hence a diffeomorphism of $\pi^{-1}(x) \to \pi^{-1}(x)$. It is easy to verify that $\mathscr{F}$ defined here satisfies both conditions (18a) and (18b). Since the function φ_i is defined over the region $\pi^{-1}(U_i)$, we can extend the domain of $\mathscr{F}$ to $\pi^{-1}(U_i)$ by writing

$$\mathscr{F} \equiv \varphi_{i,x}^{-1} \circ \varphi'_i . \tag{20b}$$

It should be noted, however, that although both φ'_i and $\varphi'_{i,x}$ are differentiable, only the latter is a diffeomorphism. Then $\mathscr{F}$ in (20b) is not a diffeomorphism unless we also define its inverse as $\mathscr{F}^{-1} \equiv \varphi'^{-1}_{i,x} \circ \varphi_i$. Henceforth, we will assume this and use (20b) instead of (20a).

Conversely if $\mathscr{F}: P \to P$ satisfies (18a) and (18b), then we can define a coordinate function φ'_i such that

$$\varphi'_i(p) = \varphi_i(\mathscr{F}(p)) , \tag{21}$$

which satisfies (17a). Indeed from (19a) we have

$$\begin{aligned} \varphi'_i(p) = \varphi_i(\mathscr{F}(p)) &= \varphi_i(p\tau(\sigma_i(x)\varphi_i(p))) \\ &= \tau(\sigma_i(x))\varphi_i(p) , \end{aligned} \tag{22}$$

from which we can identify $a(x)^{-1}$ with $\tau(\sigma_i(x))$ since $\tau(\sigma_i(x)) \in G$. Essentially, Eq. (21) means that the change in coordinates produced by an active transformation is the same as that produced by taking a different trivialization.[15,22] However, this choice of a new trivialization is not arbitrary, in that, they must satisfy the condition that the transition functions remain unchanged:

$$\begin{aligned} \phi'_{ij}(x) = \varphi'_i(p)\varphi'_j(p)^{-1} &= \varphi_i(\mathscr{F}(p))\varphi_j(\mathscr{F}(p))^{-1} \\ &= \varphi_i(p)\tau(p)\tau(p)^{-1}\varphi_j(p)^{-1} \\ &= \phi_{ij}(x), \quad \forall x \in U_i \cap U_j . \end{aligned} \tag{23}$$

Furthermore, the functions $\tau(\sigma_i(x))$ on different patches are related by the transition functions as follows: for $x \in U_i \cap U_j$, we can write $p = \sigma_i(x)\varphi_i(p) = \sigma_j(x)\varphi_j(p)$ and $\tau(\sigma_i(x)\varphi_i(p)) = \tau(\sigma_j(x)\varphi_j(p))$. Then using (19b) we have $\varphi_i(p)^{-1}\ \tau(\sigma_i(x))\ \varphi_i(p) = \varphi_j(p)^{-1}\ \tau(\sigma_j(x))\ \varphi_j(p)$ from which it follows that

$$\tau(\sigma_i(x))\phi_{ij}(x) = \phi_{ij}(x)\tau(\sigma_j(x)) . \tag{24}$$

Now $C(P,\mathscr{G})$, which denotes the space of all maps $P \to \mathscr{G}$, can be regarded as the Lie algebra of $C(P,G)$. [Recall that $C(P,G)$ is isomorphic to $GA(P)$.] It inherits the Lie algebra of G in the following manner: Define $H \in C(P,\mathscr{G})$ such that

$$H(pg) = \mathrm{Ad}(g^{-1})H(p), \quad \forall p \in P, \quad g \in G , \tag{25a}$$

$$[H_1,H_2](p) = [H_1(p),H_2(p)], \quad \forall H_1 H_2 \in C(P,\mathscr{G}) . \tag{25b}$$

Then we find that $[H_1,H_2]$ is also in $C(P,\mathscr{G})$.[20] Moreover, if exp: $\mathscr{G} \to G$ is the exponential map that maps the Lie algebra $\mathscr{G}$ into G then there exists a map Exp: $C(P,\mathscr{G}) \to C(P,G)$, defined by

$$(\mathrm{Exp}(H))(p) = \exp(H(p)) \in G, \quad \forall p \in P . \tag{26}$$

It is easy to verify that $\mathrm{Exp}(H) \in C(P,G)$, which allows us to regard Exp as the exponential map of the Lie algebra of $GA(P)$.

Next, let us consider the metric on $\pi^{-1}(U)$ in a given trivialization [Eq. (1)] and see how it can be made gauge invariant. It can be noted that under the gauge transformation $\varphi \to \varphi' = a(x)^{-1}\varphi$ the metric transforms as $\hat{\mathfrak{g}}_{\pi^{-1}(U)}(V,W) \to \hat{\mathfrak{g}}'_{\pi^{-1}(U)}(V,W)$, where

$$\begin{aligned} &\hat{\mathfrak{g}}'_{\pi^{-1}(U)}(V,W) \\ &\quad = \pi^*\mathfrak{g}_U(V,W) + \varphi'^*\mathfrak{h}(V - \hat{V}, W - \hat{W}) \\ &\quad = \pi^*\mathfrak{g}_U(V,W) + (L_{a(x)^{-1}} \circ \varphi)^*\mathfrak{h}(V - \hat{V}, W - \hat{W}) \\ &\quad = \pi^*\mathfrak{g}_U(V,W) + \varphi^* L^*_{a(x)^{-1}}\mathfrak{h}(V - \hat{V}, W - \hat{W}) , \\ &\qquad \forall V, W \in \mathfrak{X}(\pi^{-1}(U)) , \end{aligned} \tag{27}$$

where $\mathfrak{X}(\pi^{-1}(U))$ denotes the set of all smooth vector fields on $\pi^{-1}(U)$. It is obvious from (27) that the metric is gauge invariant if $\mathfrak{h}$ is left invariant.

Under the equivalence between the active and passive viewpoints, this gauge invariance implies a local isometry, $\mathscr{F}: \pi^{-1}(U) \to \pi^{-1}(U)$ for the metric (4), since

$$\begin{aligned}
&\mathscr{F}^* \hat{\mathfrak{g}}_{\pi^{-1}(U)}(V,W) \\
&= \mathscr{F}^*\pi^*\mathfrak{g}_U(V,W) + \mathscr{F}^* \sum_i k_i(x) \\
&\quad \times \varphi_i^* \mathfrak{h}(V-\hat{V}, W-\hat{W}) \\
&= (\pi \circ \mathscr{F})^* \mathfrak{g}_U(V,W) + \sum_i k_i(x) \\
&\quad \times (\varphi_i \circ \mathscr{F})^* \mathfrak{h}(V-\hat{V}, W-\hat{W}) \\
&= \pi^* \mathfrak{g}_U(V,W) + \sum_i k_i(x) \varphi'^* \mathfrak{h}(V-\hat{V}, W-\hat{W}), \\
&\quad \forall V,W \in \mathfrak{X}(\pi^{-1}(U)). \qquad (28)
\end{aligned}$$

Here we have made use of Eqs. (18b) and (21).

To obtain the infinitesimal version of (28) we must evaluate the Killing vector fields generating $\mathscr{F}$. In particular, we need to consider the one-parameter subgroups of the group of gauge transformations $GA(P)$. Since the group $GA(P)$ can be realized by the maps $\tau: P \to G$ [Eq. (19a)], these can be defined through the assignment $t \to \mathrm{Exp}(tH)$ with

$$\frac{d}{dt}\mathrm{Exp}(tH)(p)\Big|_{t=0} = \frac{d}{dt}\exp(tH(p))\Big|_{t=0} = H(p). \qquad (29)$$

Then the one-parameter subgroup $\mathscr{F}_t$ of $GA(P)$ is given by

$$\mathscr{F}_t(p) = p\tau_t(p) = p\,\mathrm{Exp}(tH)(p) \qquad (30)$$

or equivalently, using Eqs. (20b) and (22),

$$\begin{aligned}
\mathscr{F}_t(p) &= \varphi_x^{-1} \circ \varphi'(p) = \varphi_x^{-1} \circ \tau_t(\sigma(x)) \circ \varphi(p) \\
&= \varphi_x^{-1} \circ \mathrm{Exp}(tH)(\sigma(x)) \circ \varphi(p) \qquad (31)
\end{aligned}$$

in the given trivialization. Hence, from the one-parameter subgroup, the induced Killing vector field can easily be obtained:

$$\begin{aligned}
V^H(p) &\equiv \frac{d}{dt}\mathscr{F}_t(p)\Big|_{t=0} \\
&= \frac{d}{dt}\varphi_x^{-1} \circ \mathrm{Exp}(tH)(\sigma(x)) \circ \varphi(p)\Big|_{t=0} \\
&= \varphi_{x*}^{-1} R_{\varphi(p)*} H(\sigma(x)), \qquad (32)
\end{aligned}$$

where $H(\sigma(x))$ belongs to the Lie algebra of G. It is immediately obvious that these Killing vector fields are isomorphic to the set of right invariant vector fields on G. Since $\mathscr{F}$ is an isometry, it follows that the Lie derivative of the metric tensor (4b) with respect to these Killing vector fields must vanish, i.e., $(\mathscr{L}_{V^H}\hat{\mathfrak{g}})(X,Y) = 0$, $\forall X,Y \in \mathfrak{X}(P)$. In particular by assigning

$$\begin{aligned}
X = e_\mu &= R_{g*}\sigma_* \partial_\mu - A_\mu^b(x)\mathrm{Ad}(g^{-1})_b^a \xi_a^* \\
&\equiv R_{g*}\sigma_* \partial_\mu - A_\mu^a(x)\bar{\xi}_a,
\end{aligned}$$

where $\bar{\xi}_a$ is a right-invariant vector field on P defined by

$$\bar{\xi}_a = \varphi_{x*}^{-1}\frac{d}{dt}\exp(t\xi_a)\varphi_j(p)\Big|_{t=0}, \qquad (33)$$

and $Y = \xi_a^*$, the Killing equation yields

$$\hat{\mathfrak{g}}([V^H, e_\mu], \xi_a^*) = 0, \qquad (34)$$

since $\hat{\mathfrak{g}}(e_\mu, \xi_a^*) = 0$ and $[V^H, \xi_a^*] = 0$. By noting that

$$\begin{aligned}
\varphi_* e_\mu &= \varphi_*(R_{g*}\sigma_* \partial_\mu - A_\mu^a(x)\bar{\xi}_a) \\
&= R_{g*}\varphi_*\sigma_* \partial_\mu - A_\mu^a(x)\varphi_*\bar{\xi}_a = -A_\mu^a(x)\varphi_*\bar{\xi}_a
\end{aligned}$$

[since $\varphi \circ \sigma(x) = e$, $\forall x \in U$] and that $[V^H, e_\mu]$ is vertical it follows that Eq. (34) reduces to

$$-\sum_i k_i(x) A_\mu^b(x) (L_{\phi_{ij}(x)} \cdot \varphi_j)^* \mathfrak{h}([V^H, \bar{\xi}_b], \xi_a^*) = 0, \qquad (35a)$$

or equivalently,

$$-A_\mu^b(x)\hat{\mathfrak{g}}([V^H, \bar{\xi}_b], \xi_a^*) = 0, \qquad (35b)$$

where A is the potential one-form in the gauge specified by φ_j. (Here we have made use of the fact that $\varphi_{j*}[V^H, e_\mu] = [\varphi_{j*}V^H, \varphi_{j*}e_\mu]$.) Since the commutator $[V^H, \bar{\xi}_b]$ does not vanish and is vertical in general, Eq. (35b) clearly indicates a constraint on the gauge fields. In fact, this is not surprising at all. Indeed a gauge-invariant metric effectively requires the metric to be invariant under coordinate changes, since gauge transformations are really a special kind of coordinate transformation. Now on this consideration alone one would anticipate some form of constraints since a metric on any manifold is generally not required to be invariant but rather covariant with respect to coordinate changes. This naturally raises the question of its necessity. Recall that the only reason for requiring it in the first place was to ensure a gauge-invariant theory but this, as we have seen, leads to the inadmissible constraint (35). Now, it is well known that the action

$$I = \int_P d^{n+4}x\sqrt{-\hat{g}}\,\hat{R}, \qquad (36)$$

where $\hat{R}$ is the scalar curvature of the space P, is invariant under all basis transformations. A viable solution, then, would be to consider the conditions under which the above transformations can be regarded as basis transformations. In the following we will demonstrate that this is possible if the metric $\mathfrak{h}$ is right invariant. We would like to remark that this is a sufficient condition rather than a necessary one. Nevertheless, we will find that it will not only lead to a gauge-invariant theory, but also prove to be important in the context of the consistency problem as will be discussed in Sec. IV.

Since the action (36) is invariant under all basis transformations, it is sufficient to work in a particular basis. For convenience, we will use the horizontal lift basis, comprised of $\{e_\mu\}$ for the horizontal subspaces and the set of right-invariant vector fields $\{\bar{\xi}_a(p)\}$ for the vertical subspaces. Since the horizontal vectors $\{e_\mu\}$ are trivialization independent, they are unaffected by gauge transformations and we only need to consider the effect of these changes on the vertical fields. Now, under a general basis transformation $\bar{\xi}_a(p) \to \bar{\xi}'_a(p) = \Lambda_a^b(p)\bar{\xi}_b(p)$ the components of a metric tensor transform as

$$g_{ab}(p) \to g'_{ab}(p) = \Lambda_a^c(p)\Lambda_b^d(p)g_{cd}(p)\,, \tag{37}$$

where $\Lambda_a^c(p)$ are smooth functions. Let us see whether the components of the metric (4b) can be expressed in this form under a gauge transformation.

Consider a change in the trivialization $\varphi_j \to \varphi'_j = L_{\gamma_j(x)} \circ \varphi_j$, where $\gamma_j(x) = \tau(\sigma_j(x))$ [see Eq. (22)]. Under such a change, the components of $\hat{\mathfrak{g}}$, $g_{ab}(p) = \hat{\mathfrak{g}}(\bar{\xi}_a, \bar{\xi}_b)(p)$ transform as $g_{ab}(p) \to g'_{ab}(p)$ with

$$g'_{ab}(p) = \sum_i k_i(x)(L_{\phi_{ij}(x)} \circ L_{\gamma_j(x)} \circ \varphi_j)^*\mathfrak{h}(\bar{\xi}_a, \bar{\xi}_b)(p) \tag{38a}$$

[note that $\phi_{ij}(x)$ remains unchanged]. Since we are evaluating the metric at the point $p \in P$, we can write

$$\begin{aligned} g'_{ab}(p) &= \sum_i k_i(x)((L_{\phi_{ij}(x)} \circ L_{\gamma_j(x)} \circ \varphi_j)^*\mathfrak{h})_p(\bar{\xi}_a(p), \bar{\xi}_b(p)) \\ &= \sum_i k_i(x)\mathfrak{h}_{g'_i}(L_{\phi_{ij}}(x)_* L_{\gamma_j(x)*} \varphi_{j*} \bar{\xi}_a(p), L_{\phi_{ij}(x)*} \\ &\quad \times L_{\gamma_j(x)*} \varphi_{j*} \bar{\xi}_b(p))\,, \end{aligned} \tag{38b}$$

where $g'_i = L_{\phi_{ij}(x)} \circ L_{\gamma_j(x)} \circ \varphi_j(p) \in G$. Before proceeding further, let us evaluate the term $L_{\gamma_j(x)*} \varphi_{j*} \bar{\xi}_a(p)$ explicitly. Since $\varphi_{j*} \bar{\xi}_a(p) = e_a^L(\varphi_j(p))$ (a right-invariant vector field on G), we have

$$\begin{aligned} L_{\gamma_j(x)*} e_a^L(\varphi_j(p)) &= L_{\gamma_j(x)*} \frac{d}{dt} \exp(t\xi_a)\varphi_j(p)\Big|_{t=0} \\ &= \frac{d}{dt}\gamma_j(x)\exp(t\xi_a)\gamma_j(x)^{-1}\gamma_j(x)\varphi_j(p)\Big|_{t=0} \\ &= \mathrm{Ad}(\gamma_j(x))_a^b R_{\varphi_j(p)^{-1}\gamma_j(x)\varphi_j(p)*} e_b^L(\varphi_j(p))\,. \end{aligned} \tag{39}$$

Substituting into Eq. (38b), we have

$$\begin{aligned} g'_{ab}(p) = \sum_i k_i(x) &\mathrm{Ad}(\gamma_j(x))_a^c \mathrm{Ad}(\gamma_j(x))_b^d \\ &\times R^*_{\varphi_j(p)^{-1}\gamma_j(x)\varphi_j(p)} \mathfrak{h}_{g''_i}(e'_c, e'_d)\,, \end{aligned} \tag{40}$$

where $g''_i = L_{\phi_{ij}(x)} \circ \varphi_j(p)$ and $e'_c = L_{\phi_{ij}(x)*} \varphi_{j*} \bar{\xi}_c(p)$. It is obvious from (40) that if $\mathfrak{h}$ is right invariant then the metric transforms as

$$g'_{ab}(p) = \mathrm{Ad}(\gamma_j(x))_a^c\ \mathrm{Ad}(\gamma(x))_b^d g_{cd}(p)\,, \tag{41}$$

which is similar to Eq. (37) with $\Lambda_a^c = \mathrm{Ad}(\gamma(x))_a^c$. If we had used the fundamental vector fields $\{\xi_a^*\}$ as a basis for the vertical subspaces instead, then it is not difficult to show that the transformations are equivalent to a change,

$$\xi_a^*(p) \to \xi_a'^*(p) = \mathrm{Ad}(\varphi_j(p)^{-1}\gamma_j(x)\varphi_j(p))_a^b \xi_b^*(p)\,. \tag{42}$$

Unlike the previous case, the matrix elements Λ_a^b here depend on p as well. To summarize briefly, we have shown that the effect of gauge transformations on the metric can be realized as basis transformations if the metric $\mathfrak{h}$ is right invariant. In fact with a right-invariant $\mathfrak{h}$, the right action $R_g: P \to P$ for each $g \in G$, on the principal bundle, becomes an isometry. This can be shown by noting that

$$\pi \circ R_g(p) = \pi(p), \quad \forall p \in P, \quad g \in G\,,$$
$$\varphi_j \circ R_g(p) = R_g \circ \varphi_j(p)\,.$$

Then we have, for each $g \in G$,

$$\begin{aligned} R_g^*\hat{\mathfrak{g}}(V,W) &= R_g^*\pi^*\mathfrak{g}_M(V,W) + \sum_i k_i(x) \\ &\quad \times R_g^*(L_{\phi_{ij}(x)} \circ \varphi_j)^*\mathfrak{h}(V - \hat{V}, W - \hat{W}) \\ &= \pi^*\mathfrak{g}_M(V,W) + \sum_i k_i(x)(L_{\phi_{ij}(x)} \circ \varphi_j)^* \\ &\quad \times R_g^*\mathfrak{h}(V - \hat{V}, W - \hat{W}) \\ &= \hat{\mathfrak{g}}(V,W)\,. \end{aligned} \tag{43}$$

It is easy to verify that this isometry does not lead to any constraints.

IV. GLOBAL ACTION AND CONSISTENCY

Having discussed the general setting of the fields in geometric terms, which to some extent summarizes the kinematics of these fields, we now turn to their dynamics. This is primarily governed by the Lagrangian that not only determines the field equations via the variational principle but also restricts the possible field configurations. From the outset, one would expect the theory to be mathematically consistent, in the sense that the choice of the fields on which the assumed theory is built should not lead to any field equations that are inconsistent. For a realistic theory it is also natural to require that the theory be compatible with phenomenological observations at the low-energy limit. In Refs. 9 and 10, it has been shown that the effective low-energy theory, which is usually obtained by truncating the massive states, may or may not be consistent. In the context of the Kaluza–Klein theory, this means that the fields corresponding to the massless modes, which arise together with the massive ones in the harmonic expansions, should be consistent with the higher-dimensional field equations. However, Duff and others[11] have pointed out that the Kaluza–Klein *Ansätze*, which corresponds to the zero modes of the expansions, suffers from this malady. Indeed, the *Ansätze* are generally inconsistent with the $(4+n)$-dimensional field equations. To circumvent this, one has to restrict the fields to those that are invariant under a group that acts transitively on the internal manifold. This G-invariant scheme requires the vacuum state, $M_4 \times S$, to be a solution of the field equations. (Here M_4 and S denote the four-dimensional space-time and internal manifold, respectively.) In the bundle approach, one can relax this by requiring the presence of a global isometry that is also vertical. Here dimensional reduction is achieved through isometries rather than truncations and, as we will demonstrate in the following, this will lead to consistent field equations. In particular, the equations of motion,

$$\hat{R}_{AB} = 2\Lambda/(n+2)\hat{g}_{AB} ,$$

which are Einstein's equations (with a cosmological constant Λ) will be shown to be independent of the fiber coordinates.

Essentially, the scheme requires the existence of an action that is both global and transitive on the fibers. On the principal bundle, we have two such actions; the usual right action by the structure group and the vertical automorphisms. In the latter case, we have shown that when these are regarded as isometries, they impose some severe constraints on the gauge fields and are thus inadmissible. Let us consider the right action which, as we have seen, ensures a gauge-invariant theory.

Now, if the linear connection on P is Riemannian, then under the isometry $R_g: P \to P$, we have[14]

$$R_{g*}\nabla_X Y = \nabla_{R_{g*}X} R_{g*} Y, \tag{44}$$

for all vector fields X, Y on P. From this it follows that the curvature tensor defined by $\mathfrak{R}(X,Y)Z = [\nabla_X,\nabla_Y]Z - \nabla_{[X,Y]}Z$ transforms as

$$R_{g*}\mathfrak{R}(X,Y)Z = \mathfrak{R}(R_{g*}X,R_{g*}Y)R_{g*}Z, \quad \forall X,Y,Z \in \mathfrak{X}(P) . \tag{45}$$

It can be noted that if X, Y, and Z are right-invariant vector fields, then $\mathfrak{R}(X,Y)Z$ is also a right-invariant field since

$$R_{g*}\mathfrak{R}(X,Y)Z = \mathfrak{R}(R_{g*}X,R_{g*}Y)R_{g*}Z = \mathfrak{R}(X,Y)Z . \tag{46}$$

Now if the chosen basis is composed of right-invariant vector fields [which can always be found on $\pi^{-1}(U)$ in a given trivialization], then the components of the curvature tensor in this basis are independent of the fiber coordinates. We can demonstrate this as follows. Consider the basis $\{X_A\}$ with vector fields that satisfy $R_{g*}X_A = X_A$, $\forall g \in G$. Then because of right invariance, they commute with the fundamental vector fields as the latter generate right actions on P. Since $\mathfrak{R}(X_A,X_B)X_C$ is a right-invariant vector field, we have

$$[\xi^*,\mathfrak{R}(X_A,X_B)X_C] = 0 \tag{47}$$

for any fundamental vector field ξ^*. In the component form with $R^{\ D}_{CAB}X_D = \mathfrak{R}(X_A,X_B)X_C$, this yields

$$\begin{aligned}[][\xi^*,R^{\ D}_{CAB}(p)X_D] \\ &= \xi^*(R^{\ D}_{CAB}(p))X_D + R^{\ D}_{CAB}(p)[\xi^*,X_D] = 0 \end{aligned}\tag{48}$$

and hence

$$\xi^* R^{\ D}_{CAB} = 0 . \tag{49}$$

This shows that the components are indeed independent of the fiber coordinates. Similarly, the components of the Ricci tensor, which are obtained by contracting the indices A and D in (49), are also independent of the fiber coordinates.

Now, let us consider a right action on the associated bundle $(E,M,\pi_E,G/H,G,P)$. Since a point on E is characterized by a representative from a class of elements of the form (p,y), which are equivalent under the right action defined by (6), there are two ways in which we can transform a point from one equivalent class to another. We can achieve this by either applying $g \in G$ to $p \in P$ or $y \in G/H$ of the original class. However, both of these actions are not gauge invariant in the sense that they depend on the choice of the class representative. Thus in general we can say that the action induced by a right action on P or a left action on G/H is not well defined.

Next let us look at the automorphisms. Here unlike the above case, a global action on E can be induced in a gauge-invariant way. Indeed the action $\overline{\mathscr{F}}: E \to E$ given by

$$\overline{\mathscr{F}}(w) = \mathscr{F}(p)\cdot y , \tag{50}$$

where $\mathscr{F}: P \to P$ and $w = p\cdot y$, satisfies this criterion. These automorphisms, which are vertical since

$$\pi_E(\overline{\mathscr{F}}(w)) = \pi_E(\mathscr{F}(p)\cdot y) = \pi(\mathscr{F}(p)) = \pi(p) = \pi_E(w) , \tag{51}$$

also form a group that is isomorphic to $GA(P)$. This can be shown as follows: If $\overline{\mathscr{F}}_1$ and $\overline{\mathscr{F}}_2$ are two such automorphisms corresponding to $\mathscr{F}_1$ and $\mathscr{F}_2$ on P, then

$$\begin{aligned}\overline{\mathscr{F}}_1\circ\overline{\mathscr{F}}_2(w) &= \overline{\mathscr{F}}_1(\mathscr{F}_2(p)\cdot y) = \mathscr{F}_1(\mathscr{F}_2(p))\cdot y \\ &= \mathscr{F}_1\circ\mathscr{F}_2(p)\cdot y, \end{aligned}\tag{52}$$

where $\mathscr{F}_1\circ\mathscr{F}_2 \in GA(P)$. Furthermore, they are transitive on the fibers of E since the action of $\mathscr{F}$ that induces $\overline{\mathscr{F}}$ is transitive on the fibers of P. To implement these isometries, the metric on E has to be restricted. This is clear from our earlier discussion of gauge-invariant metrics. In analogy with (28), it is easy to show that the automorphisms $\overline{\mathscr{F}}: E \to E$ become isometries if the metric (7) is gauge invariant. Indeed from Eqs. (50), (8), and (21) we have

$$\begin{aligned}\overline{\varphi}_i(\overline{\mathscr{F}}(w)) &= \overline{\varphi}_i(\mathscr{F}(p)\cdot y) = \vartheta\circ\varphi_i(\mathscr{F}(p)) \\ &= \vartheta\circ\varphi'_i(p) = \overline{\varphi}'_i(w) , \end{aligned}\tag{53}$$

from which the above claim follows. Now the metric (7) is gauge invariant if $\mathfrak{h}_{G/H}$ is G invariant since under a gauge transformation $\varphi \to \varphi' = L_{a(x)}\varphi$ the corresponding trivialization $\overline{\varphi}$ on E transforms as

$$\begin{aligned}\overline{\varphi}(w) \to \overline{\varphi}'(w) &= \overline{\varphi}'\circ\mu(p) = \vartheta\circ\varphi'(p) \\ &= \vartheta\circ L_{a(x)}\circ\varphi(p) = \overline{L}_{a(x)}\circ\vartheta\circ\varphi(p) \\ &= \overline{L}_{a(x)}\circ\overline{\varphi}(w) , \end{aligned}\tag{54}$$

where we have made use of Eq. (11). From (54) and (7) it is clear that if the metric $\mathfrak{h}_{G/H}$ is G invariant then $\hat{\mathfrak{g}}_E$ remains invariant under gauge transformations. As most of the results here are similar to those obtained earlier for the principal bundle, it appears likely that the Killing vector fields that generate $\overline{\mathscr{F}}$ will also lead to the same constraints on the gauge fields. This is precisely what we obtain, if we follow the calculations of (34) and (35). It is evident, then, that both the right actions and the automorphisms on the associated bundle $(E,M,\pi_E,G/H,G,P)$ fail in one way or another. Nevertheless, we will show in the following section that the problem of implementing a global action can be resolved by considering the bundle $(E,M,\pi_E,G/H,G\times N(H)/H,P_1\circ P_2)$ instead. This fiber bundle that is associated to the principal bundle $(P_1\circ P_2,M,\pi,G\times N(H)/H)$ [obtained by splicing two bundles (P_1,M,π_1,G) and $(P_2,M,\pi_2,N(H)/H)$] allows for global automorphisms with nonvanishing gauge potentials.

V. BUNDLE SPLICING

One of the ways in which the fiber bundle approach differs from the usual formulation is that the gauge potentials are valued in the subalgebra of the full isometry group. (Here the full isometry group refers to that of the internal space.) To elaborate further, if the internal space is diffeomorphic to a coset manifold G/H, then the full isometry group that one can have is not G, but rather $G\times N(H)/H$.[23] This is because the manifold can be transformed by either a left multiplication by G or right multiplication by $N(H)/H$ since both actions are defined. So, essentially, the metric should incorporate gauge fields belonging to the Lie algebra of $G\times N(H)/H$. In the bundle approach, however, as we have seen, one does not obtain this accord. Indeed, the gauge potentials are valued only in the Lie algebra of G in the case of the bundle $(E,M,\pi_E,G/H,G,P)$.

In this section we will construct a fiber bundle with typical fiber G/H, which incorporates the full isometry group, i.e., $G\times N(H)/H$, as the gauge group.

Now, a natural way to introduce the full gauge group is to "splice" two principal bundles (P_1,M,π_1,G_1) and (P_2,M,π_2,G_2) together to form a new principal bundle $(P_1\circ P_2,M,\pi,G_1\times G_2)$ whose structure group is the direct product of G_1 and G_2. Formally this can be defined as follows[20]: If (P_1,M,π_1,G_1) and (P_2,M,π_2,G_2) are two principal fiber bundles then the set

$$P_1\circ P_2=\{(p_1,p_2)\in P_1\times P_2\,|\,\pi_1(p_1)=\pi_2(p_2)\} \tag{55}$$

with a right action of $G_1\times G_2$ defined by $(p_1,p_2)(g_1,g_2)=(p_1g_1,p_2g_2)$, can be regarded as the bundle space of a fiber bundle with base space $M=P_1\circ P_2/(G_1\times G_2)$ and structure group $G_1\times G_2$. Now, the existence of local trivializations ψ_1: $\pi_1^{-1}(U)\to U\times G_1$, ψ_2: $\pi_2^{-1}(U)\to U\times G_2$ on the two principal bundles gives rise to a trivialization ψ: $\pi^{-1}(U)\to U\times(G_1\times G_2)$ on $(P_1\circ P_2,M,\pi,G_1\times G_2)$ through

$$\psi(p_1,p_2)=(\pi(p_1,p_2),\,\varphi(p_1,p_2)),\quad \forall(p_1,p_2)\in\pi^{-1}(U)\,, \tag{56a}$$

where

$$\varphi(p_1,p_2)=(\varphi_1(p_1),\varphi_2(p_2))\,. \tag{56b}$$

Associated to this trivialization is a local section σ: $U\to P_1\circ P_2$, which can be written as

$$\begin{aligned}\sigma(x)=(p_1,p_2)\,\varphi(p_1,p_2)^{-1}&=(p_1\varphi_1(p_1)^{-1},p_2\varphi(p_2)^{-1})\\&=(\sigma_1(x),\sigma_2(x))\,,\end{aligned} \tag{57}$$

where $\sigma_1(x)$ and $\sigma_2(x)$ are the corresponding local sections on (P_1,M,π_1,G_1) and (P_2,M,π_2,G_2), respectively.

Furthermore, the projection π^1: $P_1\circ P_2\to P_1$ defined by $\pi^1(p_1,p_2)=p_1$ can be regarded as the bundle projection of a principal fiber bundle $(P_1\circ P_2,P_1,\pi^1,G_2)$, with a right action by the group $(\{e\}\times G_2)\approx G_2$ and the base manifold $P_1=P_1\circ P_2/(\{e\}\times G_2)$. Similarly $(P_1\circ P_2,P_2,\pi^2,G_1)$, where $\pi^2(p_1,p_2)=p_2$ is also a principal fiber bundle.

Now let us construct a fiber bundle that is associated to this spliced bundle. Recall from Sec. II that in order to do so, we must first identify the typical fiber and also specify the group action on it. In the present case we will take F to be the coset manifold G/H with the left action ρ: $G/H\to G/H$ defined as

$$\rho(a,b)y=ayb^{-1},\quad \forall(a,b)\in G\times N(H)/H,\quad y\in G/H\,; \tag{58}$$

it is easy to see that the group $G\times N(H)/H$ is transitive on G/H. Then by observing that the isotropy subgroup at the origin is

$$\Delta N=\{(n_1,n_2)\in G\times N(H)/H\,|\,n_1=n_2\}\,, \tag{59}$$

we can identify G/H by the coset $(G\times N(H)/H)/\Delta N$ with a projection map ϑ: $G\times N(H)/H\to G/H$ given by

$$\vartheta:\ (a,b)\to ab^{-1},\quad \forall(a,b)\in G\times N(H)/H\,. \tag{60}$$

From (58) and (60) we also have the relation

$$\vartheta\circ L_{(a,b)}=\rho(a,b)\circ\vartheta,\quad \forall(a,b)\in G\times N(H)/H\,, \tag{61}$$

where $L_{(a,b)}$ denotes the left action of $G\times N(H)/H$ on itself.

With definition (58) we can now construct the associated bundle as

$$E=(P_1\circ P_2)\times_{(G\times N(H)/H)}G/H=P_1\circ P_2/\Delta N\,. \tag{62}$$

The vertical automorphisms we discussed earlier can also be defined here. However, we must first introduce them on the spliced bundle $P_1\circ P_2$, which we do so by setting

$$\mathscr{F}(p_1,p_2)=(\mathscr{F}_1(p_2),\mathscr{F}_2(p_2))\,, \tag{63}$$

where $\mathscr{F}_1$: $P_1\to P_1$ and $\mathscr{F}_2$: $P_2\to P_2$ are the automorphisms on (P_1,M,π_1,G) and $(P_2,M,\pi_2,N(H)/H)$, respectively. It is easy to verify that (63) satisfies both requirements, (18a) and (18b). It follows, then, that the group of gauge transformations on $P_1\circ P_2$ is the direct product of $GA(P_1)$ and $GA(P_2)$, i.e., $GA(P_1\circ P_2)=GA(P_1)\times GA(P_2)$. Now consider the automorphisms on E, given by (50), that correspond to the subgroup $\{GA(P_1)\times e\}$, where e is the identity of $GA(P_2)$. It is interesting to note that these automorphisms are also transitive on the fibers. Since we are looking for isometries with this property, the automorphisms belonging to the above subgroup appear well suited for the purpose. Now, it is obvious from our discussion on gauge-invariant metrics that some restrictions will be required in implementing this isometry. First, let us look at the restrictions required on the metric $\mathfrak{h}_{G/H}$. To realize the automorphisms $\mathscr{F}$: $E\to E$, we can write, using (8), (22), and (61),

$$\begin{aligned}\bar\varphi_i'(w)&=\bar\varphi_i'\circ\mu(p)=\vartheta\circ\varphi_i'(p)\\&=\vartheta\circ L_{a_i(x)}\circ\varphi_i(p)\\&=\rho(a_i(x))\vartheta\circ\varphi_i(p)\\&=\rho(a_i(x))\bar\varphi_i(w)\,,\end{aligned} \tag{64}$$

where $a_i(x)=\tau(\sigma_i(x))=(\tau_1(\sigma_{1,i}(x)),\,\tau_2(\sigma_{2,i}(x)))$. [Here τ_1: $P_1\to G$ and τ_2: $P_2\to N(H)/H$ are the realizations of the gauge transformations and $\sigma_{1,i}(x)$, $\sigma_{2,i}(x)$ are the local sections on (P_1,M,π_1,G) and $(P_2,M,\pi_2,N(H)/H)$, respectively.] Then using (53) and (58) we have

$$\begin{aligned}\bar\varphi_i(\mathscr{F}(w))&=\rho(\tau_1(\sigma_{1,i}(x)),\tau_2(\sigma_{2,i}(x)))\bar\varphi_i(w)\\&=\bar L_{\tau_1(\sigma_{1,i}(x))}\circ\bar R_{\tau_2(\sigma_{2,i}(x))^{-1}}\bar\varphi_i(w)\,.\end{aligned} \tag{65}$$

If we restrict the automorphisms to the subgroup $\{GA(P_1)\times e\}$ then $\tau_2(\sigma_{2,i}(x))=e$ and it becomes clear that if these automorphisms are to be isometries then the metric $\mathfrak{h}_{G/H}$ must be invariant under left actions by G or in other words, it must be G invariant. Note that this is similar to the case of the associated bundle $(E,M,\pi_E,G/H,G,P)$ we discussed in the previous section. One may wonder whether these isometries will still allow an x-dependent metric on the vertical subspaces since in the earlier case this x dependence vanished when the metric on the typical fiber was taken to be left invariant. In the following, we will show that if the metric is not invariant under right actions by $N(H)/H$ then it can still retain the x dependence. Following (64), we have

$$\bar{\varphi}_i(w)=\rho(\phi_{ij}^{(1)}(x),\phi_{ij}^{(2)}(x))\bar{\varphi}_j(w)$$
$$=\bar{L}_{\phi_{ij}^{(1)}(x)}\circ\bar{R}_{\phi_{ij}^{(2)}(x)^{-1}}\bar{\varphi}_j(w)\,, \qquad (66)$$

where $\phi_{ij}^{(1)}$: $U_i\cap U_j\to G$ and $\phi_{ij}^{(2)}$ $U_i\cap U_j\to N(H)/H$ are the transition functions on (P_1,M,π_1,G) and $(P_2,M,\pi_2,N(H)/H)$, respectively. Then the metric tensor on the vertical subspaces can be written as

$$\sum_i k_i(x)\varphi_i^*\mathfrak{h}_{G/H}=\sum_i k_i(x)(\bar{L}_{\phi_{ij}^{(1)}(x)}\circ\bar{R}^{-1}_{\phi_{ij}^{(2)}}\circ\bar{\varphi}_j)^*\mathfrak{h}_{G/H}$$
$$=\sum_i k_i(x)(\bar{R}^{-1}_{\phi_{ij}^{(2)}(x)}\circ\varphi_j)^*\mathfrak{h}_{G/H}\,, \qquad (67)$$

where we have assumed that $\mathfrak{h}_{G/H}$ is G invariant. Hence it is obvious from (67) that $\bar{R}$: $G/H\to G/H$ should not be an isometry if x dependence is to be retained. It was also shown in Sec. III that a gauge-invariant metric implies constraints on the gauge potentials. Since we are now demanding partial gauge invariance, it is inevitable that some constraints will arise. However, before considering this let us construct the horizontal and vertical vector fields on E.

Now if $e_a^R\in\mathfrak{X}(G)$ is a left-invariant vector field on G then its lift $\tilde{e}_a^R$ to $G\times N(H)/H$ is given by

$$\tilde{e}_a^R(g_1,g_2)=\frac{d}{dt}(g_1\exp t\xi_a,g_2)\Big|_{t=0}\,, \qquad (68a)$$

where $\xi_a\in T_e(G)$. Similarly the lift of a left-invariant vector field on $N(H)/H$ is

$$\tilde{e}_\alpha^R(g_1,g_2)=\frac{d}{dt}(g_1,g_2\exp t\xi_\alpha)\Big|_{t=0}\,. \qquad (68b)$$

[In the following the indices $a,b,\ldots$ and $\alpha,\beta,\ldots$ will be used to denote the subspaces $T_e(G)$ and $T_e(N(H)/H)$, respectively.] Then from Eqs. (68a) and (68b), we can define the fundamental vector fields on $P_1\circ P_2$ by

$$\xi_a^*(p_1,p_2)=\varphi_{x*}^{-1}\tilde{e}_a^R(\varphi_1(p_1),\varphi_2(p_2))$$
$$=\frac{d}{dt}(p_1\exp t\xi_a,p_2)\Big|_{t=0}\,, \qquad (69a)$$

$$\xi_\alpha^*(p_1,p_2)=\varphi_{x*}^{-1}\tilde{e}_\alpha^R(\varphi_1(p_1),\varphi_2(p_2))$$
$$=\frac{d}{dt}(p_1,p_2\exp t\xi_\alpha)\Big|_{t=0}\,. \qquad (69b)$$

Although these vector fields are expressed by using a trivialization φ, they are, however, independent of the gauge choice. This can be verified by choosing a different trivialization $\varphi'=L_{\gamma(x)}\circ\varphi$.

To construct the horizontal vector fields we must first define a connection form on the principal bundle. Since both π^1 and π^2 are bundle projections they are, by definition, smooth and the connection form ω on $(P_1\circ P_2,M,\pi,G\times N(H)/H)$ can be introduced canonically through[20]

$$\omega=\pi^{1*}\omega_1+\pi^{2*}\omega_2\,, \qquad (70)$$

where ω_1 and ω_2 are the connection forms on (P_1,M,π_1,G) and $(P_2,M,\pi_2,N(H)/H)$, respectively. Now consider the image of the vector field ∂_μ on $U\subset M$ under the map σ': $U\to P_1\circ P_2$ given by

$$\sigma'(x)=R_{(g_1,g_2)}\circ\sigma(x)\,, \qquad (71)$$

where $\sigma(x)$ is the preferred section. Then from (70) we have

$$\begin{aligned}\omega(\sigma'_*\partial_\mu)&\\&=(\pi^{1*}\omega_1+\pi^{2*}\omega_2)(R_{(g_1,g_2)*}\sigma_*\partial_\mu)\\&=\omega_1(R_{g_1*}\sigma_{1*}\partial_\mu)+\omega_2(R_{g_2*}\sigma_*\partial_\mu)\\&=\mathrm{Ad}(g_1^{-1})_a^bA_\mu^a(x)\xi_b+\mathrm{Ad}(g_2^{-1})_\alpha^\beta\tilde{A}_\mu^\alpha(x)\xi_\beta\\&=(\pi^{1*}\omega_1+\pi^{2*}\omega_2)(\mathrm{Ad}(g_1^{-1})_a^bA_\mu^a(x)\xi_b^*\\&\quad+\mathrm{Ad}(g_2^{-1})_\alpha^\beta\tilde{A}_\mu^\alpha(x)\xi_\beta^*)\\&=\omega(\mathrm{Ad}(g_1^{-1})_a^bA_\mu^a(x)\xi_b^*+\mathrm{Ad}(g_2^{-1})_\alpha^\beta\tilde{A}_\mu^\alpha(x)\xi_\beta^*)\,,\end{aligned} \qquad (72)$$

where $(g_1,g_2)=\varphi(p_1,p_2)$, $A=\sigma_1^*\omega_1$, and $\tilde{A}=\sigma_2^*\omega_2$. It follows that the horizontal lift e_μ of ∂_μ can be expressed as

$$e_\mu(p_1,p_2)=R_{\varphi(p_1,p_2)*}\sigma_*\partial_\mu-\mathrm{Ad}(\varphi_1(p_1)^{-1})_a^bA_\mu^a(x)\xi_b^*$$
$$-\mathrm{Ad}(\varphi_2(p_2)^{-1})_\alpha^\beta\tilde{A}_\mu^\alpha(x)\xi_\beta^* \qquad (73a)$$

or, in terms of the right-invariant vector fields,

$$e_\mu(p_1,p_2)=R_{\varphi(p_1,p_2)*}\sigma_*\partial_\mu-A_\mu^a(x)\bar{\xi}_a-\tilde{A}_\mu^\alpha(x)\bar{\xi}_\alpha\,, \qquad (73b)$$

where

$$\bar{\xi}_a(p_1,p_2)=\varphi_{x*}^{-1}\tilde{e}_a^L(\varphi_1(p_1),\varphi_2(p_2))$$
$$=\varphi_{x*}^{-1}\frac{d}{dt}((\exp t\xi_a)\varphi_1(p_1),\varphi_2(p_2))\Big|_{t=0} \qquad (74a)$$

and

$$\bar{\xi}_\alpha(p_1,p_2)=\varphi_{x*}^{-1}\tilde{e}_\alpha^L(\varphi_1(p_1),\varphi_2(p_2))$$
$$=\varphi_{x*}^{-1}\frac{d}{dt}(\varphi_1(p_1),(\exp t\xi_\alpha)\varphi_2(p_2))\Big|_{t=0}\,. \qquad (74b)$$

On E, the corresponding horizontal vector fields are given by $\bar{e}_\mu=\mu_*e_\mu$. These can be evaluated explicitly by noting that from relations (8) and (60) we have

$$\mu_* \bar{\xi}_a(p_1,p_2) = \bar{\varphi}_{x*}^{-1}\vartheta_* \bar{e}_a^L(\varphi_1(p_1),\varphi_2(p_2)) \equiv \bar{\varphi}_{x*}^{-1} K_a(\varphi_1(p_1)\varphi_2(p_2)^{-1}), \tag{75a}$$

$$\mu_* \bar{\xi}_\alpha(p_1,p_2) = \bar{\varphi}_{x*}^{-1}\vartheta_* \bar{e}_\alpha^L(\varphi_1(p_1),\varphi_2(p_2)) \equiv -\bar{\varphi}_{x*}^{-1} K_\alpha(\varphi_1(p_1)\varphi_2(p_2)^{-1}), \tag{75b}$$

and hence

$$\bar{e}_\mu = \bar{\sigma}_* \partial_\mu - A_\mu^a(x)\bar{\varphi}_{x*}^{-1}K_a(\varphi_1(p_1)\varphi_2(p_2)^{-1}) + \tilde{A}_\mu^\alpha(x)\bar{\varphi}_{x*}^{-1}K_\alpha(\varphi_1(p_1)\varphi_2(p_2)^{-1}), \tag{76}$$

where $\bar{\sigma}$: $U \to E$ is the local section on E under the map μ: $P_1 \circ P_2 \to E$. Here the gauge potentials are valued in the Lie algebra of $G \times N(H)/H$.

Now returning to the question of constraints, we know that they become manifest when one considers the Killing equation, $(\mathscr{L}_Z \hat{g}_E)(X,Y) = 0$, $\forall X,Y \in \mathscr{X}(E)$. With $X = \bar{e}_\mu$, $Y = \mu_* \xi_a^* \equiv \bar{K}_a$, and $Z = V^H$, where

$$V^H(w) = \frac{d}{dt}\bar{\mathscr{F}}_t\Big|_{t=0} = \bar{\varphi}_{x*}^{-1}K_a(\bar{\varphi}(w)) \tag{77}$$

is a Killing vector field generating a one-parameter subgroup of $\{GA(P_1) \times e\}$, the Killing equation yields

$$\hat{g}_E([V^H,\bar{e}_\mu],\bar{K}_a) = 0. \tag{78}$$

By evaluating the commutator (see the Appendix), Eq. (78) reduces to

$$A_\mu^b(x)\hat{g}_E([V^H,\bar{\varphi}_{x*}^{-1}K_b],\bar{K}_a) = 0, \tag{79}$$

which shows that one must set $A_\mu^b(x)$ to zero if $\{GA(P_1) \times e\}$ is to be implemented as an isometry group. Essentially, this requires the gauge group to be restricted to $N(H)/H$. When we substitute $X = \bar{K}_a$ and $Y = \bar{K}_b$ into the Killing equation we obtain

$$V^H(\hat{g}_E(\bar{K}_a,\bar{K}_b)) = 0, \tag{80}$$

and this implies that the components of the metric tensor for the vertical subspaces are independent of the fiber coordinates in the basis $\{\bar{K}_a\}$, $a = 1,2,\dots,\dim(G/H)$. Similarly by choosing $X = \bar{e}_\mu$ and $Y = \bar{e}_\nu$ [with $A_\mu^a(x) = 0$], the components of the metric tensor are also independent of the internal coordinates. Hence if we choose the vector fields $\{\bar{e}_\mu,\bar{K}_a\}$ [$\mu = 1,\dots,4$, $a = 1,\dots,\dim(G/H)$] as a basis on E then the components of the metric in this basis are purely x dependent. It is worth noting that these vector fields are invariant under the automorphisms that belong to $\{GA(P_1) \times e\}$ and hence if they evaluate the components of the Ricci tensor, these, too, will turn out to be independent of the internal coordinates.

To summarize briefly, we have shown that in defining a global action that is also an isometry on an associated bundle, the gauge potentials are usually constrained. In particular, when the fiber bundle is the associated bundle $(E,M,\pi_E, G/H,G,P)$, in which G acts on the left of G/H, the global action, which consists of vertical automorphisms, is not an isometry unless we set all the gauge potentials to zero. This is clearly unacceptable. To resolve this, we have enlarged the principal bundle by splicing two principal bundles (P_1,M,π_1,G) and $(P_2,M,\pi_2,N(H)/H)$. The resulting associated bundle that has $G \times N(H)/H$ as its structure group admits gauge potentials that are valued in the Lie algebra of this group. However, if we choose $\{GA(P_1) \times e\} \approx GA(P_1)$, which is a subgroup of the full group of gauge transformations, as the isometry group, then we find that only gauge potentials that correspond to $N(H)/H$ survive. To implement this, the metric tensor on G/H is required to be left invariant. Moreover, if scalar fields are to be included, then this metric should not be right invariant. In conclusion we would like to remark that although our gauge group is similar to the G-invariant scheme of Ref. 6, our model differs in some aspects. In particular, the isometry group is $GA(P_1)$ instead of G and the structure group is $G \times N(H)/H$ rather than $N(H)/H$.

APPENDIX: DERIVATION OF EQ. (79)

To obtain Eq. (79) from Eq. (78), we need to evaluate the Lie bracket $[V^H,\bar{e}_\mu]$. By using the explicit expression for $\bar{e}_\mu$, which is given by (76), we have

$$[V^H,\bar{e}_\mu] = [V^H,\mu_* R_{(g_1,g_2)*}\sigma_* \partial_\mu] + \tilde{A}_\mu^\alpha(x)[V^H,\bar{\varphi}_{x*}^{-1}K_\alpha] - A_\mu^a(x)[V^H,\bar{\varphi}_{x*}^{-1}K_a]. \tag{A1}$$

Let us evaluate the first term on the right-hand side of (A1). By using the definition of the Lie bracket, we have

$$[V^H,\mu_* R_{(g_1,g_2)*}\sigma_* \partial_\mu] = \frac{d}{dt}\bar{\mathscr{F}}_{-t*}\mu_* R_{(g_1,g_2)*}\sigma_* \partial_\mu\Big|_{t=0}, \tag{A2}$$

where $\bar{\mathscr{F}}_t$ is the one-parameter subgroup that induces V^H. Then from (65) with $\tau_2(\sigma_{2,i}(x)) = e$, the one-parameter subgroup can be denoted by

$$\bar{\mathscr{F}}_t = \bar{\varphi}_{i,x}^{-1} \circ \bar{L}_{\tau_{1,t}(\sigma_{1,i}(x))} \circ \bar{\varphi}_i. \tag{A3}$$

Substituting (A3) into (A2), and using (8), we have

$$\begin{aligned}[V^H,\mu_* R_{(g_1,g_2)*}\sigma_* \partial_\mu] &= \frac{d}{dt}\bar{\varphi}_{i,x*}^{-1}\bar{L}_{\tau_{1,(-t)}(\sigma_{1,i}(x))*}\bar{\varphi}_{i*}\mu_* R_{(g_1,g_2)*}\sigma_* \partial_\mu\Big|_{t=0} \\ &= \frac{d}{dt}\bar{\varphi}_{i,x*}^{-1}\bar{L}_{\tau_{1,(-t)}(\sigma_{1,i}(x))*}\vartheta_* \varphi_* R_{(g_1,g_2)*}\sigma_* \partial_\mu\Big|_{t=0}. \end{aligned}\tag{A4}$$

By noting that φ commutes with the right action, the bracket vanishes since $\varphi_* \sigma_* \partial_\mu = 0$. This is because all the points on $\sigma(x)$ are mapped to the identity in G by φ.

The second bracket in (A1) also vanishes since

$$\begin{aligned} V^H(w) &= \frac{d}{dt}\bar{\mathscr{F}}_t\Big|_{t=0} \\ &= \frac{d}{dt}\bar{\varphi}_x^{-1} \circ \bar{L}_{\mathrm{Exp}(tH)(\sigma_{1,i}(x))} \circ \bar{\varphi}(w)\Big|_{t=0} \\ &= \bar{\varphi}_{x*}^{-1}\frac{d}{dt}\bar{L}_{\mathrm{Exp}(tH)(\sigma_{1,i}(x))} \circ \bar{\varphi} \circ \mu(p)\Big|_{t=0} \\ &= \bar{\varphi}_{x*}^{-1}\frac{d}{dt}\bar{L}_{\mathrm{Exp}(tH)(\sigma_{1,i}(x))} \circ \vartheta \circ \varphi(p)\Big|_{t=0}, \end{aligned}\tag{A5}$$

where in arriving at the last step we have used Eq. (8), and thus

$$[V^H, \overline{\varphi}_{x*}^{-1} K_\alpha] = -\overline{\varphi}_{x*}^{-1}[\vartheta_*(e^L,0), \vartheta_*(0,e_\alpha^L)]$$
$$= -\overline{\varphi}_{x*}^{-1}\vartheta_*[(e^L,0),(0,e_\alpha^L)]$$
$$=0. \qquad (A6)$$

Finally the last term that remains in (A1) does not vanish since

$$[V^H, \overline{\varphi}_{x*}^{-1} K_\alpha] = \overline{\varphi}_{x*}^{-1}\vartheta_*[(e^L,0),(e_\alpha^L,0)] \neq 0, \qquad (A7)$$

and the result follows.

[1]A. Trautman, Rep. Math. Phys. (Torun) **1**, 29 (1970).
[2]R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968).
[3]L. N. Chang, K. I. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976).
[4]Y. M. Cho, J. Math. Phys. **16**, 2029 (1975).
[5]R. Percacci and S. Randjbar-Daemi, J. Math. Phys. **24**, 807 (1983).
[6]R. Coquereaux and A. Jadczyk, Commun. Math. Phys. **90**, 79 (1983).
[7]A. Salam and J. Strathdee, Ann. Phys. (NY) **141**, 316 (1982).
[8]E. Witten, Nucl. Phys. B **186**, 412 (1981).
[9]M. Duff, "Kaluza–Klein-Konsistency," Preprint CERN-TH. 4296/85, 1985.
[10]M. J. Duff and C. N. Pope, Nucl. Phys. B **255**, 355 (1985).
[11]M. J. Duff, B. E. W. Nilsson, C. N. Pope, and N. P. Warner, Phys. Lett. B **149**, 90 (1984).
[12]C. N. Pope, *Truncations in Kaluza–Klein Theories*, in *Field Theory, Quantum Gravity and Strings*, Proceedings, Mendon and Paris VI, France 1984/85, edited by H. J. de Vega and N. Sanchez (Springer, Berlin, 1986).
[13]B. O'Neill, *Semi-Riemannian Geometry* (Academic, New York, 1983).
[14]S. Kobayshi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 1.
[15]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis Manifolds and Physics* (North-Holland, Amsterdam, 1982).
[16]Y. Matsushima, *Differential Manifolds* (Dekker, New York, 1972).
[17]R. Hermann, Proc. Am. Math. Soc. **11**, 236 (1960).
[18]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, NJ, 1951).
[19]G. H. Thomas, Riv. Nuovo Cimento **3**, 1 (1980).
[20]D. Bleecker, *Gauge Theory and Variational Principles* (Addison–Wesley, Reading, MA, 1981).
[21]M. Daniel and C. M. Viallet, Rev. Mod. Phys. **54**, 175 (1980).
[22]C. J. S. Clarke, "Fields, fibre bundles and gauge groups," in *Mathematical Aspects of Superspace*, edited by H. J. Seifert, C. J. S. Clarke, and A. Rosenblum (Reidel, Boston, 1984).
[23]L. Castellani, L. J. Romans, and N. P. Warner, Ann. Phys. (NY) **157**, 394 (1984).

A simplified model for orbifold compactification

Krzysztof Zabłocki
Institute for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

(Received 1 March 1988; accepted for publication 13 July 1988)

The problem of vacuum stability in orbifold compactification is addressed. The spectrum of fermions and the effective potential in a simple model of compactification on the T^2/Z_2 orbifold with a topologically nontrivial gauge field background is calculated.

I. INTRODUCTION

Recently the extra-dimensional physics has become a standard device in modern particle model building, especially for the superstring inspired models. In spite of increased interest in the field, many important questions of a fundamental nature remain unanswered, especially those involving stability of the vacuum, as well as some problems concerning the existence of chiral zero modes and the index theorem.

A new type of compactification was proposed recently,[1] where the compact space is an orbifold. This type of model is much simpler than that involving compactification on Calabi–Yau manifolds[2] where metrics are hard to find and computation of masses and mixings of the physical spectrum is very complicated[3]; unfortunately it is difficult to obtain realistic low energy models using orbifolds. On the other hand, orbifolds may supply important information about Calabi–Yau compactification after blowing up the singularities of the orbifold.[1] Unfortunately the problems met are still very complicated so a simplified model for investigating compactification on manifolds was proposed by Duncan and Segrè.[4] This model can easily be modified to allow investigations of the properties of the orbifold compactification. The simplicity of the original model is retained after the modification.

We will look for the spectrum of fermions on the T^2/Z_2 orbifold in the presence of a magnetic monopole in the background. We will also find the one-loop contribution of fermions to the effective potential in the massless case. This will allow us to find the spin structure that is preferred in the compactification as a possible vacuum state.

II. THE MODEL

We will remind the reader of the basics of the Duncan–Segrè model[4] and set our conventions, which remain very close to those of Ref. 4. We start from a model in six dimensions containing a Weyl fermion coupled to a U(1) gauge field. Two of the dimensions have compactified with the geometry of $\mathscr{O} = T^2/Z_2$.

The six-dimensional Lagrangian is[4]

$$\mathscr{L}_6 = -\tfrac{1}{4}F_{MN}F^{MN} + i/2(\overline{\Psi}\Gamma^M D_M\Psi - D_M\overline{\Psi}\Gamma^M\Psi), \tag{1}$$

where $D_M = \partial_M - ieA_M$—the covariant derivative, the metric signature is $(-\,+\,+\,+\,+\,+)$, the Dirac matrices Γ^M $(M = 0,1,2,3,5,6)$ are $\Gamma^\mu = \gamma^\mu \otimes 1$ $(\mu = 0,...,3)$, $\Gamma^5 = \gamma^5 \otimes \tau^1$, $\Gamma^6 = \gamma^5 \otimes \tau^2$ where γ are the Dirac matrices in four dimensions and τ are the Pauli matrices. Also, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and $\overline{\Gamma} = \Gamma^0\Gamma^1\Gamma^2\Gamma^3\Gamma^5\Gamma^6 = \gamma^5 \otimes \tau^3$. Fermi fields Ψ fulfill the Weyl condition $\overline{\Gamma}\Psi = +\Psi$, and the directions 5 and 6 are compact, so we set the coordinates $z^M = (x^\mu, y^i)$, $\mu = 0,1,2,3$, $i = 1,2$. The background gauge fields may only occur in compact directions since the four-dimensional space must be Poincaré invariant.

III. CONSTRUCTION OF THE ORBIFOLD

An orbifold $\mathscr{O}$ is formally a quotient of the Euclidean space $\mathbb{R}^n$ over a space group $S = \{(\theta, e^i)\}$ consisting of discrete translations e^i and discrete rotations θ that form the point group P. The action of S may leave several points fixed—these points correspond to the singularities of the orbifold. The action of the space group can be extended to the gauge degrees of freedom by embedding the space group in the gauge group; this way the complete orbifold group is formed.

We will concentrate on $M = \mathbb{R}^2$, and S containing the $\Lambda = \mathbb{Z}\times\mathbb{Z}$ subgroup defining the lattice of the torus with the lattice translations: $g_1 = e^{2\pi i k_1 a_1}$, $g_2 = e^{2\pi i k_2 a_2}$, where a_1 and a_2 are two vectors on the plane generating the torus (with lengths corresponding to the circumferences of the torus) and $P = \mathbb{Z}_2$ acting on M by rotations on π. The orbifold $\mathscr{O}$ can be alternatively defined by the action of $\overline{P} = S/\Lambda$ on T^2 ($\overline{P}$ contains a rotation on π accompanied by a translation on $a_1 + a_2$). Here $\overline{P}$ has four points fixed on T^2 that in our parametrization appear at $(0,0)$, $(0,a_2/2)$, $(a_1/2,0)$, and $(a_1/2,a_2/2)$ (see Fig. 1). There is a conical singularity of deficit angle π at each of them.

IV. THE BOUNDARY CONDITIONS

To set fields on $\mathscr{O}$ we have to impose not only the standard torus boundary conditions but also the constraints arising from identification of the fields at the edge A with that at A' and at B with that at B' (Fig. 1).

For a field Φ on $\mathscr{O}$ we have

$$\Phi(A) = R(\pi)\Phi(A'), \tag{2}$$

where $R(\pi)$ is an operator representing a rotation on π around one of the fixed points.

If a background gauge field is present then the above rotation may be accompanied by a gauge transformation $\mathscr{G}$, as well as some other transformation $\mathscr{S}$ when additional symmetries are present,

$$\Phi(A) = R(\pi)\mathscr{S}\mathscr{G}\Phi(A'). \tag{3}$$

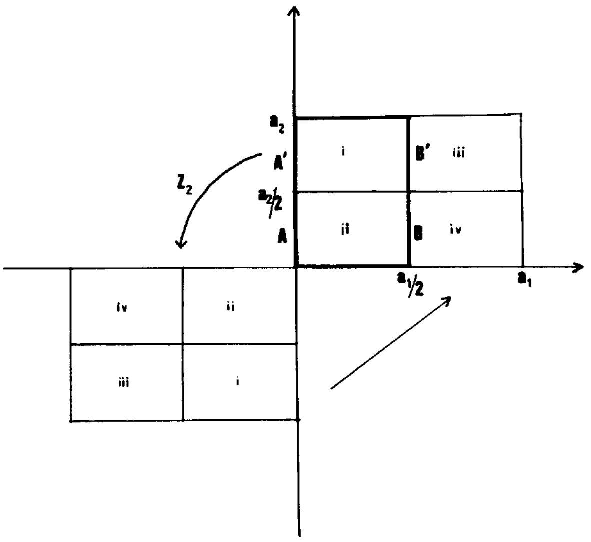


FIG. 1. Parametrization of $\mathcal{O}$ on the plane. The areas i, ii, iii, and iv parametrize T^2. To describe $\mathcal{O}$ area i is identified with iv, and ii with iii.

This case involves twisted boundary conditions for the field Φ when the point group is embedded in the gauge group. For physical purposes it is usually required that Φ is single valued, which will restrict the set of allowed boundary conditions.

V. THE BACKGROUND GAUGE FIELD CONFIGURATION

We will set a U(1) background gauge field with nonvanishing monopole number on $\mathcal{O}$. In order to achieve this, we start from the analogous configuration on T^2 (Ref. 4) and impose the condition $R(\pi)\bar{A} = \bar{A} + d\Lambda$ for every singularity.

Using the gauge freedom it is possible to set for the gauge potential on the $(++d)$ coordinate patch on $\mathcal{O}$,

$$e\bar{A}^{(++d)} = (2\pi n/(a_1a_2/2))y_1\, dy_2, \tag{4}$$

where a_1 and a_2 are circumferences of the torus.

The coordinate patches are

$$(++d),(++u),(+-d),(+-u),(-+d),$$
$$(-+u),(--d),(--u).$$

The $+-$ convention is adopted from the description of the coordinate patches on T^2 (Ref. 3):

$$\begin{aligned}
(++)&:\ 0\leqslant y_1 < a_1/2,\quad 0\leqslant y_2 < a_2,\\
(+-)&:\ 0\leqslant y_1 < a_1/2,\quad 0 < y_2\leqslant a_2,\\
(-+)&:\ 0 < y_1\leqslant a_1/2,\quad 0\leqslant y_2 < a_2,\\
(--)&:\ 0 < y_1\leqslant a_1/2,\quad 0 < y_2\leqslant a_2,
\end{aligned} \tag{5}$$

while u and d indicate that the coordinate patch contains the upper or the lower part of the boundary for $x_1 = 0$ or $x_1 = a_1/2$ [i.e., A' or B' for u and A or B for d (Fig. 1)]. For instance

$$\begin{aligned}
&(++d)\ 0\leqslant x_1 < a_1/2, \text{ for } 0\leqslant x_2 < a_2/2;\\
&\qquad 0 < x_1 < a_1/2, \text{ for } a_2/2\leqslant x_2 < a_2,
\end{aligned} \tag{6}$$

It is easy to find the transition functions (gauge transformations) required to go from one coordinate patch to the other,

$$\begin{aligned}
e\bar{A}^{(\cdots u)} &= e\bar{A}^{(\cdots d)},\\
e\bar{A}^{(-+u \text{ or } d)} &= e\bar{A}^{(++u \text{ or } d)} - d\Lambda,\\
e\bar{A}^{(+-u \text{ or } d)} &= e\bar{A}^{(++u \text{ or } d)},\\
e\bar{A}^{(--u \text{ or } d)} &= e\bar{A}^{(-+u \text{ or } d)},
\end{aligned} \tag{7}$$

where $\Lambda = (2\pi n/a_2)y_2$, n must be integer for the matter fields to be single valued.

Since $R(\pi)\bar{A} = -\bar{A}$, the boundary condition for the edge $A - A'$ is fulfilled in the $(++)$ and $(+-)$ patches $[\bar{A}(A) = \bar{A}(A') = 0]$. The same is true for the edge $B - B'$ in the $(-+)$ and $(--)$ patches. The background gauge field is

$$\bar{F} = (2\pi n/\text{vol})dy_i \wedge dy_j \tag{8}$$

(only the 56 component of the two-dimensional background gauge potential $\bar{F}$ is nonvanishing). Since

$$\frac{1}{2\pi}\int F = \text{integer} = n, \tag{9}$$

$\bar{F}$ is twice the one for T^2 with the same a_1 and a_2 and for the same monopole number since the volume of $\mathcal{O}$ is half the volume of T^2. The Dirac equation for the $(++)$ patch is

$$iD^M\Gamma_M\Psi = 0, \tag{10}$$

so that

$$(i\not{D})^2\Psi = (-D^2 - F_{56}1\otimes\tau^3)\Psi \tag{11}$$

(notice that the equation of motion of a scalar field ϕ coupled to the guage field will be $-D^2\phi = 0$).

VI. THE SCALAR EIGENMODES

In the compact subspace with the topology of a torus given by circumferences a_1 and a_2 and the background gauge field with monopole number n we have[4]

$$-D^2_{\text{comp}} = -\frac{\partial^2}{\partial y_1^2} - \left(\frac{\partial}{\partial y_2} - \frac{2\pi i n}{\text{vol}}y_1\right)^2. \tag{12}$$

The solutions of $-D^2_{\text{comp}}\phi = M^2\phi$ fulfilling the periodicity requirement $f(y_1, y_2) = f(y_1, y_2 + a_2)$ are

$$f_{Nm}(y_1, y_2) = \exp\left(\frac{2\pi i m}{a_2}y_2\right)\exp\left[-\frac{2|n|}{a_1a_2}\left(y_1 - \frac{m}{n}a_1\right)^2\right] \times H_N\left[\sqrt{\frac{2\pi|n|}{a_1a_2}}\left(y_1 - \frac{m}{n}a_1\right)\right], \tag{13}$$

where H_N is the Nth Hermite polynomial. The corresponding eigenvalues are

$$M_n^2 = (4\pi/a_1a_2)|n|(N + \tfrac{1}{2}). \tag{14}$$

Imposing the boundary condition (for the torus)

$$\phi(a_1, y_2) = \exp((2\pi i n/a_2)y_2)\phi(0, y_2) \tag{15}$$

we obtain (see Ref. 4)

$$\phi_{Nm} = \left[a_2\left(\frac{a_1a_2}{2|n|}\right)^{1/2}2^N N!\right]^{-1/2}\sum_{k=-\infty}^{+\infty} f_{N,m+kn}. \tag{16}$$

The above T^2 solutions become valid for $\mathcal{O}$ once n is replaced by $2n$ and extra boundary conditions are imposed:
for $A - A'$,

$$\phi(0, y_2) = \phi(0, a_2 - y_2); \quad (17)$$

and for $B - B'$,

$$\phi(a_1/2, y_2) = \phi(a_1/2, a_2 - y_2)e^{i\Lambda}, \quad (18)$$

where $\Lambda = [2\pi i(2n)/a_2]\, y_2$.

In so doing the following identities are useful:

$$f_{Nm}(a_1 + y_1, y_2) = \exp([2\pi i(2n)/a_2]\, y_2) f_{N,m-n}(y_1, y_2), \quad (19)$$

$$f_{Nm}(y_1, a_2 - y_2) = (-1)^N f_{N,-m}(-y_1, y_2). \quad (20)$$

This implies the shape of the boundary condition fulfilled by the eigenfunctions on the torus:

for $A - A'$,

$$\phi_{Nm}(0, a_2 - y_2) = (-1)^N \phi_{N,|2n|-m}(0, y_2), \quad (21)$$

for $B - B'$,

$$\begin{aligned} \phi_{Nm}(a_1/2, a_2 - y_2) &= (-1)^N \phi_{N,|2n|-m}(-a_1/2, y_2) \\ &= (-1)^N \exp([2\pi i(2n)/a_2] y_2) \phi_{N,|2n|-m}(a_1/2, y_2) \end{aligned} \quad (22)$$

(here $0 \leqslant m < |2n|$).

Thus the eigenfunctions on $\mathscr{O}$ are

$$\phi^{\mathscr{O}+}_{Nm}(y_1, y_2) = 1/\sqrt{2}(\phi_{N,m}(y_1, y_2) + (-1)^N \phi_{N,2|n|-m}(y_1, y_2)), \quad (23)$$

for $1 \leqslant m \leqslant |n| - 1$, and

$$\phi^{\mathscr{O}+}_{Nm}(y_1, y_2) = \phi_{N,m}(y_1, y_2), \quad (24)$$

for N even and $m = 0$ or $|n|$.

So the states with N odd are degenerated $|n| - 1$ times, and that with N even $|n| + 1$ times.

We have found simultaneously the spectrum and eigenstates of a complex scalar field coupled to a U(1) gauge field on $\mathscr{O}$. If the theory possesses an extra global U(1) guage symmetry (a phase shift of a complex field ϕ), then the boundary conditions (17) and (18) can be generalized to

$$\phi(0, y_2) = e^{i\alpha} \phi(0, a_2 - y_2), \quad (25)$$

$$\phi(a_1/2, y_2) = e^{i\alpha'} \phi(a_1/2, a_2 - y_2) e^{i\Lambda}, \quad (26)$$

but solutions exist for $\alpha = \alpha' = 0$ and $\alpha = \alpha' = \pi$ only (that is related to two possible ways $\mathbb{Z}_2$ can be embedded in the gauge group). The eigenfunctions in the second case are

$$\phi^{\mathscr{O}-}_{Nm}(y_1, y_2) = 1/\sqrt{2}(\phi_{N,m}(y_1, y_2) - (-1)^N \phi_{N,2|n|-m}(y_1, y_2)), \quad (27)$$

for $1 \leqslant m \leqslant |n| - 1$, and for N *odd* $m = 0$ or $|n|$ the eigenfunctions are given by (24).

VII. FERMIONIC FIELD EIGENSTATES

In order to obtain the boundary conditions for fermions we have to know how $R(\pi)$ acts on the fermion fields,

$$R(\omega^{MN}) = \exp(-i\omega^{MN}\sigma_{MN}/4), \quad (28)$$

where ω parametrizes the rotation, $\sigma_{MN} = i/2[\Gamma_M, \Gamma_N]$, $\sigma_{56} = i/2[\Gamma_5, \Gamma_6] = i/2\ 1 \otimes [\tau^1, \tau^2] = -1 \otimes \tau^3$ so that $R(\omega^{56}) = \exp(i\omega^{56}/2\ 1 \otimes \tau^3)$ and

$$R(\pi) = 1 \otimes i\tau^3 \text{ or } R(3\pi) = -1 \otimes i\tau^3. \quad (29)$$

Writing $\Psi = \psi_4$, where ψ_4 and ψ are four- and two-component spinors, respectively, we obtain an analog of (11) for fermions:

$$(-D^2_{\text{comp}} - F_{56}\tau^3)\psi = M^2\psi. \quad (30)$$

The solutions are labeled as before, by N,m and also by the eigenvalue of $\tau^3 - h$, $h = \pm$. For T^2 the eigenvalues are[4]

$$M^2_{N,h} = (4\pi/a_1a_2)|n|(N + \tfrac{1}{2}(1 - h\sigma)), \quad (31)$$

where $\sigma = \text{sgn}(n)$.

The boundary conditions for fermions are

$$\psi(0, y_2) = e^{i\alpha} i\tau^3 \psi(0, a_2 - y_2), \quad (32)$$

$$\psi(a_1/2, y_2) = e^{i\alpha'} e^{i\Lambda} i\tau^3 \psi(a_1/2, a_2 - y_2). \quad (33)$$

Solutions exist for $\alpha = \alpha' = -\pi/2$,

$$\psi^{\mathscr{O}}_{Nmh=+} = \phi^{\mathscr{O}+}_{Nm} \text{ [Eqs. (23) and (24)]}, \text{ degeneration } |n| + (-1)^N;$$

$$\psi^{\mathscr{O}}_{Nmh=-} = \phi^{\mathscr{O}-}_{Nm} \text{ [Eqs. (27) and (24)]}, \text{ degeneration } |n| - (-1)^N;$$

and for

$$\alpha = \alpha' = +\pi/2;$$

$$\psi^{\mathscr{O}}_{Nmh=+} = \phi^{\mathscr{O}-}_{Nm} \text{ [Eqs. (27) and (24)]}, \text{ degeneration } |n| - (-1)^N;$$

$$\psi^{\mathscr{O}}_{Nmh=-} = \phi^{\mathscr{O}+}_{Nm} \text{ [Eqs. (23) and (24)]}, \text{ degeneration } |n| + (-1)^N.$$

So we have $|n| + 1$ or $|n| - 1$ zero modes, and the massive modes of different helicities combine to form Dirac massive states, but the result does not agree with the usual form of the index theorem.

VIII. THE OTHER SPIN STRUCTURES

This way we have found two spin structures on $\mathscr{O}$, but there are four classes of noncontractible loops on $\mathscr{O}$—every one of them is associated with singularities. (Our field equations are well-defined outside the singularities only so the space we are using has all singular points removed. The solutions are accepted if there exist a limit when the singularity is approached.) For every curve one can impose two kinds of the boundary conditions labeled by a phase acquired by the field after transporting around a singularity (we will mark these two possibilities by $-$ for phase π or $+$ for phase 0). The boundary conditions are not independent since a superposition of four loops, each one going around a separate singularity, produces a contractible curve so the number of $-$ boundary conditions must be even—we expect eight distinct spin structures. Up to now we have obtained only two of them, but it is well known[5] that there are four possible spin structures on T^2, labeled by the phase acquired by the spinor field after being transported around the circles forming the torus: $\{+,+\}$—the one already used, and $\{+,-\}$, $\{-,+\}$, and $\{-,-\}$ spin structures.[5]

The $\{+,-\}$ spin structure on T^2 is obtained after setting

$$\phi(a_1,y_2) = -\exp((2\pi in/a_2)y_2)\phi(0,y_2) \quad (34)$$

instead of (15).

Then the eigenfunctions become

$$\phi^{+-}_{Nm} = \left[a_2\left(\frac{a_1a_2}{2|n|}\right)^{1/2} 2^N N!\right]^{-1/2} \times \sum_{k=-\infty}^{+\infty} (f_{N,m+2kn} - f_{N,m+(2k+1)n}), \quad (35)$$

where the functions f are given by (13) and $0\leqslant m<|n|-1$. Equations (21) and (22) with ϕ^{+-} instead of ϕ (that should be now called ϕ^{++}) will acquire an extra minus sign on the rhs. The previous formulas for eigenfunctions on $\mathscr{O}$ (23) and (24) are still valid after the above replacement, this leads to two spin structures on $\mathscr{O}$ corresponding to $\alpha=\alpha'=\pi/2$ and $\alpha=\alpha'=-\pi/2$ in (32) and (33) with the previously described spectra.

For construction of the remaining two spin structures on T^2 the functions f in (13) have to be modified to fulfill the new boundary condition:

$$f^-(y_1,0) = -f^-(y_1,a_2), \quad (36)$$

$$f^-_{Nm}(y_1,y_2) = \exp\left(\frac{2\pi i(m+1/2)}{a_2}y_2\right) \times\exp\left[-\frac{2|n|}{a_1a_2}\left(y_1-\frac{m+1/2)}{n}a_1\right)^2\right] \times H_N\left[\left(\frac{2\pi|n|}{a_1a_2}\right)^{1/2}\left(y_1-\frac{(m+1/2)}{n}a_1\right)\right]. \quad (37)$$

The wave functions for $\{-,+\}$ and $\{-,-\}$ boundary conditions are given by the formulas (16) and (35), respectively, after substitution of f^- for f. Equations (21) and (22) are replaced by

$$\phi^-_{Nm}(0,a_2-y_2) = \mp(-1)^N\phi^-_{N,|2n|-m-1}(0,y_2), \quad (38)$$

$$\phi^-_{Nm}(a_1/2,a_2-y_2) = \mp(-1)^N\exp([2\pi i(2n)/a_2]y_2) \times\phi^-_{N,|2n|-m-1}(a_1/2,y_2), \quad (39)$$

where the upper sign corresponds to $\{+,-\}$ and the lower one to $\{-,-\}$ cases. The scalar eigenfunctions on $\mathscr{O}$ for $\{-,+\}$ are ($0\leqslant m<|n|$, any N)

$$\phi^{\mathscr{O}-}_{Nm}(y_1,y_2) = 1/\sqrt{2}(\phi^-_{N,m}(y_1,y_2) \mp(-1)^N\phi^-_{N,2|n|-m-1}(y_1,y_2)), \quad (40)$$

where the upper sign corresponds to $\alpha=\alpha'=0$ and the lower one to $\alpha=\alpha'=\pi$. For $\{-,-\}$ the upper sign should be exchanged with the lower one. In both cases the degeneration of states is $|n|$. Applying this to fermions goes the same way as before and leads to $|n|$ chiral zero modes in both cases, so that the usual index theorem holds, but in the light of the previous results it is rather an accident.

We shall look at the $n=0$ case for completeness. The eigenfunctions for T^2 are given by

$$\phi_{n_1n_2} = \exp(2\pi i(n_1y_1/a_1+n_2y_2/a_2)), \quad (41)$$

for the $\{+,+\}$ patch, with masses

$$M^2 = 4\pi^2[(n_1/a_1)^2+(n_2/a_2)^2]. \quad (42)$$

In order to get the solutions for the other types of the boundary conditions one has to replace n_1 by $n_1+\frac{1}{2}$, n_2 by $n_2+\frac{1}{2}$, and both n_1,n_2 by $n_1+\frac{1}{2}$ and $n_2+\frac{1}{2}$ for $\{+,-\}$, $\{-,+\}$, and $\{-,-\}$, respectively, in (41) and (42). The massless states are present for $\{+,+\}$ only. The boundary conditions required to set the fields on $\mathscr{O}$ are unchanged [see (25),(26) and (32),(33)] but now $\Lambda=0$. The eigenfunctions for $\{+,+\}$ are

$$\phi^{\mathscr{O}}_{n_1n_2} = 1/\sqrt{2}(\phi_{n_1n_2} \pm \phi_{-n_1,-n_2}), \quad (43)$$

where $n_2\geqslant 0$, the upper sign corresponds to $\alpha=\alpha'=0$, and the lower one to $\alpha=\alpha'=\pi$. There is a zero mode for $n_1=n_2=0$ and this implies in the fermionic case the existence of a *chiral zero mode for* $n_{\rm mon}=0$ with $h=+$ for $\alpha=\pi/2$ and with $h=-$ for $\alpha=-\pi/2$. Degenerations of the higher excited zero modes are 1 for $n_1=0$, $n_2>0$ and 2 in the remaining cases. The remaining eigenfunctions are

for $\{+,-\}$, $\phi^{\mathscr{O}}_{n_1n_2} = 1/\sqrt{2}(\phi_{n_1n_2} \mp \phi_{-n_1,-n_2-1})$,

for $\{-,+\}$, $\phi^{\mathscr{O}}_{n_1n_2} = 1/\sqrt{2}(\phi_{n_1n_2} \pm \phi_{-n_1-1,-n_2})$,

for $\{-,-\}$, $\phi^{\mathscr{O}}_{n_1n_2} = 1/\sqrt{2}(\phi_{n_1n_2} \mp \phi_{-n_1-1,-n_2-1})$.

It is interesting to know which kind of the boundary conditions is in fact chosen during compactification. It is probably desirable to have single-valued fermion fields on $\mathscr{O}$ since after blowing up the singularities the loops around them will become contractible; moreover we are working then with a common manifold and the index theorem must hold. Because of that it is probably a good idea to choose the $\{+,+\}$ type of the boundary conditions.

IX. ONE-LOOP EFFECTIVE POTENTIAL

Let us look at the one-loop effective potential for various boundary conditions we have analyzed (we are working here with the massless case, the massive case involves dealing with logarithmic divergences that survive the dimensional regularization and will be considered elsewhere together with contributions of other fields). From the Appendix we have

$$V_{\rm eff} = -4(2\pi)^{-4}\zeta(3)(4\pi n/{\rm vol})^2V(n), \quad (44)$$

where ${\rm vol}=a_1a_2/2$, $V(n)$ depends on the degeneration,

$$V(n) = \begin{cases} |n|+7, & d_N = n+(-1)^N, \\ |n|, & d_N = n, \\ |n|-7, & d_N = n-(-1)^N. \end{cases}$$

The lowest value appears for $\{+,+\}$, $\alpha=\pi/2$ and $\{-,-\}$, $\alpha=-\pi/2$. For $n=0$ the result is

$$V_{\rm eff} = 2[1/(a_1a_2)^2][1/(4\pi)]\pi^{-3}V, \quad (45)$$

where $\beta=a_1/a_2$, and $W(\beta) = \Sigma[(1/\beta)n_1^2+\beta n_2^2]^{-3}$,

for $V^{\{+,-\}}_{\rm eff} - V^{\{+,+\}}_{\rm eff}$,

V becomes $2(W(\beta)-W(2\beta)/8)$;

for $V^{\{-,+\}}_{\rm eff} - V^{\{+,+\}}_{\rm eff}$,

V becomes $2(W(\beta)-W(\beta/2)/8)$;

for $V_{\rm eff}^{\{-,-\}} - V_{\rm eff}^{\{+,+\}}$,

V becomes $2(W(2\beta)/8 + W(\beta/2)/8 - W(\beta)/4)$;

so the $\{+,+\}$ case corresponds again to the minimal value of the effective potential. Unfortunately it is impossible to distinguish the spin structure uniquely in this way.

X. CONCLUSIONS

We have found the spectra of fermions for all possible spin structures on $\mathscr{O}$. The effective potential seems to distinguish the $\{+,+\}$ spin structure in agreement with the intuition and the requirements imposed by blowing up the singularities. This procedure inevitably involves a complicated change of the spin structure. There is more interesting information—the usual index theorem is not respected and there exists a chiral zero mode in spite of the topologically trivial gauge field background. This may be used to construct a model with $N=1$ supersymmetry (SUSY) in four dimensions[6] and may have some relevance for constructing Kaluza–Klein theories with chiral fermions. The question of vacuum stability will be addressed more completely in a forthcoming paper; the case of Yang–Mills and supersymmetric field theories[6,7] will be considered there, as well as divergences appearing in the effective potential and the counterterms needed to renormalize the theory at the one-loop level. The possibility of fractional monopole number[8] and anomalies will be considered later.

ACKNOWLEDGMENTS

I would like to thank Professor G. Segrè for suggesting investigation of the problem and introducing me to the subject. I am also grateful to the Department of Theoretical Physics of Oxford University where this work was begun, for hospitality.

APPENDIX: CALCULATION OF THE EFFECTIVE POTENTIAL

The contribution to the one-loop effective potential from chiral fermions is

$$V_{\rm eff} = -\tfrac{1}{2}\ln\det((i\not{D})^2). \tag{A1}$$

Introducing dimensional regularization,

$$V_{\rm eff} = -\mu^{-(d-4)}\int\frac{d^dk}{(2\pi)^d}\sum_{Nh} d_{Nh}\ln(k^2+M_{Nh}^2), \tag{A2}$$

for $\{+,+\}$, $d_{N+} = |n| + (-1)^N$, $d_{N-} = |n| - (-1)^N$, using

$$\int\frac{d^dk}{(2\pi)^d}\ln(k^2+M^2) = -\Gamma\left(\frac{-d}{2}\right)(4\pi)^{-d/2}M^d \tag{A3}$$

and

$$\sum_{N=1}^{\infty} N^{d/2} = \zeta\left(\frac{-d}{2}\right),$$
$$\sum_{N=1}^{\infty}(-1)^N N^{d/2} = (2^{(d+2)/2}-1)\zeta\left(\frac{-d}{2}\right), \tag{A4}$$

and the reflection formula

$$\zeta(-d/2) = -(1/\pi)(2\pi)^{-d/2}\times\sin(\pi d/4)\Gamma(1+d/2)\zeta(1+d/2), \tag{A5}$$

for $d=4$, we have

$$V_{\rm eff} = -4(2\pi)^{-4}\zeta(3)(4\pi n/{\rm vol})^2(|n|+7), \quad \text{for } d_N = |n| + (-1)^N,$$
$$V_{\rm eff} = -4(2\pi)^{-4}\zeta(3)(4\pi n/{\rm vol})^2|n|, \quad \text{for } d_N = |n|, \tag{A6}$$
$$V_{\rm eff} = -4(2\pi)^{-4}\zeta(3)(4\pi n/{\rm vol})^2(|n|-7), \quad \text{for } d_N = |n| - (-1)^N.$$

The $n=0$ case is a bit more complicated technically since it involves Epstein Z_p functions,[9]

$$V_{\rm eff} = \mu^{-(d-4)}(4\pi)^{-d/2}\Gamma\left(\frac{-d}{2}\right)(2\pi)^{d/2}\times\sum_{n_1n_2}{}'\left[\left(\frac{n_1}{a_1}\right)^2+\left(\frac{n_2}{a_2}\right)^2\right]^{d/2}, \tag{A7}$$

where the contributions of both helicities were added, and the prime denotes omission of the $n_1 = n_2 = 0$ term in the summation. The sum is conveniently rewritten as

$$(a_1a_2)^{-d/2}\sum_{n_1n_2}{}'\left[\frac{n_1^2}{\beta}+\beta n_2^2\right]^{d/2}, \tag{A8}$$

where $\beta = a_1/a_2$. A sum like

$$\sum_{n_1n_2}{}'[(n_1+p)^2\beta^{-1}+(n_2+q)^2\beta]^{d/2} \tag{A9}$$

can be represented with use of the Jacobi theta functions,

$$\left(\Gamma\left(\frac{-d}{2}\right)\right)^{-1}\int_0^\infty\frac{dx}{x}x^{-d/2}\vartheta\left(\frac{p}{\pi\beta},\frac{x}{\pi\beta}\right)\vartheta\left(\frac{q\beta}{\pi},\frac{x\beta}{\pi}\right)\times\exp\left(\frac{-p^2x}{\beta}-q^2x\beta\right). \tag{A10}$$

Then[9]

$$\left(\Gamma\left(\frac{-d}{2}\right)\pi\right)^{-1}\int_0^\infty\frac{dx}{x}x^{-(d/2+1)}\vartheta\left(\frac{p}{i},\frac{\beta\pi}{x}\right)\vartheta\left(\frac{q}{i},\frac{\pi}{\beta x}\right). \tag{A11}$$

Changing the integration variables we have

$$\left(\Gamma\left(\frac{-d}{2}\right)\pi\right)^{-1}\int_0^\infty\frac{dx}{x}x^{(d/2+1)}\vartheta\left(\frac{p}{i},\beta\pi x\right)\vartheta\left(\frac{q}{i},\frac{\pi}{\beta}x\right). \tag{A12}$$

Using $\vartheta(1/2i,x) = 2\vartheta(0,4x) - \vartheta(0,x)$ we are able to produce formulas for every case considered. For the $\{+,+\}$ case we have

$$V_{\rm eff}^{\{+,+\}} = -\frac{i}{4\pi}\pi^{-3}\Gamma(3)(a_1a_2)^{-2}\times\int_0^\infty\frac{dx}{x}x^{(d/2+1)}\vartheta(0,\beta\pi x)\vartheta\left(0,\frac{\pi}{\beta}x\right)$$
$$= -\frac{1}{4\pi}\pi^{-3}\Gamma(3)(a_1a_2)^{-2}\times\sum_{n_1n_2}{}'\left[\frac{n_1^2}{\beta}+\beta n_2^2\right]^{-3}. \tag{A13}$$

[1]L. Dixon, J. A. Harvey, C. Vafa, and E. Witten, Nucl. Phys. B **261**, 678 (1985); **274**, 285 (1986).
[2]P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. B **258**, 46 (1985).
[3]A. Strominger and E. Witten, Commun. Math. Phys. **101**, 341 (1985); A. Strominger, Phys. Rev. Lett. **55**, 2547 (1985).
[4]M. J. Duncan and C. G. Segrè, Phys. Lett. B **195**, 36 (1987).
[5]C. Vafa, Nucl. Phys. B **273**, 592 (1986).
[6]A. Salam and E. Sezgin, Phys. Lett. B **147**, 47 (1984).
[7]S. Randjbar-Daemi, A. Salam, and J. Stradhee, Nucl. Phys. B **214**, 491 (1983).
[8]N. S. Manton, Ann. Phys. (NY) **159**, 220 (1985).
[9]J. Ambjorn and S. Wolfram, Ann. Phys. (NY) **147**, 1 (1983).

Evaluation of Feynman diagrams in the logarithmic approach to quantum field theory

Carl M. Bender
Department of Physics, Washington University, St. Louis, Missouri 63130
H. F. Jones
Physics Department, Imperial College, London, SW7 2BZ, England

(Received 29 June 1988; accepted for publication 17 August 1988)

The techniques necessary to compute the $O(\delta^2)$ contributions to the Green's functions of a scalar field theory with self-interaction $\lambda(\phi^2)^{1+\delta}$ in d-dimensional space-time are developed. The resulting expressions are evaluated explicitly for $d=1$, 0, and some negative even dimensions. For $d=3$ and 4 we calculate their leading behavior in terms of a spatial cutoff a.

I. INTRODUCTION

A new approach to quantum field theory—a perturbation expansion in the powers of the interaction—has recently been proposed and to some extent developed in Refs. 1 and 2. To take the specific case of a scalar $\lambda\phi^4$ theory in d dimensions, the interaction is written as $\lambda\phi^{2(1+\delta)}$ and δ is taken as the expansion parameter. This is not necessarily a weak-coupling expansion and it therefore falls into the same category as other nonperturbative techniques, which include the large N expansion or lattice calculations, that can be used to investigate problems whose solutions are not accessible via ordinary perturbation theory.

In Refs. 1 and 2 the basic computational rules of the δ expansion were explained. When $\phi^{2\delta}$ is expanded it produces interaction terms of the form $\delta^k(\ln\phi^2)^k$. The procedure to evaluate the Green's functions for such a logarithmic interaction is to construct a provisional Lagrangian $\widetilde{L}$, which contains polynomial Lagrangians of arbitrary *integer* powers α_i. Having evaluated the Green's functions of $\widetilde{L}$ to any given order K, say, in δ one must then continue the expression to real values of the α_i and apply a certain differential operator D_K to the result in order to obtain the Green's functions of the logarithmic Lagrangian. The purpose of the present paper is to develop techniques that will enable us to evaluate such diagrams through order δ^2 for arbitrary d.

Actually, each two-vertex diagram represents a finite sum over the number of possible propagators joining the two vertices. The two tasks that need to be addressed are, first, to perform the sum to obtain an analytic function of $\alpha\equiv\alpha_1$ and $\beta\equiv\alpha_2$, which were originally integers, and, second, to apply to this function the differential operator $D_{K=2}$. These are achieved in Sec. II.

The resulting integral expressions cannot be evaluated in closed form for general dimension d. However, there are some special cases where this can be done, notably when $d=1$, a field theory in one space dimension, namely quantum mechanics, and $d=0$, corresponding to the evaluation of one-dimensional integrals that provided the original impetus for the development of the method in Ref. 1. These cases are evaluated as field theories in Sec. III, where we also look at the soluble cases of even negative dimensions, $d=-2,-4,\ldots$.

For dimensions $d\geqslant 2$ the theory contains infinite quantities and needs to be regularized as a preliminary to renormalization. In Sec. IV we propose one such method of regularization and evaluate the asymptotic behavior of the integral expressions obtained in Sec. II for the two-point and four-point Green's functions in dimensions $d=3$ and 4. These results form the basis of the discussion of renormalization in Ref. 3.

We conclude the paper in Sec. V with a brief discussion of the results obtained and possibilities for further development.

II. EVALUATION OF GREEN'S FUNCTIONS TO ORDER δ^2

A. Vacuum diagrams

We consider first the calculation of the connected vacuum diagrams shown in Fig. 1. There are effectively two distinct diagrams corresponding to the direct and crossed terms arising from a double application of the provisional Lagrangian[1]

$$\begin{aligned}\widetilde{L}_{K=2} &= \tfrac{1}{2}(\partial\phi)^2 + \tfrac{1}{2}m^2\phi^2 + \delta\lambda M^d\big((\phi^2M^{2-d})^{\alpha+1} \\ &\quad - (\phi^2M^{2-d})^{\beta+1}\big) \\ &\quad + \delta^2\lambda M^d\big((\phi^2M^{2-d})^{\alpha+1} + (\phi^2M^{2-d})^{\beta+1}\big).\end{aligned} \tag{2.1}$$

Each diagram represents a finite sum in which the number of pairs of lines, l, joining the two vertices can vary from 1 to $l_{\max}$. Since an α vertex has $\alpha+1$ pairs of lines attached to it, and similarly for a β vertex, $l_{\max}=\alpha+1$ for Fig. 1(a) and $\min(\alpha+1,\beta+1)$ for Fig. 1(b). We will evaluate the diagrams in configuration space, taking one vertex at the origin and the other at x. Then each propagator gives a propagator

FIG. 1. Two-vertex vacuum diagrams. (a) Direct $(\alpha\alpha)$ contribution. (b) Crossed $(\alpha\beta)$ contribution.

 0022-2488/88/122659-07$02.50

factor $\Delta(x)$, while each closed loop, or "petal" gives $\Delta(0)$, a quantity which will have to be regularized in dimensions $d \geqslant 2$.

The l summation: Consider first Fig. 1(a). For a given l, the symmetry factor associated with this diagram is

$$S^{l}_{\alpha\alpha} = \tfrac{1}{2}(2^{2\alpha+2-2l}(2l)!((\alpha+1-l)!)^2)^{-1}. \tag{2.2}$$

Thus the contribution to the connected vacuum functional $\widetilde{W}[0]$ is

$$\widetilde{W}_{\alpha\alpha}[0] = -\delta^2\lambda^2 M^{2d}\int d^dx(\Delta(0)M^{2-d})^{2\alpha+2} \times((2\alpha+2)!)^2 \sum_{l=1}^{\alpha+1} S^{l}_{\alpha\alpha}\left(\frac{\Delta(x)}{\Delta(0)}\right)^{2l}. \tag{2.3}$$

To proceed we first use the duplication formula[4] on the factor $(2l)! = \Gamma(2l+1)$ occurring in $S^{l}_{\alpha\alpha}$ to write

$$\frac{1}{(2l)!(\alpha+1-l)!} = \frac{1}{(\alpha+1)!}\frac{\sqrt{\pi}}{2^{2l}}\binom{\alpha+1}{l}\frac{1}{\Gamma(l+\frac{1}{2})}. \tag{2.4}$$

Using the reflection formula[4] on the remaining factor of $1/(\alpha+1-l)!$ we can manipulate (2.3) into the form

$$\widetilde{W}_{\alpha\alpha}[0] = -\tfrac{1}{2}\sqrt{\pi}\delta^2\lambda^2 M^{2d}(\tfrac{1}{2}\Delta(0)M^{2-d})^{2(\alpha+1)} \times\frac{(\Gamma(2\alpha+3))^2}{\Gamma(\alpha+2)\Gamma(\alpha+\frac{3}{2})}\frac{\sin(\pi\alpha)}{\pi} \times\int d^dx \sum_{l=1}^{\alpha+1}\binom{\alpha+1}{l}(-z)^l \times B(l-\alpha-1,\alpha+\tfrac{3}{2}), \tag{2.5}$$

where $z = (\Delta(x)/\Delta(0))^2$.

If we now express the beta function as a one-dimensional integral,

$$B(l-\alpha+\tfrac{3}{2}) = \int_0^1 dy(1-y)^{\alpha+1/2}y^{l-\alpha-2}, \tag{2.6}$$

the sum over l can be performed:

$$\sum_{l=1}^{\alpha+1}\binom{\alpha+1}{l}(-yz)^l = (1-yz)^{\alpha+1} - 1. \tag{2.7}$$

The subsequent y integration is expressed in terms of a hypergeometric function[4]:

$$\int_0^1 dy(1-y)^{\alpha+1/2}y^{-\alpha-2}((1-yz)^{\alpha+1}-1) = \frac{\sqrt{\pi}}{\sin(\pi\alpha)}\frac{\Gamma(\alpha+\frac{3}{2})}{\Gamma(\alpha+2)} \times(F(-1-\alpha,-1-\alpha;\tfrac{1}{2};z)-1). \tag{2.8}$$

Upon using the duplication formula for $\Gamma(2\alpha+3)$ we can write $\widetilde{W}_{\alpha\alpha}[0]$ in the final form

$$\widetilde{W}_{\alpha\alpha}[0] = -\tfrac{1}{2}\delta^2\lambda^2 M^4(\Delta(0))^2X_\alpha^2\int d^dx \times(F(-1-\alpha,-1-\alpha;\tfrac{1}{2};z)-1), \tag{2.9}$$

where

$$X_\alpha = (2M^{2-d}\Delta(0))^\alpha\Gamma(\alpha+\tfrac{3}{2})/\Gamma(\tfrac{3}{2}). \tag{2.10}$$

Although we have been a little cavalier in our derivation of (2.9), treating α sometimes as an integer, sometimes not, one can check *a posteriori* that the hypergeometric series in (2.9) indeed reproduces the original series in (2.3) when α is integral. In what follows, (2.9) provides an expression continuable to general α to which we apply the differential operator $D_{K=2}$ of Ref. 1.

In a completely parallel fashion the contribution of Fig. 1(b) to the connected vacuum functional of the provisional Lagrangian is

$$\widetilde{W}_{\alpha\beta}[0] = \delta^2\lambda^2M^2(\Delta(0))^2X_\alpha X_\beta\int d^dx \times(F(-1-\alpha,-1-\beta;\tfrac{1}{2};z)-1). \tag{2.11}$$

In the derivation of (2.11) one has to assume a definite ordering of the integers α and β, but, as can be seen, the final result is symmetric in α and β, and completely independent of that choice.

B. Differentiation

To obtain the Green's functions of the logarithmic Lagrangian arising from the expansion of $\exp(\delta\ln(\phi^2M^{2-d}))$ one has to apply the operator D to those of the provisional Lagrangian, as explained in Refs. 1 and 2. As we are working to second order, the appropriate differential operator is

$$D_{K=2} = \frac{1}{2}\left(\frac{\partial}{\partial\alpha}-\frac{\partial}{\partial\beta}\right)+\frac{1}{4}\left(\frac{\partial^2}{\partial\alpha^2}+\frac{\partial^2}{\partial\beta^2}\right). \tag{2.12}$$

After the differentiation α and β are to be set to zero.

Consider first the evaluation of $W_{\alpha\alpha}[0]$. After the first term, $2(1+\alpha)^2z$, in the expansion of $F(-1-\alpha,-1-\alpha;\frac{1}{2};z)-1$ all subsequent terms contain a factor of α^2. Thus the single derivative picks up a nonzero contribution from the first term only:

$$\frac{1}{2}\frac{\partial}{\partial\alpha}[X_\alpha^2(F-1)] = 2Sz, \tag{2.13}$$

where

$$S = 1+\psi(\tfrac{3}{2})+\ln(2M^{2-d}\Delta(0)). \tag{2.14}$$

However, this single derivative term will cancel against the corresponding terms in $W_{\beta\beta}[0]$.

When operating on all but the first term, the action of the double derivative $\frac{1}{4}\partial^2/\partial\alpha^2$ is simply to replace the factor α^2 by 2, thus generating the infinite series

$$\sum_{k=2}^{\infty} B(k-1,\tfrac{3}{2})\frac{z^k}{k(k-1)}.$$

By writing the beta function as a one-dimensional integral as in (2.6) and expanding in partial fractions, this series can be expressed as

$$\int_0^1\frac{dt}{t^2}(1-t)^{1/2}[(1-zt)\ln(1-zt)+zt].$$

The final expression for $W_{\alpha\alpha}[0]$ is

$$W_{\alpha\alpha}[0] = -\tfrac{1}{2}\delta^2\lambda^2M^4(\Delta(0))^2 \times\int d^dx\left\{\left(2S^2-1+\psi'\left(\frac{3}{2}\right)\right)z + \int_0^1 \frac{dt}{t^2}(1+t)^{1/2}[(1-zt)\ln(1-zt)+zt]\right\}, \tag{2.15}$$

with an identical contribution from a diagram where each vertex is a β vertex. In (2.15) we have omitted the (canceling) single derivative term.

The evaluation of $W_{\alpha\beta}[0]$ is easier. In this case every term beyond the first in $F(-1-\alpha,-1-\beta;\frac{1}{2},z)-1$ contains a factor of $\alpha\beta$. Since there are no mixed derivatives in D, all these terms give zero when α and β are set to zero after the differentiation. Thus

$$W_{\alpha\beta}[0] = \tfrac{1}{2}\delta^2\lambda^2M^4(\Delta(0))^2\int d^dx\left(S^2-1+\psi'\left(\frac{3}{2}\right)\right)z, \tag{2.16}$$

again with an identical contribution $W_{\beta\alpha}[0]$. Altogether, then (2.15) and (2.16) and their $\alpha\leftrightarrow\beta$ counterparts give

$$W_1[0] = -\delta^2\lambda^2M^4(\Delta(0))^2\int d^dx\left\{S^2z + \int_0^1\frac{dt}{t^2}(1-t)^{1/2}[(1-zt)\ln(1-zt)+zt]\right\}. \tag{2.17}$$

To this should be added the contribution of the single-vertex diagram of Fig. 2, which arises from a single application of the last term in $\tilde{L}_{k=2}$ [Eq. (2.11)]. Here there is no summation to be performed and we merely have to apply D to

$$\delta^2\lambda M^d(\Delta(0)M^{2-d})^{\alpha+1}\frac{\Gamma(2\alpha+3)}{2^{\alpha+1}\Gamma(\alpha+2)},$$

giving

$$W_2[0] = \tfrac{1}{2}\delta^2\lambda m^2\Delta(0)((S-1)^2+\psi'(\tfrac{3}{2})). \tag{2.18}$$

Note that in (2.17) and (2.18) the (mass) dimensions of $W[0]$ are correctly given as d. The propagator $\Delta(0)$ has dimension $d-2$, and in (2.16) the x-space integration has dimension $-d$.

C. Two-point and four-point functions

There are three topologically distinct types of diagrams contributing to the proper two-point functions, or self-energy part, as shown in Fig. 3. The first, Fig. 3(a), is rather easily dealt with. Before differentiation the expression is

$$-\delta^2\lambda M^2(\Delta(0)M^{2-d})^\alpha\Gamma(2\alpha+3)/[2^\alpha\Gamma(\alpha+1)],$$

which yields

$$\Pi^{(a)}(p^2) = -\delta^2\lambda M^2(S^2-1+\psi'), \tag{2.19}$$

the same combination as appears in (2.16).

The evaluation of Fig. 3(b) proceeds along very similar

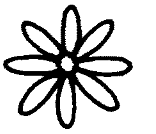


FIG. 2. Single-vertex vacuum diagram.

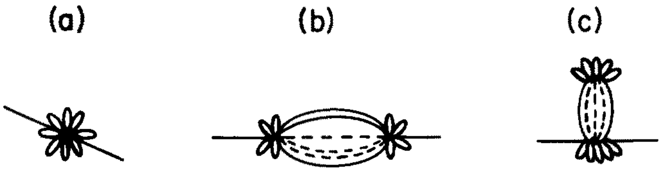


FIG. 3. Contributions to the proper two-point function. (a) Single-vertex diagram. (b) and (c) Double-vertex diagram. Only the $(\alpha\beta)$ contributions are shown explicitly.

lines to that of the vacuum diagrams of Fig. 1. Before differentiation the contribution of the mixed $\alpha\beta$ term is

$$\tilde{\Pi}^{(b)}_{\alpha\beta} = -2\delta^2\lambda^2M^4\Delta(0)(\alpha+1)(\beta+1)X_\alpha X_\beta \times\int d^dx\, e^{ip\cdot x}z^{1/2}[F(-\alpha,-\beta;\tfrac{3}{2};z)-1], \tag{2.20}$$

from which $\tilde{\Pi}^{(b)}_{\alpha\alpha}$ can be obtained by setting $\beta=\alpha$ and supplying the appropriate minus sign [from Eq. (2.1)].

In fact, it is only $\tilde{\Pi}^{(b)}_{\alpha\alpha}$ (and $\tilde{\Pi}^{(b)}_{\beta\beta}$) that gives a nonvanishing contribution upon differentiation since each term of (2.20) contains a factor $\alpha\beta$ that cannot be removed by the unmixed derivatives contained in D. In the case of $\tilde{\Pi}^{(b)}_{\alpha\alpha}$ each term contains a factor α^2 upon which both derivatives must act. This generates the series

$$\sum_{k=1}^{\infty}\frac{z^k}{k}B(k,\tfrac{3}{2}),$$

which can again be expressed as an integral over an auxiliary variable t to give

$$\Pi^{(b)} = -4\delta^2\lambda^2M^4\Delta(0)\int d^dx\, e^{ip\cdot x}z^{1/2} \times\int_0^1\frac{dt}{t}(1-t)^{1/2}\ln(1-zt). \tag{2.21}$$

In Fig. 3(c) we encounter for the first time a diagram that is not symmetric in α and β, for the reason that the two vertices are no longer equivalent. Before differentiation we obtain

$$\tilde{\Pi}^{(c)}_{\alpha\beta} = -2\delta^2\lambda^2M^4\Delta(0)(\beta+1)X_\alpha X_\beta \times\int d^dX\left(F\left(-1-\alpha,-\beta;\frac{1}{2};z\right)-1\right), \tag{2.22}$$

as is easily seen by comparison with the vacuum graph of Fig. 1(b). As usual, the mixed graphs are simpler to differentiate, while the $(\alpha\alpha)$ and $(\beta\beta)$ graphs give rise to an infinite series expressible as an integral over t. Altogether the contribution from Fig. 3(c) and its counterparts is

$$\Pi^{(c)} = 4\delta^2\lambda^2M^4\Delta(0)\int d^dx\left\{Sz+\int_0^1\frac{dt}{t^2} \times(1-t)^{1/2}[\ln(1-zt)+zt]\right\}. \tag{2.23}$$

The only momentum dependence comes from $\Pi^{(b)}$. For purposes of renormalization[4] one needs $\Pi(0)$ and its derivative $d\Pi(p^2)/dp^2$ evaluated at zero momentum. In that case one can expand $e^{ip\cdot x}$ under the integral as $1-x^2p^2/2d+\dots$.

At this point the general procedure should be clear, and we merely list the contributions of the four diagrams of Fig. 4 to the proper four-point vertex Γ_4. For simplicity we have taken all momenta zero, which is all that is needed for renormalization. Figures 4(a) and 4(d) do not contain any momentum dependence. The contributions are

$$\Gamma_4^{(a)} = -4\delta^2\lambda M^2 S/\Delta(0)\,, \tag{2.24}$$

$$\Gamma_4^{(b)} = -12\delta^2\lambda^2 M^4 \int d^d x \int_0^1 \frac{dt}{t}(1-t)^{-1/2}\ln(1-zt)\,, \tag{2.25}$$

$$\Gamma_4^{(c)} = -32\delta^2\lambda^2 M^4 \int d^d x\, z^{3/2} \int_0^1 \frac{dt(1-t)^{1/2}}{1-zt}\,, \tag{2.26}$$

$$\Gamma_4^{(d)} = 8\delta^2\lambda^2 M^4 \int d^d x \left\{ -Sz + z^2 \int_0^1 \frac{dt(1-t)^{1/2}}{1-zt} \right\}. \tag{2.27}$$

III. EXPLICIT FORMS FOR FINITE THEORIES

In this section we show how the x-space integral in the expression (2.17) for $W_1[0]$ can be explicitly evaluated for the cases $d=1$, corresponding to the ground-state energy of the anharmonic oscillator in quantum mechanics, and $d=0$, corresponding to the evaluation of the one-dimensional integral

$$\int_{-\infty}^{\infty} dx \exp\left(-\frac{1}{2}m^2x^2 - \lambda M^4 x^4\right).$$

The integral can also be evaluated in negative even dimensions, $d=-2,-4,\ldots$. This is of more than academic interest, since the work of Halliday and co-workers[5] has shown that a study of negative dimensions can yield information about the positive-dimensional case. However, what would be required is a general formula for negative (even) dimensions, which we have so far been unable to obtain.

A. $d=1$ (quantum mechanics)

We are concerned with the evaluation of the integral

$$I_d = \int d^d x \int_0^1 \frac{dt}{t^2}(1-t)^{1/2}[(1-zt)\ln(1-zt) + zt] \tag{3.1}$$

occurring in (2.17). Similar integrals occur in the expressions for other Green's functions. Before specializing to $d=1$ we can convert the x-space integration measure to polar coordinates according to $d^d x = \Omega_d\, x^{d-1}\, dx$, where x now represents the "radial" coordinate and Ω_d is the solid angle in d dimensions, $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$. On integration by parts we can rewrite (3.1) as

$$I_d = \Omega_d \int_0^\infty dx \frac{x^d}{d}\frac{dz}{dx}\int_0^1 \frac{dt}{t}(1-t)^{1/2}\ln(1-zt)\,. \tag{3.2}$$

Recall that $z=(\Delta(x)/\Delta(0))^2$, where $\Delta(x)$, the scalar propagator in d dimensions, is given by a modified Bessel function

$$\Delta(x) = (2\pi)^{-d/2}(x/m)^{1-d/2}K_{1-d/2}(mx)\,, \tag{3.3}$$

where m is the shifted mass and $m^2=\mu^2+2\lambda M^2$, which is a feature of the δ expansion.

For $d<2$, $\Delta(0)$ exists and is given by the limiting form[4] of the Bessel function

$$\Delta(0) = [1/(4\pi)^{d/2}]m^{d-2}\Gamma(1-d/2)\,. \tag{3.4}$$

Thus

$$z = 2^d((mx)^\nu K_\nu(mx)/\Gamma(\nu))^2\,, \tag{3.5}$$

where $\nu = 1-d/2$. For $d=1$ the Bessel function is $K_{1/2}(mx) = (\pi/(2mx))^{1/2}e^{-mx}$, so that

$$z = e^{-2mx} \quad (d=1)\,. \tag{3.6}$$

In Eq. (3.2) it is then convenient to charge variables from x to z and integrate again by parts on z, leading to

$$mI_1 = -\int_0^1 dt(1-t)^{1/2}\left[\frac{1}{t}\ln(1-t) + \int_0^1 \frac{z\,dz}{1-zt}(1-\ln(zt))\right]. \tag{3.7}$$

After various charges of variables and integrations by parts, this can be reduced to standard integrals tabulated by Barbieri *et al.*[6] and expressible in terms of the Riemann zeta functions $\zeta(2)$ and $\zeta(3)$, or equivalently $\psi'(\frac{3}{2})$ and $\psi''(\frac{3}{2})$:

$$mI_1 = 2(\psi'(\tfrac{3}{2}) + \tfrac{1}{2}\psi'(\tfrac{3}{2})\ln 2 + \tfrac{1}{16}\psi''(\tfrac{3}{2}) - 2 + 2\ln 2 - \tfrac{1}{2})\,. \tag{3.8}$$

The other integral $\int d^d x\, z$ occurring in (2.17) can be evaluated for general $d<2$ as

$$m^d\Omega_d \int_0^\infty dx\, x^{d-1}z = \frac{1}{(4\pi)^{d/2}}\frac{\Gamma(2-d/2)}{(m^{2-d}\Delta(0))^2}\,. \tag{3.9}$$

To compare with the result obtained in Ref. 1 for the ground-state energy of the anharmonic oscillator using Rayleigh–Schrödinger perturbation theory we set $\mu=0$ and $\lambda=\frac{1}{2}$, so that $m=M$. Then, including the $O(\delta)$ contribution,

$$W[0] = \tfrac{1}{4}\delta M\psi(\tfrac{3}{2}) - \tfrac{1}{16}\delta^2 M\,[2\psi(\tfrac{3}{2}) - \psi^2(\tfrac{3}{2}) + \psi'(\tfrac{3}{2})\ln 2 + \tfrac{1}{8}\psi''(\tfrac{3}{2}) - 4 + 4\ln 2]\,. \tag{3.10}$$

Equation (3.10) is identical with Eq. (13) of Ref. 1 when $W[0]$ is identified with ΔE_0.

The shift in the energy of the first excited state can in principle be obtained from an evaluation of $\Pi(p^2)$. Work is in progress on this problem both in the field-theoretic framework and by ordinary perturbation theory.

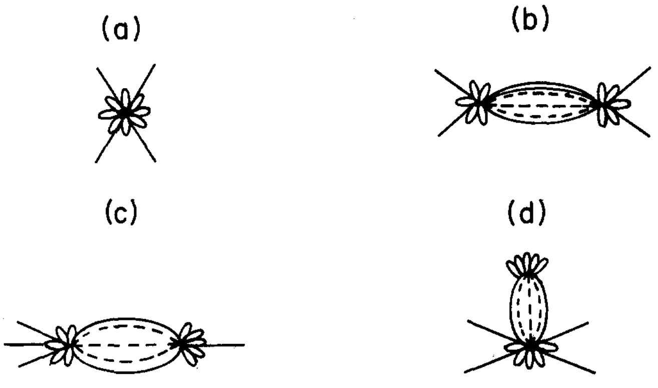


FIG. 4. Contributions to the four-point function. (a) Single-vertex diagram. (b)–(d) Double-vertex diagrams.

B. $d=0$ (one-dimensional integrals)

In Eq. (3.2) the apparent singularity at $\alpha=0$ is canceled by a zero of Ω_d, and, in fact, $\Omega_d/d\to 1$ as $d\to 0$. Thus

$$I_0=\int_0^1 dz\int_0^1\frac{dt}{t}(1-t)^{1/2}\ln(1-zt)\,, \tag{3.11}$$

which becomes

$$I_0=\int_0^1\frac{dt}{t^2}(1-t)^{1/2}[(1-t)\ln(1-t)+t]\,, \tag{3.12}$$

on performing the z integration. Note that this is the second integral occurring in Eq. (3.1) evaluated at $z=1$. It is easily evaluated in terms of $\psi'(\frac{3}{2})$, namely,

$$I_0=\tfrac{3}{2}\psi'(\tfrac{3}{2})-1\,. \tag{3.13}$$

Again putting $\mu=0$, $\lambda=\frac{1}{2}$, $m=M=1$, and including the $O(\delta)$ contribution, we obtain

$$W[0]=\tfrac{1}{2}\delta\psi-\tfrac{1}{4}\delta^2(2\psi(\tfrac{3}{2})+\tfrac{1}{2}\psi'(\tfrac{3}{2}))\,, \tag{3.14}$$

which are the first two terms in the expansion[1] of $-\ln Z$, where

$$Z=\int_{-\infty}^{\infty}\frac{dx}{\sqrt{\pi}}e^{-(x^2)^{1+\delta}}\,.$$

We can evaluate I_0 in an alternative way that generalizes to negative even dimensions. Define

$$F(x)=\int_0^1\frac{dt}{t^2}(1-t)^{1/2}[(1-zt)\ln(1-zt)+zt]\,. \tag{3.15}$$

Then

$$I_d=\Omega_d\int_0^1 dx\,x^{d-1}F(x)\,. \tag{3.16}$$

Integrating by parts gives

$$I_d=\frac{\Omega_d}{d}\left\{[x^dF(x)]_0^\infty-\int_0^\infty dx\,x^dF'(x)\right\}. \tag{3.17}$$

Taking the limit as $d\to 0$ from above, the first term in { } vanishes and we get

$$I_0=-\int_0^1 dx\,F'(x)=F(0)\,, \tag{3.18}$$

as remarked above ($x=0$ corresponds to $z=1$).

C. Negative even dimensions $d=-2n$

The reason that these dimensions are rather special is that in such cases the factor Ω_d has a zero arising from the pole in $\Gamma(d/2)$. Then I_d can be evaluated by successive integration by parts until a denominator $d+2n$ is produced to cancel the zero.

For example, near $d=-2$, Ω_d has the behavior $\Omega_d\sim-(1/\pi)(d+2)$. Thus

$$I_d\sim-\frac{1}{\pi}(d+2)\int_0^1 dx\,x^dF(x)\,. \tag{3.19}$$

On integrating by parts three times and keeping $d>0$ until the final step one obtains

$$I_{-2}=(1/2\pi)F''(0)\,. \tag{3.20}$$

Now

$$F'(x)=-z'\int_0^1\frac{dt}{t}(1-t)^{1/2}\ln(1-zt) \tag{3.21}$$

so that

$$F''(0)=-z''(0)\int_0^1\frac{dt}{t}(1-t)^{1/2}\ln(1-t)+2(z'(0))^2\,. \tag{3.22}$$

Recall from (3.5) that $z(x)=y(\tilde{x})^2$, where

$$y(\tilde{x})=[2^{d/2}/\Gamma(\nu)]\tilde{x}^\nu K_\nu(\tilde{x}) \tag{3.23}$$

and $\tilde{x}=mx$. From the formula[4]

$$\frac{d}{d\tilde{x}}(\tilde{x}^\nu K_\nu(\tilde{x}))=-\tilde{x}(\tilde{x}^{\nu-1}K_{\nu-1}(\tilde{x}))\,, \tag{3.24}$$

it is readily seen that $z'=0$, for $\nu>1$, i.e., for arbitrary negative d. Moreover

$$z''(0)=2m^2/d\,.$$

The integral occurring in (3.22) is just $-\psi'(\frac{3}{2})$; thus

$$I_{-2}=-(1/2\pi)m^2\psi'(\tfrac{3}{2})\,. \tag{3.25}$$

For $d=-4$ one needs the fourth derivative of $F(x)$ evaluated at $x=0$,

$$I_{-4}=(1/4\pi^2)F^{(4)}(0)\,. \tag{3.26}$$

Introducing the notation

$$J_L(x)=\int_0^1\frac{dt}{t}(1-t)^{1/2}\ln(1-zt) \tag{3.27}$$

and

$$J_r(x)=\int_0^1\frac{dt}{t}(1-t)^{1/2}\frac{t^r}{(1-zt)^r}\quad(r\geqslant 1)\,, \tag{3.28}$$

$F^{(4)}(x)$ is given by

$$F^{(4)}(x)=-z^{(4)}J_L+(4z'z^{(3)}+3(z'')^2)J_1+6(z')^2z''J_2+2(z')^4J_3\,. \tag{3.29}$$

Greater care is now needed since the integrals J_2 and J_3 are divergent as $x\to 0$, like x^{-1} and x^{-3}, respectively. However, there are sufficient factors of z' in front of these integrals to cancel the singularities, and

$$F^{(4)}(0)=z^{(4)}(0)\psi'(\tfrac{3}{2})+6(z''(0))^2\,, \tag{3.30}$$

inserting the values $J_L^{(0)}=-\psi'(\frac{3}{2})$, $J_1^{(0)}=2$.

In this case, $z''(0)=-\frac{1}{2}m^2$, while

$$z^{(4)}(0)=6m^4/[d(d+2)]=3m^4/4.$$

Thus

$$I_{-4}=(1/12\pi^2)\,F^{(4)}(0)=(M^4/16\pi^2)(\psi'(\tfrac{3}{2})+2)\,. \tag{3.31}$$

When we go to $d=-6$ a further complication arises: the divergent integrals $J_2^{(0)}-J_5^{(0)}$ are no longer individually tamed by the accompanying factors of z' and $z^{(3)}$. However, the coefficients are such that the divergences collectively cancel, with a finite result that again depends on just J_L and J_1. The calculation becomes rapidly more involved for $d<-6$, and we are unable to give a general form. As mentioned above, a general expression for negative even dimensions may well be equivalent to solving the problem for all positive dimensions as well.[5]

IV. RENORMALIZABLE THEORIES: $d=3,4$

In dimensions greater than or equal to 2, $\Delta(0)$ does not exist, and we must regularize the theory prior to renormalization. A Lorentz covariant regularization prescription in Euclidean configuration space is to impose a short-distance cutoff a on the radial variable x, equivalent to a momentum-space cutoff $\Lambda = 1/a$. Then $\Delta(0)$ is replaced by

$$\Delta(a) = [\Gamma(|\nu|)/4\pi](\pi a^2)^{-|\nu|}, \tag{4.1}$$

where we have used the limiting form of the Bessel function for small argument. Its index is $|\nu| = d/2 - 1$, for $d \geqslant 2$. Correspondingly, the variable z occurring in the integrals for the obvious Green's functions is replaced by

$$\tilde{z}(x) = \left(\frac{\Delta(x)}{\Delta(a)}\right)^2 = \left(\frac{2}{\Gamma(|\nu|)}\left(\frac{ma^2}{2x}\right)^{|\nu|} K_{|\nu|}(mx)\right)^2. \tag{4.2}$$

It seems impossible to perform an asymptotic analysis of the integrals for general $d+2$. Instead, each dimension must be treated separately. We now discuss in turn the cases $d=3$ and $d=4$. The borderline case $d=2$ ($|\nu|=0$) is considerably more difficult [Eq. (4.1) gives a logarithmic rather than power behavior], and is reserved for a future publication.

A. $d=3$

The relevant Bessel function in this case is $K_{1/2}$, which has a rather simple form, and (4.2) becomes just[7]

$$\tilde{z} = (a^2/x^2)e^{-2mx}, \tag{4.3}$$

and $\Delta(a)$ reduces to $\Delta(a) = 1/(4\pi a)$.

Consider first the various contributions to the proper self-energy part $\Pi(p^2)$. For $\Pi^{(a)}$ in Eq. (2.19), we merely note that S diverges like $\ln(aM)$.

In the expression for $\Pi^{(c)}$ in Eq. (2.23) we encounter two integrals, the first of which is very easy to evaluate:

$$\Pi_1^{(c)} = 4\delta^2\lambda^2 M^4 \Delta(a) S \Omega_d \int_a^\infty dx\, x^{d-1}\tilde{z}$$
$$= 2\delta^2\lambda^2 M^4 \frac{a}{m} S. \tag{4.4}$$

Contrary to appearances (4.4) does not vanish as $a \to 0$. One must remember that λ is the unrenormalized coupling constant, which tends to infinity in such a way[4] as to ensure that the renormalized coupling constant is finite. In fact, $\Pi_1^{(c)}$ represents the dominant[8] contribution to $\Pi^{(c)}$.

To estimate the second integral in (2.23),

$$\Pi_2^{(c)} = 4\delta^2\lambda^2 M^4 \Delta(a)\Omega_d \int_a^\infty dx\, x^{d-1}$$
$$\times \int_0^1 \frac{dt}{t^2}(1-t)^{1/2}[\ln(1-\tilde{z}t) + \tilde{z}t], \tag{4.5}$$

we first perform a scaling, setting $x = y/m$. Then we expand the logarithm and perform the t integration to obtain the beta-function series

$$-\sum_{n=2}^\infty \frac{1}{n} B\left(n-1, \frac{3}{2}\right)\tilde{z}^n,$$

where $\tilde{z} = (\epsilon^2/y^2)e^{-2y}$, with $\epsilon = am$. Each term of the series gives a contribution of order ϵ^3 under the y integration, and

$$\Pi_2^{(c)} \sim -4\delta^2\lambda^2 M^4 a^2 \sum_{n=2}^\infty \frac{1}{n(2n-3)} B\left(n-1, \frac{3}{2}\right), \tag{4.6}$$

which is suppressed by a power of am relative to $\Pi_1^{(c)}$. By taking partial fractions and rewriting the beta function as an integral, the n summation in (4.6) can be performed:

$$\sum_{n=2}^\infty \frac{1}{n} B\left(\frac{3}{2}, n-1\right)$$
$$= -\int_0^1 \frac{dt}{t^2}(1-t)^{1/2}(t + \ln(1-t)) = 1 - \frac{1}{2}\psi'\left(\frac{3}{2}\right), \tag{4.7}$$

$$\sum_{n=2}^\infty \frac{1}{2n-3} B\left(\frac{3}{2}, n-1\right)$$
$$= -\int_0^1 \frac{dt}{2\sqrt{t}}(1-t)^{1/2} \ln\left(\frac{1+\sqrt{t}}{1-\sqrt{t}}\right)$$
$$= 2\,\mathrm{Cl}(\pi/2) - 1, \tag{4.8}$$

where $\mathrm{Cl}(\theta)$ is the Clausen function,[4]

$$\mathrm{Cl}(\theta) = -\int_0^\theta \ln\left(2\sin\frac{1}{2}\theta'\right)d\theta' = \sum_{n=1}^\infty \frac{\sin n\theta}{n^2}.$$

We obtain

$$\Pi_2^{(c)} \sim -\tfrac{4}{3}\delta^2\lambda^2 M^4 a^2 [4\,\mathrm{Cl}(\pi/2) - 3 + \tfrac{1}{2}\psi'(\tfrac{3}{2})]. \tag{4.9}$$

Turning now to $\Pi^{(b)}$ in Eq. (2.21), we expand the exponential $e^{ip\cdot x}$ to second order in p, which is all that is needed for the renormalization procedure. Under symmetric integration the linear term vanishes, while $(p\cdot x)^2$ is replaced by p^2x^2d.

The asymptotic behavior of the first term,

$$\Pi_2^{(b)} = -4\delta^2\lambda^2 M^4 \int_a^\infty dx\, x^{d-1}\tilde{z}^{1/2}$$
$$\times \int_0^1 \frac{dt}{t}(1-t)^{1/2}\ln(1-\tilde{z}t), \tag{4.10}$$

is easily evaluated by the same techniques. In this case the first term of the beta-function series dominates, and we obtain

$$\Pi_1^{(b)} \sim \tfrac{8}{3}\delta^2\lambda^2 M^4 a^2 \ln(am). \tag{4.11}$$

In a similar fashion we obtain

$$\Pi_2^{(b)} \sim -\tfrac{8}{81}\delta^2\lambda^2 M^4 a^2 p^2/m^2 \tag{4.12}$$

for the asymptotic behavior of the second term,

$$\Pi_2^{(b)} = \frac{4}{d}\delta^2\lambda^2 M^4 \int_a^\infty dx\, x^{d+1}p^2\tilde{z}^{1/2}$$
$$\times \int_0^1 \frac{dt}{t}(1-t)^{1/2}\ln(1-\tilde{z}t). \tag{4.13}$$

The integrals for the four-point function, Eqs. (2.25)–(2.27), are estimated similarly. Thus

$$\Gamma_4^{(b)} \sim 48\pi\delta^2\lambda^2 M^4 a^2/m \tag{4.14}$$

and

$$\Gamma_4^{(c)} \sim \tfrac{128}{3}\pi\delta^2\lambda^2 M^4 a^3 \ln(am). \tag{4.15}$$

The first integral of $\Gamma_4^{(d)}$ is proportional to that evaluated in (4.4),

$$-8\delta^2\lambda^2 M^4 S\Omega_3 \int_a^\infty dx\, x^2\bar{z} = -16\pi\delta^2\lambda^2 M^4 \frac{a^2}{m} S, \tag{4.16}$$

and is the dominant term unless canceled by a mass counterterm from Fig. 4(a). The second integral is proportional to a^3,

$$8\delta^2\lambda^2 M^4\Omega_3 \int_a^\infty dx\, x^2\bar{z}^2 \int_0^1 \frac{dt(1-t)^{1/2}}{1-\bar{z}t} \sim 32\pi\delta^2\lambda^2 M^4 a^3(\mathrm{Cl}(\pi/2)-1). \tag{4.17}$$

B. $d=4$

Now $\Delta(a)$ is $1/(2\pi a)^2$, but $\bar{z}$ does not have such a simple form, since the relevant Bessel function is K_1. Its behavior for small argument is $K_1(y) \sim 1/y$, so we write

$$K_1(y) = (1/y) f_1(y), \tag{4.18}$$

so that

$$\bar{z}(x) = ((a^2/y^2) f_1(y))^2. \tag{4.19}$$

The function f_1 plays the role that e^{-y} took for the case $d=3$. Although a much more complicated function, it falls off exponentially for large y and behaves like[4]

$$f_1(y) = 1 + O(y^2 \ln y), \tag{4.20}$$

for small y, so that $f_1'(0)$ exists. These properties are sufficient for us to be able to derive the asymptotic forms of the integrals for Π and Γ_4 using integration by parts as we did for the exponential.

The first integral in the expression for $\Pi^{(c)}$ is again the dominant one: it now gives

$$\Pi_1^{(c)} = 4\delta^2\lambda^2 M^4 \Delta(a) S\Omega_4 \int_a^\infty dx\, x^3\bar{z} \sim 2\delta^2\lambda^2 M^4 S a^2 \ln(am), \tag{4.21}$$

whereas

$$\Pi_2^{(c)} \sim \tfrac{1}{2}\delta^2\lambda^2 M^4 a^2(\tfrac{3}{2}\Psi'(\tfrac{3}{2})-1), \tag{4.22}$$

using the result

$$\sum_{n=2}^\infty \frac{1}{n-1} B\left(\frac{3}{2}, n-1\right) = -\int_0^1 \frac{dt}{t}(1-t)^{1/2}\ln(1-t) = \Psi'\left(\frac{3}{2}\right). \tag{4.23}$$

For the first two terms in the expansion of $\Pi^{(b)}$ in powers of p^2 we obtain

$$\Pi_1^{(b)} \sim -2\delta^2\lambda^2 M^4 a^2(2\,\mathrm{Cl}(\pi/2) - 1 - \tfrac{1}{2}\Psi'(\tfrac{3}{2})) \tag{4.24}$$

and

$$\Pi_2^{(b)} = \tfrac{1}{3}\delta^2\lambda^2 M^4 p^2 a^2 \ln(am). \tag{4.25}$$

Turning to the four-point function, the asymptotic forms of Eqs. (2.25) and (2.26) are now

$$\Gamma_4^{(b)} \sim -48\pi^2\delta^2\lambda^2 M^4 a^4 \ln(am) \tag{4.26}$$

and

$$\Gamma_4^{(c)} \sim -32\pi^2\delta^2\lambda^2 M^4 a^4(2\,\mathrm{Cl}(\pi/2)-1). \tag{4.27}$$

The first integral of $\Gamma_4^{(d)}$ is proportional to that given in (4.19),

$$-8\delta^2\lambda^2 M^4 S\Omega_4 \int_a^\infty dx\, x^3\bar{z} \sim -32\pi^2\delta^2 M^4 a^4 S \ln(am), \tag{4.28}$$

and finally the last integral is

$$8\delta^2\lambda^2 M^4\Omega_4 \int_a^\infty dx\, x^3\bar{z}^2 \int_0^1 \frac{dt(1-t)^{1/2}}{1-\bar{z}t} \sim 4\pi^2\delta^2\lambda^2 M^4 a^4 \Psi'(\tfrac{3}{2}). \tag{4.29}$$

V. DISCUSSION

In the preceding sections we have shown how to sum and differentiate the Feynman diagrams arising in the δ expansion and to evaluate them for certain dimensions $d<2$, where no regularization is required. For $d=3,4$ we introduced a short-distance cutoff a and identified the leading contributions as $a \to 0$. Points that remain for further investigation are the borderline case $d=2$, the search for general expressions for $d=-2m$, and an incorporation of the full momentum dependence $e^{ip\cdot x}$. Extension of the field theory expansion to $O(\delta^3)$ is a very difficult task.

In view of the strong indications of triviality of $\lambda(\phi^4)_4$, reinforced by the application of the above results in Ref. 3, this paper should be regarded as a laying the groundwork for future developments in tackling more realistic, nontrivial four-dimensional theories. Some initial progress[9] has been made in that direction in the context of models of Nambu–Jona-Lasinio[10] type. The ultimate aim must, of course, be to incorporate gauge fields.

The numerical accuracy of the δ expansion has been shown[2] to be very good in $d=0$ in the massless case, $\mu=0$. The same is true[11] for massive theories, with both $\mu^2>0$ and $\mu^2<0$. This gives one hope that the δ expansion may be used to give meaningful information about such nonperturbative phenomena as phase transitions, confinement, and spontaneous symmetry breaking.

[1] C. M. Bender, K. A. Milton, M. Moshe, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Rev. Lett. **58**, 2615 (1987).

[2] C. M. Bender, K. A. Milton, M. Moshe, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Rev. D **37**, 1472 (1988).

[3] C. M. Bender and H. F. Jones, Imperial College preprint TP/87—88/11.

[4] See, for example, M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[5] I. G. Halliday and R. W. Ricotta, Imperial College preprint TP/86-87-13; G. V. Dunne and I. G. Halliday, Imperial College preprint TP/86-87/110.

[6] R. Barbieri, J. A. Mignaco, and E. Remiddi, Nuovo Cimento A **11**, 824 (1972).

[7] Here we are assuming that $am \ll 1$, i.e., that the shifted bare mass m is much less than the momentum cutoff $\Lambda = 1/a$. This is case 1 of Ref. 3. The asymptotics of the integrals in the opposite case (case 3), $ma \gg 1$, can be obtained by straightforward scaling.

[8] This contribution is precisely canceled by a mass insertion in Fig. 3(a) if one adopts the counterterm procedure for mass renormalization.

[9] C. M. Bender and H. F. Jones, work in progress.

[10] Y. Nambu and J. Jona-Lasinio, Phys. Rev. **122**, 345 (1961).

[11] H. F. Jones and M. Monoyios, to be published.

On the Painlevé property of nonlinear field equations in 2+1 dimensions: The Davey–Stewartson system

R. A. Leo, G. Mancarella, G. Soliani, and L. Solombrino
Dipartimento di Fisica dell'Universitá di Lecce, 73100 Lecce, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, 73100 Lecce, Italy

(Received 8 June 1988; accepted for publication 17 August 1988)

With the purpose of clarifying some aspects of the complete integrability of nonlinear field equations, a singular-point analysis is performed of the Davey–Stewartson system, which can be considered as an extension in 2 + 1 dimensions of the nonlinear Schrödinger equation. It is found that the system under consideration possesses the Painlevé property and allows a set of Bäcklund transformations obtained by truncating the series expansions of the solutions about the singularity manifold.

I. INTRODUCTION

The search for possible connections among different approaches to nonlinear field equations (NFE) may help to clarify some aspects concerning the complete integrability of the equations under consideration.

For NFE in 1 + 1 dimensions, encouraging results in this direction have been achieved recently in Refs. 1–3. Precisely, the authors of Ref. 1 find a link between the Bäcklund transformations of certain NFE derived from the Weiss–Tabor–Carnevale (WTC) Painlevé analysis,[4] and those obtained by Hirota's technique.[5–8] In Ref. 2, the role of the Estabrook–Wahlquist (EW) prolongation scheme[9] is examined in terms of the WTC procedure. In Ref. 3, it is shown that symmetries and recurrence operators for NFE can be obtained by the Painlevé expansion.

Following the above-mentioned line of reasoning, one could carry out a singular-point analysis in the more general case of NFE in 2 + 1 dimensions, keeping in mind the program of establishing a possible connection between the Painlevé property[4] and Hirota's method. According to this order of ideas, in this paper we study the Davey–Stewartson (DS) system[10]

$$iQ_t + (\tfrac{1}{2})(Q_{xx} + Q_{yy}) = -\sigma|Q|^2Q + qQ, \tag{1.1a}$$

$$q_{xx} - q_{yy} = 2\sigma(|Q|^2)_{xx}, \tag{1.1b}$$

where $Q = Q(x,y,t)$, $q = q(x,y,t)$, $\sigma = \pm 1$, and subscripts denote partial derivatives.

Equations (1.1) are of physical interest, since they describe the propagation of two-dimensional water waves of finite depth. The DS equations were studied in different theoretical frameworks. For example, the initial value problem associated with them, was linearized by Fokas and Ablowitz,[11] while Champagne and Winternitz,[12] by means of a symmetry reduction technique, discovered that Eqs. (1.1) possess a loop algebra structure.

In Sec. II a set of Bäcklund transformations for Eqs. (1.1) are determined by applying a singular-point analysis. In this context, some propositions are also proved concerning the invariance property of certain relations under the Möbius group. In Sec. III a bilinear formulation of the DS system is used to give an N-soliton solution. Finally, Sec. IV contains some conclusions, while in the Appendix details of the calculation are reported.

II. BÄCKLUND TRANSFORMATIONS VIA THE SINGULAR-POINT ANALYSIS

Informally, one says that a partial differential equation possesses the Painlevé property when its solutions are single-valued about the movable singularity manifold.[4] The reader interested in formal mathematical preliminaries is referred to the wide bibliography quoted, for example, in Refs. 1 and 4.

In order to perform the Painlevé analysis of Eqs. (1.1), it is convenient to start from the system

$$iQ_t + \tfrac{1}{2}(Q_{xx} + Q_{yy}) = -\sigma Q^2R + qQ, \tag{2.1a}$$

$$iR_t - \tfrac{1}{2}(R_{xx} + R_{yy}) = \sigma R^2Q - qR, \tag{2.1b}$$

$$q_{xx} - q_{yy} = 2\sigma(QR)_{xx}, \tag{2.1c}$$

which coincides with (1.1) when $R = R(x,y,t) \equiv Q^*$.

Now we make the ansatz that the variables Q, R, and q can be expanded about the singularity manifold $\phi(x,y,t) = 0$ as

$$Q = \phi^\alpha \sum_{k=0}^{\infty} u_k\phi^k, \tag{2.2a}$$

$$R = \phi^\beta \sum_{k=0}^{\infty} w_k\phi^k, \tag{2.2b}$$

$$q = \phi^\gamma \sum_{k=0}^{\infty} v_k\phi^k, \tag{2.2c}$$

where $\phi = \phi(x,y,t)$, $u_k = u_k(x,y,t)$, $w_k = w_k(x,y,t)$, and $v_k = v_k(x,y,t)$ are analytic functions in a neighborhood of $\phi = 0$, and α, β, and γ are integers.

Inserting (2.2) in Eqs. (2.1), a leading-order analysis uniquely provides $\alpha = \beta = -1$ and $\gamma = -2$. Consequently, Eqs. (2.1) yield the recursion relations

 0022-2488/88/122666-06$02.50

$$[\tfrac{1}{2}(\phi_x^2+\phi_y^2)(k-1)(k-2)+2\sigma u_0w_0-v_0]u_k+\sigma u_0^2w_k-u_0v_k$$
$$=-i\phi_t(k-2)u_{k-1}-iu_{k-2,t}-\tfrac{1}{2}(\phi_{xx}+\phi_{yy})(k-2)u_{k-1}-(k-2)(\phi_xu_{k-1,x}+\phi_yu_{k-1,y})$$
$$-\tfrac{1}{2}(u_{k-2,xx}+u_{k-2,yy})-\sigma\sum_{j,l=0}^{k-1}{}'\,u_jw_lu_{k-(j+l)}+\sum_{j=1}^{k-1}u_jv_{k-j}, \tag{2.3a}$$
$$[\tfrac{1}{2}(\phi_x^2+\phi_y^2)(k-1)(k-2)+2\sigma w_0u_0-v_0]w_k+\sigma w_0^2u_k-w_0v_k$$
$$=i\phi_t(k-2)w_{k-1}+iw_{k-2,t}-\tfrac{1}{2}(\phi_{xx}+\phi_{yy})(k-2)w_{k-1}-(k-2)(\phi_xw_{k-1,x}+\phi_yw_{k-1,y})$$
$$-\tfrac{1}{2}(w_{k-2,xx}+w_{k-2,yy})-\sigma\sum_{j,l=0}^{k-1}{}'\,w_ju_lw_{k-(j+l)}+\sum_{j=1}^{k-1}w_jv_{k-j}, \tag{2.3b}$$
$$(\phi_x^2-\phi_y^2)(k-2)(k-3)v_k-2\sigma\phi_x^2(k-2)(k-3)(w_0u_k+u_0w_k)$$
$$=(\phi_{yy}-\phi_{xx})(k-3)v_{k-1}-2(k-3)(\phi_xv_{k-1,x}-\phi_yv_{k-1,y})-(v_{k-2,xx}-v_{k-2,yy})$$
$$+2\sigma\Big[\phi_x^2(k-3)\sum_{j=2}^{k-1}(j-1)(u_jw_{k-j}+w_ju_{k-j})+2\phi_x(k-3)\sum_{j=0}^{k-1}(u_jw_{k-1-j,x}+w_ju_{k-1-j,x})$$
$$+\phi_{xx}\sum_{j=0}^{k-1}(j-1)(w_ju_{k-1-j}+u_jw_{k-1-j})+\frac{1}{2}\sum_{j=0}^{k-2}(w_ju_{k-2-j}+u_jw_{k-2-j})_{xx}\Big], \tag{2.3c}$$

where the indexes j,l in the summation Σ' are such that $0<j+l\leqslant k$.

Now we observe that Eqs. (2.3) can be considered as an inhomogeneous system in the unknowns u_k, w_k, and v_k. The determinant D of the coefficients of such a system, namely,

$$D=\tfrac{1}{4}(\phi_x^2-\phi_y^2)(\phi_x^2+\phi_y^2)\times k(k-2)(k-3)^2(k-4)(k+1), \tag{2.4}$$

vanishes for

$$k=-1,0,2,3,3,4, \tag{2.5}$$

which are therefore the resonances of the recursion relations (2.3), i.e., those values of k that correspond to points where arbitrary functions of (x,y,t) are introduced into the expansions (2.2).

For $k=0,1,2,3,4$, Eqs. (2.3) give rise to the following constraints.

For $k=0$,

$$\phi_x^2+\phi_y^2=-\sigma u_0w_0+v_0, \tag{2.6a}$$

$$(\phi_x^2-\phi_y^2)v_0=2\sigma\phi_x^2u_0w_0. \tag{2.6b}$$

For $k=1$,

$$i\phi_tu_0+\tfrac{1}{2}(\phi_{xx}+\phi_{yy})u_0+\phi_xu_{0x}+\phi_yu_{0y}=\sigma(2u_0w_0u_1+u_0^2w_1)-u_0v_1-u_1v_0, \tag{2.7a}$$

$$-i\phi_tw_0+\tfrac{1}{2}(\phi_{xx}+\phi_{yy})w_0+\phi_xw_{0x}+\phi_yw_{0y}=\sigma(2w_0u_0w_1+w_0^2u_1)-w_0v_1-w_1v_0, \tag{2.7b}$$

$$(\phi_x^2-\phi_y^2)v_1-2(\phi_xv_{0x}-\phi_yv_{0y})-(\phi_{xx}-\phi_{yy})v_0=-2\sigma[2\phi_x(u_0w_0)_x+\phi_{xx}u_0w_0-\phi_x^2(w_0u_1+u_0w_1)]. \tag{2.7c}$$

For $k=2$,

$$(v_0-2\sigma u_0w_0)u_2-\sigma u_0^2w_2+u_0v_2=iu_{0t}+\tfrac{1}{2}(u_{0xx}+u_{0yy})+2\sigma u_1w_1u_0+\sigma u_1^2w_0-u_1v_1, \tag{2.8a}$$

$$-\sigma w_0^2u_2+(v_0-2\sigma u_0w_0)+w_0v_2=-iw_{0t}+\tfrac{1}{2}(w_{0xx}+w_{0yy})+2\sigma w_1u_1w_0+\sigma w_1^2u_0-w_1v_1, \tag{2.8b}$$

$$-2\phi_xv_{1x}-\phi_{xx}v_1+v_{0xx}+2\phi_yv_{1y}+\phi_{yy}v_1-v_{0yy}=2\sigma[(u_0w_0)_{xx}-2\phi_x(u_0w_{1x}+w_0u_{1x})-2\phi_x(w_{0x}u_1+u_{0x}w_1)-\phi_{xx}(w_0u_1+u_0w_1)]. \tag{2.8c}$$

For $k=3$,

$$(v_0-2\sigma u_0w_0-\phi_x^2-\phi_y^2)u_3-\sigma u_0^2w_3+u_0v_3=i\phi_tu_2+\tfrac{1}{2}(\phi_{xx}+\phi_{yy})u_2+\phi_xu_{2x}+\phi_yu_{2y}+2\sigma(u_0w_1+u_1w_0)u_2+2\sigma u_0u_1w_2-u_2v_1+[iu_{1t}+\tfrac{1}{2}(u_{1xx}+u_{1yy})+\sigma u_1^2w_1-u_1v_2], \tag{2.9a}$$

$$-\sigma w_0^2u_3+(v_0-2\sigma u_0w_0-\phi_x^2-\phi_y^2)w_3+w_0v_3=-i\phi_tw_2+\tfrac{1}{2}(\phi_{xx}+\phi_{yy})w_2+\phi_xw_{2x}+\phi_yw_{2y}+2\sigma(w_0u_1+w_1u_0)w_2+2\sigma w_0w_1u_2-w_2v_1+[-iw_{1t}+\tfrac{1}{2}(w_{1xx}+w_{1yy})+\sigma w_1^2u_1-w_1v_2], \tag{2.9b}$$

$$v_{1xx}-v_{1yy}=2\sigma(u_0w_1+w_0u_1)_{xx}. \tag{2.9c}$$

For $k=4$,

$$-2i\phi_t u_3 - iu_{2t} - (\phi_{xx}+\phi_{yy})u_3 - 2\phi_x u_{3x} - 2\phi_y u_{3y} - \tfrac{1}{2}(u_{2xx}+u_{2yy}) - \sigma[2(u_0w_1+w_0u_1)u_3 + 2u_0u_1w_3 + 2u_0u_2w_2 + 2u_1w_1u_2 + u_1^2w_2 + w_0u_2^2] + u_1v_3 + u_2v_2 + u_3v_1 = [3(\phi_x^2+\phi_y^2) + 2\sigma u_0w_0 - v_0]u_4 + \sigma u_0^2 w_4 - u_0v_4, \quad (2.10a)$$

$$2i\phi_t w_3 + iw_{2t} - (\phi_{xx}+\phi_{yy})w_3 - 2\phi_x w_{3x} - 2\phi_y w_{3y} - \tfrac{1}{2}(w_{2xx}+w_{2yy}) - \sigma[2(u_0w_1+w_0u_1)w_3 + 2w_0w_1u_3 + 2w_0w_2u_2 + 2u_1w_1w_2 + w_1^2u_2 + u_0w_2^2] + w_1v_3 + w_2v_2 + w_3v_1 = [3(\phi_x^2+\phi_y^2) + 2\sigma u_0w_0 - v_0]w_4 + \sigma w_0^2 u_4 - w_0v_4, \quad (2.10b)$$

$$(\phi_{xx}-\phi_{yy})v_3 + 2(\phi_x v_{3x} - \phi_y v_{3y}) + v_{2xx} - v_{2yy} - 4\sigma[\phi_x^2(u_2w_2+u_3w_1+u_1w_3) + \phi_x(u_0w_{3x}+u_1w_{2x}+u_2w_{1x}+u_3w_{0x} + w_0u_{3x}+w_1u_{2x}+w_2u_{1x}+w_3u_{0x}) + \tfrac{1}{2}\phi_{xx}(w_2u_1+u_2w_1+w_3u_0+u_3w_0)] - 2\sigma[(w_0u_2+w_2u_0)_{xx} + (u_1w_1)_{xx}] = 2\sigma(w_0v_0u_4 + u_0v_0w_4 - u_0w_0v_4). \quad (2.10c)$$

From Eqs. (2.6) we find

$$u_0w_0 = \sigma(\phi_x^2 - \phi_y^2), \quad (2.11a)$$

$$v_0 = 2\phi_x^2, \quad (2.11b)$$

where the quantities $\phi_x^2+\phi_y^2$ and $\phi_x^2-\phi_y^2$ are taken different from zero.

After some manipulations, Eqs. (2.7) yield

$$u_1 = [1/(\phi_x^2+\phi_y^2)][-i\phi_t u_0 + \tfrac{1}{2}u_0(\phi_{xx}+\phi_{yy}) - \phi_x u_{0x} - \phi_y u_{0y}], \quad (2.12a)$$

$$w_1 = [1/(\phi_x^2+\phi_y^2)][i\phi_t w_0 + \tfrac{1}{2}w_0(\phi_{xx}+\phi_{yy}) - \phi_x w_{0x} - \phi_y w_{0y}], \quad (2.12b)$$

$$v_1 = -2\phi_{xx}, \quad (2.12c)$$

and

$$u_0w_1 + u_1w_0 = \sigma(\phi_{yy} - \phi_{xx}). \quad (2.13)$$

On the other hand, we notice that (2.8c) is identically satisfied, while from (2.8a) and (2.8b) we obtain

$$v_2 = (1/\Delta)[(v_0 - 2\sigma u_0w_0)A + \sigma u_0^2 B] - (u_2/u_0)(v_0 - 3\sigma u_0w_0), \quad (2.14a)$$

$$w_2 = (1/\Delta)(u_0B - Aw_0) + (u_2/u_0)w_0, \quad (2.14b)$$

where u_2 is arbitrary, and

$$\Delta = u_0(\phi_x^2+\phi_y^2), \quad (2.15)$$

$$A = iu_{0t} + \tfrac{1}{2}(u_{0xx}+u_{0yy}) + 2\sigma u_1u_0w_1 + \sigma w_0u_1^2 - u_1v_1, \quad (2.16)$$

$$B = -iw_{0t} + \tfrac{1}{2}(w_{0xx}+w_{0yy}) + 2\sigma u_1w_1w_0 + \sigma u_0w_1^2 - w_1v_1. \quad (2.17)$$

Exploring Eqs. (2.9), we infer that (2.9c) is identically satisfied, while the arbitrary choice of the three functions u_3, w_3, and v_3 implies the compatibility condition

$$w_0H = u_0K, \quad (2.18)$$

where H and K are expressed by the right-hand sides of (2.9a) and (2.9b), respectively.

Condition (2.18) entails

$$v_3 = H/u_0 + \sigma(w_0u_3 + u_0w_3), \quad (2.19)$$

where u_3 and w_3 are arbitrary functions.

Finally, concerning the resonance occurring for $k=4$, we notice that Eqs. (2.10) lead to the compatibility condition

$$Fw_0 + Gu_0 = \sigma E, \quad (2.20)$$

where F, G, and E are the left-hand sides of (2.10a), (2.10b), and (2.10c).

We have checked that the relations (2.18) and (2.20), corresponding, respectively, to the resonances $k=3$ and $k=4$, are identically satisfied. From the above considerations we deduce that the DS system has the Painlevé property.

At this point, we are ready to show that the expansions (2.2) can be truncated to yield a Bäcklund transformation for the DS system (1.1). In doing so, let us assume $u_2 = u_3 = w_3 = u_4 = 0$ and impose that $w_2 = 0$, and

$$(u_1, w_1, v_2) \in S, \quad (2.21)$$

where S is the manifold of the solutions of Eqs. (2.1), i.e.,

$$iu_{1t} + \tfrac{1}{2}(u_{1xx}+u_{1yy}) + \sigma u_1^2 w_1 - u_1v_2 = 0, \quad (2.22a)$$

$$iw_{1t} - \tfrac{1}{2}(w_{1xx}+w_{1yy}) - \sigma w_1^2 u_1 + w_1v_2 = 0, \quad (2.22b)$$

$$v_{2xx} - v_{2yy} = 2\sigma(u_1w_1)_{xx}. \quad (2.22c)$$

Then we obtain $H=F=G=E=0$, so that $v_3 = v_4 = w_4 = 0$, and

$$v_2 = (1/u_0)[iu_{0t} + \tfrac{1}{2}(u_{0xx}+u_{0yy}) + u_1(\phi_{xx}+\phi_{yy}) + \sigma u_1u_0w_1]. \quad (2.23)$$

Keeping in mind Eqs. (2.3), one easily finds $u_k = w_k = v_k = 0$ for all values of k greater than 4. To conclude, we have the following proposition.

Proposition 2.1: The DS system (1.1) possesses the Bäcklund transformation

$$Q = u_1 + u_0/\phi, \quad (2.24a)$$

$$q = v_2 + \frac{v_1}{\phi} + \frac{v_0}{\phi^2} = v_2 - 2\frac{\partial^2}{\partial x^2}\ln\phi, \quad (2.24b)$$

where $(u_1, v_2) \in S$, u_0, v_0, and v_1 are given by (2.11a) (with $w_0 = u_0^*$), (2.11b), and (2.12c), respectively,[13] and ϕ satisfies the equation

$$u_0[iu_{1t}+\tfrac{1}{2}(u_{1xx}+u_{1yy})]=u_1[iu_{0t}+\tfrac{1}{2}(u_{0xx}+u_{0yy})+\sigma u_0u_1(u_1^*+u_1)-u_1v_1], \tag{2.25}$$

where u_1 is expressed by (2.12a).

Equation (2.25) is obtained combining (2.8a) (with $u_2=w_2=0$ and $w_0=u_0^*$, $w_1=u_1^*$) with (2.22a) (where $w_1=u_1^*$).

Other results coming from the singular-point analysis of Eqs. (1.1) are stated in the following proposition.

Proposition 2.2: Equation (2.25) is invariant under the Möbius group, in the sense that if ϕ obeys Eq. (2.25), then so does ψ, expressed by

$$\psi=(a\phi+b)/(c\phi+d),\quad ad-bc\neq 0. \tag{2.26}$$

This invariant property arises straightforwardly using the constraints (2.7a), (2.11b), (2.12c), and (2.12a).

By a direct calculation we also have the following proposition.

Proposition 2.3: Let (u_1,v_2) be a pair of solutions of Eqs. (2.22) (with $w_j=u_j^*$), where

$$u_1\equiv u_1[\phi]=u_0\left[-i\frac{\phi_t}{\phi_x^2+\phi_y^2}+\frac{\phi_{yy}-\phi_{xx}}{2(\phi_x^2-\phi_y^2)}\right], \tag{2.27a}$$

$$v_2\equiv v_2[\phi]=(1/u_0)[iu_{0t}+\tfrac{1}{2}(u_{0xx}+u_{0yy})+\sigma u_0(2|u_1|^2+u_1^2)+2u_1\phi_{xx}], \tag{2.27b}$$

and

$$u_0=[\sigma(\phi_x^2-\phi_y^2)]^{1/2}. \tag{2.28}$$

If we demand that Eqs. (2.22) be invariant under the Möbius transformation (2.26), then also the pair $\tilde{u}_1\equiv u_1[\psi]$, $\tilde{v}_2\equiv v_2[\psi]$ fulfills Eqs. (2.22), where

$$\tilde{u}_1=u_1+u_0/(\phi+\lambda), \tag{2.29a}$$

$$\tilde{v}_2=v_2-2\partial^2/\partial x^2\ln(\phi+\lambda), \tag{2.29b}$$

and $\lambda=d/c$ is an arbitrary constant.

We notice that Eqs. (2.29) reduce to the Bäcklund transformations (2.24) for $\lambda=0$. Thus the requirement of invariance of the DS system under the Möbius group represents a simple way of introducing a free parameter into the Bäcklund transformations themselves. This is the starting point for trying to build up, in the WTC context, a spectral problem for the given equations.[14]

III. BILINEAR FORMULATION AND *N*-SOLITON SOLUTIONS

The Bäcklund transformations (2.24) derived by the Painlevé analysis of Eqs. (1.1) suggest the way of writing Eqs. (1.1) in the Hirota's bilinear form.[5–8] To this aim, let us set

$$Q=g/f,\quad q=h/f^2, \tag{3.1}$$

where $f=f^*$, and

$$|g|^2=\sigma f^2\left(\frac{\partial^2}{\partial y^2}-\frac{\partial^2}{\partial x^2}\right)\ln f=\frac{1}{2}\sigma(D_y^2-D_x^2)f\cdot f, \tag{3.2a}$$

$$h=-2f^2\frac{\partial^2}{\partial x^2}\ln f=-D_x^2f\cdot f. \tag{3.2b}$$

Inserting the quantities (3.1) in Eqs. (1.1), we find that Eq. (1.1b) is identically satisfied, while Eq. (1.1a) becomes

$$[iD_t+\tfrac{1}{2}(D_x^2+D_y^2)]g\cdot f=0, \tag{3.3}$$

where the operators D are defined by

$$D_x^lD_y^mD_t^ng(x,y,t)\cdot f(x,y,t)=\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial x'}\right)^l\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial y'}\right)^m\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial t'}\right)^n\times g(x,y,t)f(x',y',t')\Big|_{x=x',y=y',t=t'}. \tag{3.4}$$

Equations (3.3) and (3.2a) can be used to find exact N-soliton solutions of the DS system (1.1).[15] Following Hirota's formalism,[16] for the reader's convenience in the Appendix we have checked that Eqs. (1.1) possess the N-envelope-soliton solution

$$Q=A\exp[i\theta](\tilde{g}/f),\quad q=-2\frac{\partial^2}{\partial x^2}\ln f, \tag{3.5}$$

with

$$\tilde{g}=\sum_{\nu=0,1}^{(1)}\exp\left\{\sum_{i=1}^{2N}\nu_i\eta_i+\sum_{i<j}^{(2N)}\nu_i\nu_j\varphi(i,j)\right\}, \tag{3.6a}$$

$$f=\sum_{\mu=0,1}^{(0)}\exp\left\{\sum_{i=1}^{2N}\mu_i\eta_i+\sum_{i<j}^{(2N)}\mu_i\mu_j\varphi(i,j)\right\}, \tag{3.6b}$$

where

$$\theta=kx+ly-\omega t, \tag{3.7}$$

$$\eta_i=K_ix+L_iy-\Omega_it-\eta_i^0, \tag{3.8}$$

$$-i(\Omega_i+i\omega)+\tfrac{1}{2}(K_i+ik)^2+\tfrac{1}{2}(L_i+il)^2=0, \tag{3.9}$$

for $i=1,2,\dots,2N$,

$$\eta_{i+N}=\eta_i^*,\quad \text{for } i=1,2,\dots,N,$$

$$\exp[\varphi(i,j)]=(\sigma A^2)/[(L_i+L_j)^2-(K_i+K_j)^2], \tag{3.10}$$

for $1\leqslant i\leqslant N$ and $N+1\leqslant j\leqslant 2N$ or $N+1\leqslant i\leqslant 2N$ and $1\leqslant j\leqslant N$,

$$\exp[\varphi(i,j)]=(\sigma A^2)^{-1}[(L_i-L_j)^2-(K_i-K_j)^2], \tag{3.11}$$

for $1\leqslant i\leqslant N$ and $1\leqslant j\leqslant N$ or $N+1\leqslant i\leqslant 2N$ and $N+1\leqslant j\leqslant 2N$.

The quantities A, ν_i, μ_i, k, L, ω, K_i, L_i, Ω_i, and η_i^0 are constants. The symbol $\Sigma_{\mu=0,1}^{(m)}$ denotes the summation over all possible combinations of $\mu_i=0,1$, under the condition $\Sigma_{i=1}^{N}(\mu_i-\mu_{i+N})=m$ while $\Sigma_{i<j}^{(2N)}$ means the summation over all possible pairs taken from $2N$ elements with the condition $i<j$.

IV. CONCLUSIONS

With the purpose of looking for possible connections among different methods for studying nonlinear field equations in more than one spatial and one temporal dimension, we have carried out a singular-point analysis of the Davey–Stewartson system (1.1). We have found that (i) Eqs. (1.1) pass the Painlevé test[17] and (ii) the generalized Laurent series (2.1) can be truncated to yield a set of Bäcklund transformations.

As is well known, a crucial role in the singular-point analysis of NFE in 1 + 1 dimensions is played by the invariance property of certain relations under the Möbius group. We remark that this characteristic is exhibited as well in our case, where (2 + 1)-dimensional equations are dealt with. Of special interest is the requirement of invariance under the Möbius group of Eqs. (2.22), which allows us to introduce a free parameter in the Bäcklund transformation [see (2.29)]. This is the basis for trying to formulate a Lax pair for the given equations. Anyway, the important problem of finding Lax pairs for NFE in more than 1 + 1 dimensions within the WTC scheme is only beginning, and deserves further extensive investigations.

Finally, we observe that for many integrable NFE in 1 + 1 dimensions, the Bäcklund transformations obtained in the WTC framework are directly related to those derived by Hirota's method. We point out that this feature is also a property of the nonlinear Schrödinger equation,[1] which can be considered as a reduced version in 1 + 1 dimensions of the Davey–Stewartson system. Hence it should be interesting to look for the existence of a connection of this type also in the case of NFE in 2 + 1 dimensions. This problem will be tackled in the near future.

APPENDIX: DETERMINATION OF THE *N*-SOLITON SOLUTION

Here we show that (3.5) [where $\tilde{g}$ and f are given by (3.6a) and (3.6b)] is an N-envelope-soliton solution for Eqs. (1.1). In doing so, let us insert (3.6a) and (3.6b) into Eqs. (3.3) and (3.2a). We have

$$\sum_{\mu=0,1}^{(0)} \sum_{\nu=0,1}^{(1)} \left\{ \sum_{i=1}^{2N} \left[-i(\Omega_i + i\omega) + ikK_i + ilL_i - \frac{1}{2}k^2 - \frac{1}{2}l^2 \right](\nu_i - \mu_i) + \left(\sum_{i=1}^{2N} K_i(\nu_i - \mu_i) \right)^2 + \left(\sum_{i=1}^{2N} L_i(\nu_i - \mu_i) \right)^2 \right\} \exp\left\{ \sum_{i=1}^{2N} (\nu_i + \mu_i)\eta_i + \sum_{i<j}^{(2N)} (\nu_i\nu_j + \mu_i\mu_j)\varphi(i,j) \right\} = 0 \tag{A1}$$

and

$$\sum_{\mu=0,1}^{(0)} \sum_{\mu'=0,1}^{(0)} \frac{1}{2}\left[\left(\sum_{i=1}^{2N} L_i(\mu_i - \mu_i') \right)^2 - \left(\sum_{i=1}^{2N} K_i(\mu_i - \mu_i') \right) \right] \exp\left\{ \sum_{i=1}^{2N} (\mu_i + \mu_i')\eta_i + \sum_{i<j}^{(2N)} (\mu_i\mu_j + \mu_i'\mu_j')\varphi(i,j) \right\} - \sigma A^2 \sum_{\nu=0,1}^{(1)} \sum_{\nu'=0,1}^{(-1)} \exp\left\{ \sum_{i=1}^{2N} (\nu_i + \nu_i')\eta_i + \sum_{i<j}^{(2N)} (\nu_i\nu_j + \nu_i'\nu_j')\varphi(i,j) \right\} = 0. \tag{A2}$$

By virtue of (3.9) and introducing the quantities

$$P_i = L_i - K_i, \quad Q_i = L_i + K_i, \tag{A3}$$

Eqs. (A1) and (A2) become

$$\sum_{\mu=0,1}^{(0)} \sum_{\nu=0,1}^{(1)} \frac{1}{2}\left\{ \left(\sum_{i=1}^{2N} P_i(\nu_i - \mu_i) \right)^2 + \left(\sum_{i=1}^{2N} Q_i(\nu_i - \mu_i) \right)^2 - \frac{1}{2}\sum_{i=1}^{2N} P_i^2(\nu_i - \mu_i) - \frac{1}{2}\sum_{i=1}^{2N} Q_i^2(\nu_i - \mu_i) \right\} \exp\left\{ \sum_{i=1}^{2N} (\nu_i + \mu_i)\eta_i + \sum_{i<j}^{(2N)} (\nu_i\nu_j + \mu_i\mu_j)\varphi(i,j) \right\} = 0 \tag{A4}$$

and

$$\sum_{\mu=0,1}^{(0)} \sum_{\mu'=0,1}^{(0)} \frac{1}{2}\left[\left(\sum_{i=1}^{2N} P_i(\mu_i - \mu_i') \right)\left(\sum_{j=1}^{(2N)} Q_j(\mu_j - \mu_j') \right) \right] \exp\left\{ \sum_{i=1}^{2N} (\mu_i + \mu_i')\eta_i + \sum_{i<j}^{(2N)} (\mu_i\mu_j + \mu_i'\mu_j')\varphi(i,j) \right\} - \sigma A^2 \sum_{\nu=0,1}^{(1)} \sum_{\nu'=0,1}^{(-1)} \exp\left[\sum_{i=1}^{2N} (\nu_i + \nu_i')\eta_i + \sum_{i<j}^{(2N)} (\nu_i\nu_j + \nu_i'\nu_j')\varphi(i,j) \right] = 0. \tag{A5}$$

Now, according to Hirota's notation,[16] we indicate by $D_1(L,M,L',M')$ and $D_2(L,M,L',M')$ the coefficients of the factor

$$\exp\left[\sum_{i=1}^{L} \eta_i + \sum_{i=1}^{L'} \eta_{i+N} + \sum_{i=L+1}^{L+M} 2\eta_i + \sum_{i=L'+1}^{L'+M'} 2\eta_{i+N} \right]$$

in (A4) and (A5), respectively. Then, putting $\sigma_i = 1 - 2\mu_i$, $\sigma_{i+N} = -1 + 2\mu_{i+N}$, and following a procedure similar to that used in Ref. 16, we find that D_1 and D_2 are different from zero only if $L + L'$ is odd or even, respectively. Furthermore, we have

$$D_1(L,M,L',M') = \text{const}\, \hat{D}_1(\hat{P}_1,\hat{Q}_1,\ldots,\hat{P}_{L+L'},\hat{Q}_{L+L'}),$$
$$D_2(L,M,L',M') = \text{const}\, \hat{D}_2(\hat{P}_1,\hat{Q}_1,\ldots,\hat{P}_{L+L'},\hat{Q}_{L+L'}),$$

with

$$\hat{D}_1 = \sum_{\hat{\sigma}=\pm 1}{}' h_1(\hat{P}_1\hat{\sigma}_1,\hat{Q}_1\hat{\sigma}_1,\ldots,\hat{P}_{L+L'}\hat{\sigma}_{L+L'},\hat{Q}_{L+L'}\hat{\sigma}_{L+L'})\, b(\hat{P}_1,\hat{Q}_1,\hat{\sigma}_1,\ldots,\hat{P}_{L+L'},\hat{Q}_{L+L'},\hat{\sigma}_{L+L'}) \tag{A6}$$

and

$$\hat{D}_2 = \sum_{\hat\sigma=\pm1}{}'' h_2(\hat{P}_1\hat\sigma_1,\hat{Q}_1\hat\sigma_1,\ldots,\hat{P}_{L+L'}\cdot\hat\sigma_{L+L'},\hat{Q}_{L+L'}\cdot\hat\sigma_{L+L'})b(\hat{P}_1,\hat{Q}_1,\hat\sigma_1,\ldots,\hat{P}_{L+L'},\hat{Q}_{L+L'},\hat\sigma_{L+L'})$$

$$-\sigma A^2 \sum_{\hat\sigma=\pm1}{}''' b(\hat{P}_1,\hat{Q}_1,\hat\sigma_1,\ldots,\hat{P}_{L+L'},\hat{Q}_{L+L'},\hat\sigma_{L+L'}), \qquad \text{(A7)}$$

where

$$h_1 = \Big(\sum_{i=1}^{L+L'} \hat{P}\hat\sigma_i\Big)^2 + \Big(\sum_{i=1}^{L+L'} \hat{Q}_i\hat\sigma_i\Big)^2 - \frac{1}{2}\sum_{i=1}^{L+L'} \hat{P}_i^2\hat\sigma_i - \frac{1}{2}\sum_{i=1}^{L+L'} \hat{Q}_i^2\hat\sigma_i,$$

$$b = \prod_{i<j}^{(L+L')} \left[\frac{1}{\sigma A^2}(P_i - P_j)(Q_i - Q_j)\right]^{(1+\hat\sigma_i\hat\sigma_j)/2},$$

$$h_2 = \sum_{i,j=1}^{L+L'} \hat{P}_i\hat{Q}_j\hat\sigma_i\hat\sigma_j,$$

with

$$\hat{P}_i = P_i, \quad \hat{Q}_i = Q_i, \quad \hat\sigma_i = \sigma_i,$$

for $i = 1,2,\ldots,L$,

$$\hat{P}_{i+L} = -P_i^*, \quad \hat{Q}_{i+L} = -Q_i^*, \quad \hat\sigma_{i+L} = \sigma_{i+N},$$

for $i = 1,2,\ldots,L'$.

Here,

$$\sum_{\hat\sigma=\pm1}{}', \quad \sum_{\hat\sigma=\pm1}{}'', \quad \text{and} \quad \sum_{\hat\sigma=\pm1}{}'''$$

imply the summation over all possible combinations of $\hat\sigma = \pm 1$ under the condition

$$\sum_{i=1}^{L+L'} \hat\sigma_i = 1,0,-2,$$

respectively; and

$$\prod_{i<j}^{(L+L')}$$

indicates the products of all the combinations of pairs chosen among $L + L'$ elements.

From (A6) and (A7) one deduces that $\hat{D}_1$ and $\hat{D}_2$ are invariant with respect to the exchanges $\hat{P}_i \to \hat{Q}_i$ and $(\hat{P}_i,\hat{Q}_i) \leftrightarrow (\hat{P}_j,\hat{Q}_j)$, for any $i<j$. Furthermore, when $\hat{P}_1 = \hat{P}_2$ and $\hat{Q}_1 = \hat{Q}_2$ we have

$$\hat{D}_1(\hat{P}_1,\hat{Q}_1,\hat{P}_2 = \hat{P}_1,\hat{Q}_2 = \hat{Q}_1,\ldots) = 2\prod_{j=3}^{n}\left[\left(\frac{1}{\sigma A^2}\right)(\hat{P}_1 - \hat{P}_j)(\hat{Q}_1 - \hat{Q}_j)\right] \times \hat{D}_1(\hat{P}_3,\hat{Q}_3,\ldots,\hat{P}_n,\hat{Q}_n)$$

and

$$\hat{D}_2(\hat{P}_1,\hat{Q}_1,\hat{P}_2 = \hat{P}_1,\hat{Q}_2 = \hat{Q}_1,\ldots) = 2\prod_{j=3}^{n}\left[\left(\frac{1}{\sigma A^2}\right)(\hat{P}_1 - \hat{P}_j)(\hat{Q}_1 - \hat{Q}_j)\right] \times \hat{D}_2(\hat{P}_3,\hat{Q}_3,\ldots,\hat{P}_n,\hat{Q}_n).$$

Using the above properties of $\hat{D}_1$ and $\hat{D}_2$ and keeping in mind that $\hat{D}_1 = 0$ for $n = 1$ and $\hat{D}_2 = 0$ for $n = 2$, one obtains that $\hat{D}_1$ and $\hat{D}_2$ can be factorized by the polynomial

$$\prod_{k<l}^{(n)} (\hat{P}_k - \hat{P}_l)(\hat{Q}_k - \hat{Q}_l)$$

of degree $\frac{1}{2}n(n-1)$ in the variables $\hat{P}$ (or in the variables $\hat{Q}$). On the other hand, from (A6) and (A7) we see that $\hat{D}_1$ and $\hat{D}_2$ are polynomials in the variables $\hat{P}$ of degrees $\frac{1}{4}(n-1)^2 + 2$ and $\frac{1}{4}n(n-2) + 1$, respectively. Such conditions on $\hat{D}_1$ and $\hat{D}_2$ are compatible only for $n = 3$ and $n = 2$, respectively. Thus we infer that for $n > 3$, $\hat{D}_1$ and $\hat{D}_2$ are identically zero. Moreover, we have directly verified that $\hat{D}_1 = 0$ also for $n = 3$. Finally, remembering that $\hat{D}_2 = 0$ for $n = 2$, we have checked that (A1) and (A2) are identically satisfied.

[1] J. D. Gibbon, P. Radmore, M. Tabor, and D. Wood, Stud. Appl. Math. **72**, 39 (1985).

[2] R. A. Leo, G. Mancarella, G. Soliani, and L. Solombrino, Nuovo Cimento B **100**, 139 (1987).

[3] W. Strampp and F. Renner, in *Topics in Soliton Theory and Exactly Solvable Nonlinear Equations*, edited by M. Ablowitz, B. Fuchssteiner, and M. Kruskal (World Scientific, Singapore, 1987), p. 284.

[4] J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).

[5] R. Hirota, in *Bäcklund Transformations, The Inverse Scattering Methods, Solitons and their Applications*, edited by R. M. Miura, *Lecture Notes in Mathematics*, Vol. 515 (Springer, Berlin, 1974), p. 40.

[6] R. Hirota, Prog. Theor. Phys. **52**, 1498 (1974).

[7] R. Hirota and J. Satsuma, Prog. Theor. Phys. Suppl. **59**, 64 (1976).

[8] R. Hirota, in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980), p. 157.

[9] F. B. Estabrook and H. D. Wahlquist, J. Math. Phys. **17**, 1293 (1976).

[10] A. Davey and K. Stewartson, Proc. R. Soc. London Ser. A **338**, 101 (1974).

[11] A. S. Fokas and M. J. Ablowitz, J. Math. Phys. **25**, 2494 (1984).

[12] B. Champagne and P. Winternitz, "On the infinite dimensional symmetry group of the Davey–Stewartson equations," Preprint CRM-1414, October 1986.

[13] Without loss of generality, we shall take $u_0 = u_0^*$. In fact, this choice does not affect the nature of the complex variable u_1 [see (2.12)].

[14] J. Weiss, J. Math. Phys. **24**, 1405 (1983).

[15] J. Satsuma and M. J. Ablowitz, J. Math. Phys. **20**, 1497 (1979).

[16] R. Hirota, J. Math. Phys. **14**, 805 (1973).

[17] M. Tabor and J. D. Gibbon, Physica D **18**, 180 (1986).

Periodic reduction of self-dual Yang–Mills equations

Yoshimasa Nakamura
Department of Mathematics, Faculty of Education, Gifu University, Gifu 501-11, Japan

(Received 27 April 1988; accepted for publication 27 July 1988)

A class of GL(∞)-invariant self-dual Yang–Mills (SDYM) equations is considered. It is shown that the GL(∞) SDYM equations are reduced to the GL(n,$\mathbb{C}$) SDYM hierarchy by imposing a periodic condition. This reduction procedure makes clear a relationship between our GL(∞) SDYM equations and the usual infinite matrix representation of a single GL(n,$\mathbb{C}$) SDYM equation.

I. INTRODUCTION

The success of Sato and Sato[1] in the theory of soliton equations gave us a better understanding of a class of nonliner integrable systems (see also Refs. 2). It was pointed out that every soliton equation [the Korteweg–de Vries (KdV) equation, the Boussinesq equation, and so on] is embedded in the Kadomtsev–Petviashvili (KP) equation hierarchy whose solution space is identified with an infinite-dimensional Grassman manifold (see, for example, Ref. 3). Here the KP hierarchy is a GL(∞)-invariant infinite system of compatible completely integrable nonlinear evolution equations depending on an infinite number of time variables $t=(t_m)_{m\in\mathbb{N}}$, where $\mathbb{N}=\{1,2,...\}$. However, there are some nonlinear integrable systems that are outside of the KP hierarchy but have properties very similar to those of soliton equations. They are the stationary axially symmetric vacuum Einstein equation, the Bogomolny equation, and the self-dual Yang–Mills (SDYM) equation. The first two are derived from the last by specializations.[4]

Recently, the author[5] obtained a GL(n,$\mathbb{C}$) SDYM hierarchy. Here the GL(n,$\mathbb{C}$) SDYM hierarchy is an infinite ($k\in\mathbb{N}$) system of compatible GL(kn,$\mathbb{C}$) SDYM equations on which a subgroup of GL(∞), called the loop group, acts. But it has not been clear how to define GL(∞)-invariant SDYM equations in the spirit of Sato and Sato.

The first purpose of this paper is to present a class of GL(∞)-invariant SDYM equations as a higher-dimensional analog of the KP hierarchy. Here each GL(k,$\mathbb{C}$) SDYM equation ($k\in\mathbb{N}$) is embedded in the SDYM equations. The second purpose is to show that by imposing an algebraic constraint called the n-periodic condition the GL(n,$\mathbb{C}$) SDYM hierarchy found in Ref. 5 is derived from our GL(∞) SDYM equations. This fact will make clear a relationship between the GL(∞) SDYM equations and the infinite matrix representation of a single GL(n,$\mathbb{C}$) SDYM equation given by Takasaki.[6]

II. GL(∞)-INVARIANT SELF-DUAL YANG–MILLS EQUATIONS

Let $A=(a_{ij})_{i,j\in\mathbb{Z}}$ denote an infinite matrix whose elements are arrayed as

$$A=\begin{bmatrix} & & \cdots & & \\ \cdots & a_{-1-1} & a_{-10} & a_{-11} & \cdots \\ \cdots & a_{0-1} & a_{00} & a_{01} & \cdots \\ \cdots & a_{1-1} & a_{10} & a_{11} & \cdots \\ & & \cdots & & \end{bmatrix}. \quad (1)$$

Let $\Xi=(\xi_{ij})_{i,j\in\mathbb{Z}}$ be a projection matrix ($\Xi^2=\Xi$) with $\xi_{ii}=1$ for $i\geqslant 0$, $\xi_{ij}=\xi_{ij}(y_k,\bar{y}_k,z_k,\bar{z}_k)$ for $i\leqslant -1$ and $j\geqslant 0$, $\xi_{ij}=0$ for others, where the y_k's are complex variables ($k\in\mathbb{N}$). It is noted that columns for $j\geqslant 0$ of Ξ span infinite-dimensional subspaces $\mathbb{C}^{\mathbb{N}}$ in $\mathbb{C}^{\mathbb{Z}}$, and Ξ is identified with an affine coordinate of an infinite-dimensional Grassman manifold GM(∞).[1,2,5,6] Let Λ be a shift matrix defined by $\Lambda=(\delta_{ij+1})_{i,j\in\mathbb{Z}}$. The product of infinite matrices is defined by

$$(a_{ij})_{i,j\in\mathbb{Z}}(b_{ij})_{i,j\in\mathbb{Z}}=\Big(\sum_{l\in\mathbb{Z}}a_{il}b_{lj}\Big)_{i,j\in\mathbb{Z}}.$$

Let $\mathbb{H}$ be a subgroup of GL(∞) of matrices $H=(h_{ij})_{i,j\in\mathbb{Z}}$, where $h_{ii}=1$ for $i\in\mathbb{Z}$, $h_{ij}=h_{ij}(y_k,\bar{y}_k,z_k,\bar{z}_k)$ for $i\leqslant -1, j\geqslant 0$, and $h_{ij}=0$ for others.

Solutions to the usual SU(k) SDYM equation are given by solving the GL(k,$\mathbb{C}$) SDYM equation

$$\partial_{\bar{y}_k}(\partial_{y_k}Q_k\cdot Q_k^{-1})+\partial_{\bar{z}_k}(\partial_{z_k}Q_k\cdot Q_k^{-1})=0, \quad (2)$$

under the reality condition,[7] where $Q_k\in$GL(k,$\mathbb{C}$). Here $\partial_{y_k}=\partial/\partial_{y_k}$ and so on. We consider a system of linear equations

$$D_kU=U\Lambda^k\partial_{\bar{z}_k}\Xi,\quad D_k=\partial_{y_k}+\Lambda^k\partial_{\bar{z}_k},$$
$$D_k^*U=-U\Lambda^k\partial_{\bar{y}_k}\Xi,\quad D_k^*=\partial_{z_k}-\Lambda^k\partial_{\bar{y}_k}, \quad (3)$$

where $U\in\mathbb{H}$. The integrability conditions of U are found to be second-order equations on GM(∞),

$$\partial_{y_k}(\Lambda^k\partial_{\bar{y}_k}\Xi)+\partial_{z_k}(\Lambda^k\partial_{\bar{z}_k}\Xi)+[\Lambda^k\partial_{\bar{z}_k}\Xi,\Lambda^k\partial_{\bar{y}_k}\Xi]=0, \quad (4)$$

$$(\partial_{\bar{y}_k}\partial_{\bar{z}_k}-\partial_{\bar{z}_k}\partial_{\bar{y}_k})\Xi=0. \quad (5)$$

Conversely, if Ξ satisfies (4) and (5) then the linear equations in (3) are compatible to each other. Hence given any solution to (4) satisfying the integrability condition (5) there is a solution U to (3). If there is a GL(∞)-valued

 0022-2488/88/122672-03$02.50

matrix function Q such that $\Lambda^k \partial_{\bar{z}_k}\Xi = -\partial_{y_k}Q\cdot Q^{-1}$ and $\Lambda^k \partial_{\bar{y}_k}\Xi = \partial_{z_k}Q\cdot Q^{-1}$, then (4) is automatically satisfied. This implies from (5) that Q should satisfy a GL(∞) SDYM equation

$$\partial_{\bar{y}_k}(\partial_{y_k}Q\cdot Q^{-1}) + \partial_{\bar{z}_k}(\partial_{z_k}Q\cdot Q^{-1}) = 0. \qquad (6)$$

We also note that each GL(k,$\mathbb{C}$) SDYM equation (2) is embedded in (4). To see this let us define $k\times k$ matrices

$$\Xi_k = \begin{bmatrix} \xi_{-k0} & \cdots & \xi_{-kk-1} \\ \vdots & & \vdots \\ \xi_{-10} & \cdots & \xi_{-1k-1} \end{bmatrix}. \qquad (7)$$

It is not hard to see that the Ξ_k satisfy

$$\partial_{y_k}\partial_{\bar{y}_k}\Xi_k + \partial_{z_k}\partial_{\bar{z}_k}\Xi_k + [\partial_{\bar{z}_k}\Xi_k,\partial_{\bar{y}_k}\Xi_k] = 0, \qquad (8)$$

from (4). Since Eq. (8) is a zero curvature condition on the (y_k,z_k) plane, there are GL(k,$\mathbb{C}$)-valued functions Q_k such that $\partial_{\bar{z}_k}\Xi_k = -\partial_{y_k}Q_k\cdot Q_k^{-1}$ and $\partial_{\bar{y}_k}\Xi_k = \partial_{z_k}Q_k\cdot Q_k^{-1}$, and consequently, we have the GL(k,$\mathbb{C}$) SDYM equation (2) from (5) and (8). The remaining part of (4) determines a sequence of infinite potentials for (2). These infinite potentials are different from the usual ones discussed in Refs. 5, 6, and 8 in the sense in which the infinite matrix Ξ for the usual potentials should satisfy an additional algebraic constraint as shown in the followings. It is shown here that all solutions to (3) lead to solutions to the GL(∞) SDYM equation (6) where each GL(k,$\mathbb{C}$) SDYM equations is embedded. We say that the linear system (3) defines the GL(∞) SDYM equations (6).

We discuss a gauge transformation of the GL(∞) SDYM equation. Let G be an $\mathbb{H}$-valued matrix function. If $U(\in\mathbb{H})$ is a solution to (3), then $\tilde{U} = UG$ $(\in\mathbb{H})$ is also a solution to (3) corresponding to a solution to (4) and (5) denoted as $\tilde{\Xi}$. Solutions Ξ and $\tilde{\Xi}$ are related by the formula

$$\Lambda^k\partial_{\bar{z}_k}\tilde{\Xi} = G^{-1}\Lambda^k\partial_{\bar{z}_k}\Xi\cdot G + G^{-1}(\partial_{y_k} + U^{-1}\Lambda^k U\partial_{\bar{z}_k})G,$$
$$\Lambda^k\partial_{\bar{y}_k}\tilde{\Xi} = G^{-1}\Lambda^k\partial_{\bar{y}_k}\Xi\cdot G - G^{-1}(\partial_{z_k} - U^{-1}\Lambda^k U\partial_{\bar{y}_k})G. \qquad (9)$$

III. PERIODIC REDUCTION

Let us consider a subhierarchy of the GL(∞) SDYM equations ($k\in\mathbb{N}$) by imposing an additional algebraic constraint, which will be called the n-periodic condition. Let Ξ depend also on infinite parameters $t = (t_m)_{m\in\mathbb{N}}$ through

$$\partial_{t_m}\Xi = [\Lambda^m\Xi,\Xi], \qquad (10)$$

for $m\in\mathbb{N}$. These are (locally in t) solved to

$$\Xi = \exp\Big(\sum_{m\in\mathbb{N}} t_m\Lambda^m\Big)\Xi^0\Big\{\mathrm{I} - \Xi^0 + \exp\Big(\sum_{m\in\mathbb{N}} t_m\Lambda^m\Big)\Xi^0\Big\}^{-1}, \qquad (11)$$

where $\mathrm{I} = (\delta_{ij})_{i,j\in\mathbb{Z}}$, $\Xi = \Xi(y_k,\bar{y}_k,z_k,\bar{z}_k;t)$, $\Xi = \Xi(y_k,\bar{y}_k,z_k,\bar{z}_k;0)$. A set $\{\exp(\Sigma_{m\in\mathbb{N}} t_m\Lambda^m);\ t\in\mathbb{R}^{\mathbb{N}}\}$ forms an Abelian subgroup of GL(∞). Remark that the exponential operator $\exp(\Sigma_{m\in\mathbb{N}} t_m\Lambda^m)$ describes the time evolution of the KP hierarchy.[1,2] It will be shown that the parameters $t = (t_m)_{m\in\mathbb{N}}$ have some hidden meaning in our theory of SDYM equations. We assume that Ξ satisfies, for a positive integer n,

$$[\Lambda^n\Xi,\Xi] = 0, \qquad (12)$$

or, equivalently, $\partial_{t_n}\Xi = 0$. After a calculation we have $[\Lambda^m\Xi,\Xi] = 0$ and $\partial_{t_m}\Xi = 0$ for $m = 0 \pmod n$. We refer to (12) as the n-periodic condition for the GL(∞) SDYM equations. Following Ref. 5, let us regard the $j\geqslant 0$ part of $W = (w_{ij})_{i,j\in\mathbb{Z}}$ as a homogeneous coordinate on GM(∞) corresponding to Ξ, i.e.,

$$(\xi_{ij})_{i<-1,\, j\geqslant 0} = (w_{ij})_{i<-1,\, j\geqslant 0}(w_{ij})^{-1}_{i,j\geqslant 0}. \qquad (13)$$

For $k = 0 \pmod n$ we can reduce (3) to

$$D_k W = 0, \quad D_k^* W = 0, \qquad (14)$$

respectively. Equation (12) implies that W takes the form of a (block) Toeplitz matrix: $W = (W_{j-i})_{i,j\in\mathbb{Z}}$, where the W_{j-i} are $n\times n$ matrices. We note that the system $D_n W = 0$, $D_n^* W = 0$, and $[\Lambda^n, W] = 0$ is an infinite matrix representation of a linear system for the single GL(n,$\mathbb{C}$) SDYM equation found by Takasaki.[6] Indeed, Takasaki discussed an infinite matrix representation (in our notation)

$$\partial_{y_n}(\Lambda^n\partial_{\bar{y}_n}\Xi) + \partial_{z_n}(\Lambda^n\partial_{\bar{z}_n}\Xi) + [\Lambda^n\partial_{\bar{z}_n}\Xi,\Lambda^n\partial_{\bar{y}_n}\Xi] = 0, \qquad (15)$$

$$[\Lambda^n\Xi,\Xi] = 0,$$

of the GL(n,$\mathbb{C}$) SDYM equation. Thus the usual GL(n,$\mathbb{C}$) SDYM equation (15) is embedded in the GL(∞) SDYM equations satisfying (12) through the GL(k,$\mathbb{C}$) SDYM equations (8) with $k = 0 \pmod n$. We remark that the infinite ($k\in\mathbb{N}$) system (4) and (5) satisfying the n-periodic condition (12) is called the GL(n,$\mathbb{C}$) SDYM hierarchy.[5] Let us remember the well-known fact that the KdV and the Boussinesq hierarchies are derived from the KP hierarchy by two- and three-periodic reductions, respectively, as subhierarchies.[1]

From (12) and (14) we have the ($n\times n$ matrix) Laplace equations $(\partial_{y_k}\partial_{\bar{y}_k} + \partial_{z_k}\partial_{\bar{z}_k})\ W_{j-i} = 0$, for $k = 0 \pmod n$, $i,j\in\mathbb{Z}$. The formula (13) implies that solutions to the GL(n,$\mathbb{C}$) SDYM hierarchy are given by a nonlinear superposition of solutions to the Laplace equations. If $W_{j-i} = 0$ for $|j-i| > l\geqslant 0$, then the ξ_{ij} are well defined and take the form of a ratio of Toeplitz determinants. Setting $k = 2$ and $n = 1$, we obtain the celebrated Atiyah–Ward ansatz solutions[9] $\mathscr{A}_l$ to the SU(2) SDYM equation.

Finally let us consider power series solutions. Two types of Cauchy problems for (14) are (locally) solved by using Lie transforms of initial values. The results are

$$W = \exp\Big\{\sum_{l\in\mathbb{N}}(z_{ln}\Lambda^{ln}\partial_{\bar{y}_{ln}} - y_{ln}\Lambda^{ln}\partial_{\bar{z}_{ln}})\Big\}W(0,\bar{y},0,\bar{z}),$$
$$W = \exp\Big\{\sum_{l\in\mathbb{N}}(\bar{y}_{ln}\Lambda^{-ln}\partial_{z_{ln}} - \bar{z}_{ln}\Lambda^{-ln}\partial_{y_{ln}})\Big\}W(y,0,z,0), \qquad (16)$$

where $\bar{y} = (\bar{y}_{ln})_{l\in\mathbb{N}}$ and so on. The corresponding formal power series solutions to the GL(n,$\mathbb{C}$) SDYM hierarchy are given through (13). We see that there are three (mutually commuting) independent "time evolution operators," (11) and (16). The infinite flows on GM(∞) defined by (3) generally do not commute to each other without using the periodic condition (12). These facts look rather promising and it is reasonable to expect that we may prove a complete integrability of the GL(n,$\mathbb{C}$) SDYM hierarchy (4), (5), and (12). Indeed, it is shown[5] that an infinite-dimensional subgroup of GL(∞) called the loop group acts on the solution space to the ($n \geqslant 2$) GL(n,$\mathbb{C}$) SDYM hierarchy as a symmetry group.

ACKNOWLEDGMENTS

The author would like to thank Professor M. Sato, Dr. K. Ueno, and Dr. K. Takasaki for valuable suggestions and discussions.

[1]M. Sato, RIMS Kokyuroku, Kyoto Univ. **439**, 30 (1981); M. Sato and Y. Sato, in *Nonlinear Partial Differential Equations in Applied Science; Proceedings of the U.S.–Japan Seminar, Tokyo 1982*, edited by H. Fujita, P. D. Lax, and G. Strang (North-Holland, Amsterdam, 1983); M. Sato, *Soliton Equations and the Universal Grassmann Manifold*, with notes by M. Noumi (Sophia University, Tokyo, 1984) (in Japanese).

[2]K. Ueno, in *Infinite Dimensional Groups with Applications*, edited by V. Kac (Springer, New York, 1985); M. Mulase, Adv. Math. **54**, 57 (1984); K. Ueno and K. Takasaki, in *Group Representation and Systems of Differential Equations*, edited by K. Okamoto (Kinokuniya, Tokyo, 1984).

[3]J. Satsuma and R. Hirota, J. Phys. Soc. Jpn. **51**, 3390 (1982).

[4]N. S. Manton, Nucl. Phys. B **135**, 319 (1978); L. Witten, Phys. Rev. D **19**, 718 (1979).

[5]Y. Nakamura, J. Math. Phys. **29**, 144 (1988).

[6]K. Takasaki, Commun. Math. Phys. **94**, 35 (1984), Saitama Math. J. **3**, 11 (1985).

[7]Y. Brihaye, D. B. Fairlie, J. Nuyts, and R. G. Yates, J. Math. Phys. **19**, 2528 (1978).

[8]K. Ueno and Y. Nakamura, Phys. Lett. B **109**, 273 (1982); Publ. Res. Inst. Math. Sci., Kyoto Univ. **19**, 519 (1983).

[9]E. F. Corrigan, D. B. Fairlie, R. G. Yates, and P. Goddard, Commun. Math. Phys. **58**, 223 (1978).

Electromagnetic scattering for a class of anisotropic layered media

Thomas M. Roberts, Harold A. Sabbagh, and L. David Sabbagh
Sabbagh Associates Inc., 4639 Morningside Drive, Bloomington, Indiana 47401

(Received 2 May 1988; accepted for publication 10 August 1988)

The problem of computing fields produced by currents in the presence of unflawed layered media is reduced to quadrature. Each layer either has single-axis-type anisotropy or is isotropic. Integral equations are derived for fields in the presence of layered media with flaws.

I. INTRODUCTION

Some interesting and useful media have nontrivial structure: the Earth is layered, some superconductors are anisotropic, and many reinforced composites are anisotropic in each layer. There are many recent innovations in scattering for anisotropic materials.[1-12] Although the medium often is the motivation, there are scientific reasons for studying anisotropic scattering. Anisotropy is crucial in at least one surprising electromagnetic effect.[13] Also, solving the direct scattering problem is a first step in setting up an inverse scattering problem: in the direct problem, scattering from "flaws" (which, in the Earth, could be great golden lodes) usually is given at least cursory attention.

We compute exactly the electromagnetic fields produced by currents in the presence of layered media whose layers have single-axis-type anisotropy. The writing mainly is in terms of anisotropic conductors and steady state alternating currents. However, it is shown that some or all layers could be dielectrics or dielectric conductors; the results apply also to pulsed currents.

The problem is solved in steps, with one section of the main text devoted to each sentence remaining in this paragraph. The problem is defined with Maxwell equations for steady state alternating currents in the presence of conductors. (The problem applies to other media and to pulsed currents.) Maxwell equations are worked into a form for which a preexisting Green's matrix technique[12] applies. The Green's matrix represents the field produced in one layer by a point current in another, or the same, layer; the field is a linear combination of four eigenmodes. The problem of computing coefficients—four per layer—in the eigenmode expansion is equivalent to a boundary value problem. The boundary value problem is solved either by inverting a banded matrix or by a more elegant technique developed here. Finally, there is a cursory treatment of electromagnetic scattering for media with flaws.

The heart of this paper is an eigenmode approach to Green's matrices. Compared with other eigenmode techniques, ours is efficiently computable and unusually explicit; it has borne unexpected fruit.[13]

II. MAXWELL EQUATIONS

Partition $\mathbf{R}^3$ into layers M_j $(j=0,1,\ldots,N-1)$ with parallel planar boundaries. Let $\hat{z}$ be perpendicular to all boundaries and z_j be the position of the lower boundary of M_j $(j=0,1,\ldots,N-2;\, z_{j+1}<z_j)$. (See Fig. 1.) Assume, for each M_j, that there are rectangular coordinates $\hat{x}_j$, $\hat{y}_j$, $\hat{z}$ in which the matrix conductivity is uniaxial: $\mathbf{J}_j=\mathrm{diag}\{\sigma_{xj},\sigma_j,\sigma_j\}\cdot\mathbf{E}_j$. Thus conductivity (σ_j) along the $\hat{y}_j$ axis of M_j is the same as along $\hat{z}$, but conductivity along $\hat{x}_j$ is distinct.

Take $e^{i\omega t}$ time dependence in Maxwell equations for steady state alternating currents and define $\epsilon_{xj}\equiv\epsilon_0-i\sigma_{xj}/\omega$ and $\epsilon_j\equiv\epsilon_0-i\sigma_j/\omega$. Then rewrite

$$\begin{pmatrix}\hat{x}_j\cdot\nabla\times\mathbf{H}\\ \hat{y}_j\cdot\nabla\times\mathbf{H}\\ \hat{z}\cdot\nabla\times\mathbf{H}\end{pmatrix}=i\omega\begin{pmatrix}\epsilon_{xj}\hat{x}_j\cdot\mathbf{E}\\ \epsilon_j\hat{y}_j\cdot\mathbf{E}\\ \epsilon_j\hat{z}\cdot\mathbf{E}\end{pmatrix}+\begin{pmatrix}\hat{x}_j\cdot\mathbf{J}\\ \hat{y}_j\cdot\mathbf{J}\\ \hat{z}\cdot\mathbf{J}\end{pmatrix}$$

with the matrix $\epsilon_j\equiv\mathrm{diag}\{\epsilon_{xj},\epsilon_j,\epsilon_j\}$ and a dot-product operator, $\rho_j\equiv\mathrm{diag}\{\hat{x}_j\cdot,\hat{y}_j\cdot,\hat{z}\cdot\}$, as $\rho_j\circ\nabla\times\mathbf{H}=i\omega\epsilon_j\cdot\rho_j\circ\mathbf{E}+\rho_j\circ\mathbf{J}$. Define one fixed coordinate system ρ and let θ_j be the angle by which ρ_j is rotated counterclockwise away from ρ: $\rho\equiv\mathrm{diag}\{\hat{x}\cdot,\hat{y}\cdot,\hat{z}\cdot\}=\mathbf{T}_j^{-1}\cdot\rho_j$, where

$$\mathbf{T}_j\equiv\begin{pmatrix}\cos\theta_j & \sin\theta_j & 0\\ -\sin\theta_j & \cos\theta_j & 0\\ 0 & 0 & 1\end{pmatrix}.$$

(Refer to ρ as global coordinates and ρ_j as local coordinates.) Substitute $\mathbf{T}_j\cdot\rho$ for ρ_j and premultiply by $\mathbf{T}_j^{-1}$ to obtain $\rho\circ\nabla\times\mathbf{H}=i\omega\mathbf{T}_j^{-1}\cdot\epsilon_j\cdot\mathbf{T}_j\cdot\rho\circ\mathbf{E}+\rho\circ\mathbf{J}$: that is,

$$\begin{pmatrix}\hat{x}\cdot\nabla\times\mathbf{H}\\ \hat{y}\cdot\nabla\times\mathbf{H}\\ \hat{z}\cdot\nabla\times\mathbf{H}\end{pmatrix}=i\omega\zeta_j\cdot\begin{pmatrix}E_x\\ E_y\\ E_z\end{pmatrix}+\begin{pmatrix}J_x\\ J_y\\ J_z\end{pmatrix},\qquad(2.1)$$

where $\zeta_j\equiv\mathbf{T}_j^{-1}\cdot\epsilon_j\cdot\mathbf{T}_j$ has nonzero off-diagonal elements. There is a second set of Maxwell curl equations:

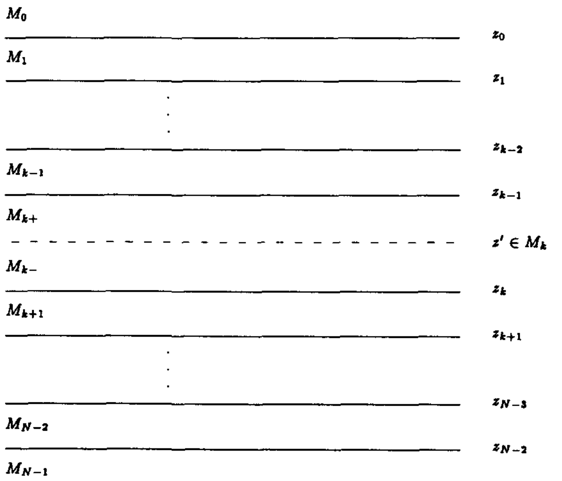


FIG. 1. Layers and boundaries.

$$\begin{pmatrix}\hat{x}\cdot\nabla\times\mathbf{E}\\ \hat{y}\cdot\nabla\times\mathbf{E}\\ \hat{z}\cdot\nabla\times\mathbf{E}\end{pmatrix} = -i\mu_0\omega\begin{pmatrix}H_x\\ H_y\\ H_z\end{pmatrix}. \tag{2.2}$$

Our goal is to solve curl equations for **E** and **H** in global coordinates; the solution uses local coordinates.

III. SCOPE OF THE PROBLEM

Section II defines a scattering problem for stacked layers. The layers can be of dissimilar materials: the only restriction is that each is uniaxial or isotropic; some layers could be free space. The problem allows each $\hat{x}_j$ axis to point in any direction perpendicular to $\hat{z}$. Layers can have unequal thicknesses.

The variables σ_{xj} and σ_j can be complex valued and ω dependent. If $\sigma_{xj} = i\omega(p_{xj} - \epsilon_0) + s_{xj}$ and $\sigma_j = i\omega(p_j - \epsilon_0) + s_j$ (with $p_{xj}, p_j, s_{xj}, s_j \in \mathbf{R}$), then $\epsilon_{xj} = p_{xj} - is_{xj}/\omega$ and $\epsilon_j = p_j - is_j/\omega$ represent layers with uniaxial permittivities $\text{diag}\{p_{xj}, p_j, p_j\}$ and uniaxial conductivities $\text{diag}\{s_{xj}, s_j, s_j\}$.

For each ω, a Green's matrix $\mathbf{g} = \mathbf{g}(k_x, k_y, \omega, z, z')$ will be constructed $\ni \int_{-\infty}^{\infty} dz'\, \mathbf{g}\cdot\tilde{\mathbf{J}}(z')$ is the xy-Fourier transform of the x and y components of **E** and **H**. [$\tilde{\mathbf{J}}(z) = \tilde{\mathbf{J}}(k_x, k_y, \omega, z)$ is the xy-Fourier transform of current.] The xy transforms of z components of **E** and **H** are linear combinations of the transforms of x and y components, as in (4.2). The scattering problem thereby will be solved in the frequency domain. The solution for pulsed currents

$$\int_{-\infty}^{\infty} d\omega\, e^{i\omega t}\mathbf{J}(x,y,z,\omega)$$

evolves in similar fashion from

$$\int_{-\infty}^{\infty} d\omega\, e^{i\omega t}\int_{-\infty}^{\infty} dz'\mathbf{g}\cdot\tilde{\mathbf{J}}.$$

Our scattering problem has a wide scope. For simplicity, we revert to the terminology of steady state currents in the presence of conductors.

IV. TRANSFORMED MAXWELL EQUATIONS

In previous work for single-layer conductors,[12] we used xy-Fourier transforms to equate Maxwell curl equations—they are six coupled partial differential equations—with a system of four ordinary differential equations (ODE's) and two linear equations. In this section, we apply the transform technique to multiple layers and work the resulting ODE's into a form that contains local coordinates only.

An xy-Fourier transform of global coordinates

$$\tilde{f}(k_x, k_y, z) \equiv \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx\, dy\, e^{i(k_x x + k_y y)} f(x,y,z) \tag{4.1}$$

takes ∂_x into multiplication by $-ik_x$, ∂_y into multiplication by $-ik_y$, and leaves ∂_z unchanged. The z components of curls in (2.1) and (2.2) have no z derivatives, so their transforms yield two algebraic equations:

$$\begin{aligned}\tilde{E}_z &= \frac{k_y}{\omega\epsilon_j}\tilde{H}_x - \frac{k_x}{\omega\epsilon_j}\tilde{H}_y - \frac{1}{i\omega\epsilon_j}\tilde{J}_z,\\ \tilde{H}_z &= \frac{-k_y}{\mu_0\omega}\tilde{E}_x + \frac{k_x}{\mu_0\omega}\tilde{E}_y.\end{aligned} \tag{4.2}$$

The four other transformed equations are ODE's. Use (4.2) to eliminate $\tilde{E}_z$ and $\tilde{H}_z$ from the ODE's and obtain

$$\partial_z\tilde{e} = \mathbf{S}_j\cdot\tilde{e} + \mathbf{U}_j\cdot\tilde{\mathbf{J}}, \tag{4.3}$$

with $\tilde{e} \equiv (\tilde{E}_x, \tilde{E}_y, \tilde{H}_x, \tilde{H}_y)^T$, $\tilde{\mathbf{J}} = (\tilde{J}_x, \tilde{J}_y, \tilde{J}_z)^T$, and ^{T} = transpose. The matrices $\mathbf{S}_j$ and $\mathbf{U}_j$ can be computed with straightforward algebra

$$\mathbf{S}_j = \frac{i}{\omega}\begin{pmatrix} 0 & 0 & -k_xk_y/\epsilon_j & -\mu_0\omega^2 + k_x^2/\epsilon_j\\ 0 & 0 & \mu_0\omega^2 - k_y^2/\epsilon_j & k_xk_y/\epsilon_j\\ \omega^2\zeta_{12j} + k_xk_y/\mu_0 & \omega^2\zeta_{22j} - k_x^2/\mu_0 & 0 & 0\\ -\omega^2\zeta_{11j} + k_y^2/\mu_0 & -\omega^2\zeta_{12j} - k_xk_y/\mu_0 & 0 & 0\end{pmatrix},$$

$$\mathbf{U}_j = \begin{pmatrix} 0 & 0 & k_x/(\omega\epsilon_j)\\ 0 & 0 & k_y/(\omega\epsilon_j)\\ 0 & 1 & 0\\ -1 & 0 & 0\end{pmatrix}, \tag{4.4}$$

where ζ_{lkj} is the lk component of ζ_j.

Equation (4.3) mixes local and global coordinates; we seek an expression in local coordinates only. Toward that end, define $\tilde{e}_j \equiv \mathbf{R}_j\cdot\tilde{e}$ with

$$\mathbf{R}_j \equiv \begin{pmatrix}\cos\theta_j & \sin\theta_j & 0 & 0\\ -\sin\theta_j & \cos\theta_j & 0 & 0\\ 0 & 0 & \cos\theta_j & \sin\theta_j\\ 0 & 0 & -\sin\theta_j & \cos\theta_j\end{pmatrix}. \tag{4.5}$$

Substitute $\mathbf{R}_j^{-1}\cdot\tilde{e}_j$ for $\tilde{e}$ in (4.3) and premultiply by $\mathbf{R}_j$ to obtain

$$\partial_z\tilde{e}_j = \mathbf{R}_j\cdot\mathbf{S}_j\cdot\mathbf{R}_j^{-1}\cdot\tilde{e}_j + \mathbf{R}_j\cdot\mathbf{U}_j\cdot\tilde{\mathbf{J}}. \tag{4.6}$$

Let $\Sigma_j \equiv \mathbf{R}_j\cdot\mathbf{S}_j\cdot\mathbf{R}_j^{-1}$, $k_{xj} \equiv k_x\cos\theta_j + k_y\sin\theta_j$, and $k_{yj} \equiv -k_x\sin\theta_j + k_y\cos\theta_j$. A derivation in Appendix A yields

$$\Sigma_j = \frac{i}{\omega}\begin{pmatrix} 0 & 0 & -k_{xj}k_{yj}/\epsilon_j & -\mu_0\omega^2 + k_{xj}{}^2/\epsilon_j \\ 0 & 0 & \mu_0\omega^2 - k_{yj}{}^2/\epsilon_j & k_{xj}k_{yj}/\epsilon_j \\ k_{xj}k_{yj}/\mu_0 & \omega^2\epsilon_j - k_{xj}{}^2/\mu_0 & 0 & 0 \\ -\omega^2\epsilon_{xj} + k_{yj}{}^2/\mu_0 & -k_{xj}k_{yj}/\mu_0 & 0 & 0 \end{pmatrix}. \tag{4.7}$$

Define

$$\tilde{\mathbf{J}}_j(k_{xj},k_{yj},z') \equiv \mathbf{T}_j\cdot\tilde{\mathbf{J}}[k_x(k_{xj};k_{yj});k_y(k_{xj},k_{yj});z'],$$

for each $j = 0,1,\ldots,N-1$ and for each $z'\in\mathbf{R}$. (The role of j is clarified in the following paragraph.) Rewrite (4.6) as $\partial_z\tilde{e}_j = \Sigma_j\cdot\tilde{e}_j + \mathbf{R}_j\cdot\mathbf{U}_j\cdot\mathbf{T}_j^{-1}\cdot\tilde{\mathbf{J}}_j$: a statement in which every term is in local coordinates. Straightforward multiplication yields

$$\partial_z\tilde{e}_j = \Sigma_j\cdot\tilde{e}_j + \Upsilon_j\cdot\tilde{\mathbf{J}}_j, \tag{4.8}$$

$$\Upsilon_j \equiv \mathbf{R}_j\cdot\mathbf{U}_j\cdot\mathbf{T}_j^{-1} = \begin{pmatrix} 0 & 0 & k_{xj}/(\omega\epsilon_j) \\ 0 & 0 & k_{yj}/(\omega\epsilon_j) \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

Maxwell curl equations are solved in seven steps.

(1) Given $\mathbf{J}(x,y,z')$, compute $\tilde{\mathbf{J}}(k_x,k_y,z')$.

(2) Identify a layer M_j for which $\mathbf{E}$ and $\mathbf{H}$ are to be computed.

(3) Compute

$$\tilde{\mathbf{J}}_j(k_{xj},k_{yj},z') = \mathbf{T}_j\cdot\tilde{\mathbf{J}}[k_x(k_{xj},k_{yj});k_y(k_{xj},k_{yj});z'],$$

for all $z'\in(-\infty,\infty)$, where j is defined in the second step.

(4) Solve $\partial_z\tilde{e}_j = \Sigma_j\cdot\tilde{e}_j + \Upsilon_j\cdot\tilde{\mathbf{J}}_j$ for $\tilde{e}_j(k_{xj},k_{yj},z)$, with $z\in M_j$.

(5) Compute

$$(\tilde{E}_x,\tilde{E}_y,\tilde{H}_x,\tilde{H}_y)^T = \tilde{e}(k_x,k_y,z) = \mathbf{R}_j^{-1}\cdot\tilde{e}_j[k_{xj}(k_x,k_y);k_{yj}(k_x,k_y);z].$$

(6) Use (4.2) to compute $\tilde{E}_z(k_x,k_y,z)$ and $\tilde{H}_z(k_x,k_y,z)$.

(7) Invert xy transforms to obtain $\mathbf{E}(x,y,z)$ and $\mathbf{H}(x,y,z)$.

The fourth step is the only nontrivial one; it is the subject of the next three sections.

V. GREEN'S MATRIX DEFINITION

Theorem 5.1 solves (4.8) by quadrature:

$$\tilde{e}_j(z) = \int_{-\infty}^{\infty} dz'\,\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_{\ell(z')}(z')$$

yields

$$(\tilde{E}_x,\tilde{E}_y,\tilde{H}_x,\tilde{H}_y)^T = \tilde{e} = \mathbf{R}_j^{-1}\cdot\tilde{e}_j,$$

which is continuous across boundaries.

Theorem 5.1: Let ℓ be a layer index function—$z\in M_{\ell(z)}$ and $z'\in M_{\ell(z')}$ can be in different layers; $\mathbf{g}(z,z')$ be a function of k_x and k_y implicitly; $[\partial_z - \Sigma_{\ell(z)}\mathbf{g}(z,z') = 0$, $\forall z\in\mathbf{R}\setminus\{z_0,z_1,\ldots,z_{N-2},z'\}$; $\mathbf{R}_{\ell(z)}^{-1}\cdot\mathbf{g}(z,z')$ be z continuous except at $z=z'$; and $\mathbf{g}(z,z'\uparrow z) - \mathbf{g}(z,z'\downarrow z) = \Upsilon_{\ell(z)}$.

Then

$$\mathbf{R}_{\ell(z)}^{-1}\cdot\int_{-\infty}^{\infty} dz'\,\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_{\ell(z')}$$

is z continuous and

$$[\partial_z - \Sigma_{\ell(z)}]\cdot\int_{-\infty}^{\infty} dz'\,\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_{\ell(z')}(z')$$

$$= \Upsilon_{\ell(z)}\cdot\tilde{\mathbf{J}}_{\ell(z)}(z), \quad \forall z\in\mathbf{R}\setminus\{z_0,z_1,\ldots,z_{N-2}\}.$$

Remark:

$$\tilde{\mathbf{J}}_{\ell(z')}(z') = \tilde{\mathbf{J}}_{\ell(z')}[k_{x\ell(z')}(k_x,k_y);k_{y\ell(z')}(k_x,k_y);z']$$

in the previous integral.

Proof: Continuity of $R^{-1}\int gJ$ results from the presumed continuity of $R^{-1}g$.

In the remainder of the proof, assume $k = \ell(z')$, $z_{N-1} \equiv -\infty$, $z_{-1} \equiv +\infty$, and $z\in\text{Int}\,M_j = (z_j,z_{j-1})$.

Let

$$\Delta \equiv (\partial_z - \Sigma_j)\cdot\int_{-\infty}^{\infty} dz'\,\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k(z').$$

Then

$$\Delta = (\partial_z - \Sigma_j)\cdot\left(\int_{-\infty}^{z} + \int_{z}^{\infty}\right) dz'\,\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k(z')$$

$$= (\partial_z - \Sigma_j)\cdot\left(\int_{z_j}^{z} + \int_{z}^{z_{j-1}}\right) dz'\,\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_j(z') + \delta,$$

where δ is a sum of terms like

$$(\partial_z - \Sigma_j)\cdot\int_a^b dz'\,\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k(z'),$$

with each term involving integration over a single layer and the sum (δ) involving integration over every layer except M_j. Each term in δ has z-independent bounds of integration and a z-continuous integrand; therefore, δ is a sum of terms like

$$\int_a^b dz'(\partial_z - \Sigma_j)\cdot\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k(z') = 0:$$

$\delta = 0$, too. Evaluate Δ using $\delta = 0$ and elementary facts[14]; the result is

$$\Delta = \lim_{z'\uparrow z}[\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k] + \int_{z_j}^{z} dz'(\partial_z - \Sigma_j)\cdot\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k(z') - \lim_{z'\downarrow z}[\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k] + \int_{z}^{z_{j-1}} dz'(\partial_z - \Sigma_j)\cdot\mathbf{g}(z,z')\cdot\tilde{\mathbf{J}}_k(z').$$

Then $\lim_{z'\to z} k = j$, $(\partial_z - \Sigma_j)\cdot\mathbf{g} = 0$, and z' discontinuity of $\mathbf{g}(z,z')$ imply $\Delta = \Upsilon_j\cdot\tilde{\mathbf{J}}_j$. ■

Let $j = \ell(z)$, $k = \ell(z')$, $M_{k-} = (z_k,z')$, $M_{k+} = (z',z_{k-1})$, and let $\mathbf{g}$ be as in Theorem 5.1. Then $(\partial_z - \Sigma_j)\cdot\mathbf{g}(z,z') = 0$.

One supposition in the theorem equates z' discontinuity of $\mathbf{g}$ with the 3-column × 4-row matrix Υ_j. Thus $\mathbf{g}$ also is a

3×4 matrix. Let $\tilde{\gamma}_{\hat{x}}, \tilde{\gamma}_{\hat{y}}$, and $\tilde{\gamma}_{\hat{z}}$ be consecutive columns of **g**; then $(\partial_z - \Sigma_j)\cdot\tilde{\gamma}_{\hat{e}}(z,z') = 0$, for $\hat{e} = \hat{x},\hat{y},\hat{z}$.

If v is an eigenvector of Σ_j with eigenvalue λ then $(\partial_z - \Sigma_j)\cdot(ve^{\lambda z}) = 0$. Therefore, each column $\tilde{\gamma}_{\hat{e}}$ is a linear combination of four linearly independent eigenmodes. In Sec. IV, we deliberately worked Maxwell equations into a form with a matrix Σ_j whose eigenvectors and eigenvalues are simple

$$v_{1j} = \begin{pmatrix} +\Sigma_{14j} \\ +\Sigma_{24j} \\ 0 \\ \lambda_{1j} \end{pmatrix}, \quad v_{2j} = \begin{pmatrix} -\Sigma_{14j} \\ -\Sigma_{24j} \\ 0 \\ \lambda_{1j} \end{pmatrix},$$

$$v_{3j} = \begin{pmatrix} 0 \\ \lambda_{3j} \\ +\Sigma_{32j} \\ -\Sigma_{31j} \end{pmatrix}, \quad v_{4j} = \begin{pmatrix} 0 \\ \lambda_{3j} \\ -\Sigma_{32j} \\ +\Sigma_{31j} \end{pmatrix},$$

$$\pm\lambda_{1j} = \pm[-\mu_0\omega^2\epsilon_{xj} + (\epsilon_{xj}/\epsilon_j)k_{xj}{}^2 + k_{yj}{}^2]^{1/2},$$
$$\pm\lambda_{3j} = \pm(-\mu_0\omega^2\epsilon_j + k_{xj}{}^2 + k_{yj}{}^2)^{1/2}, \qquad (5.1)$$

where Σ_{lkj} is the lkth component of Σ_j. (Any $\surd$ is acceptable, but we define $\mathrm{Arg}\surd\in(-\pi/2, +\pi/2]$.) The eigenvector-eigenvalue correspondence in (5.1) is $v_{1j}\to+\lambda_{1j}$, $v_{2j}\to-\lambda_{1j}$, $v_{3j}\to+\lambda_{3j}$, and $v_{4j}\to-\lambda_{3j}$. Therefore,

$$\tilde{\gamma}_{\hat{e}} = \alpha_j v_{1j}e^{\lambda_{1j}z} + \beta_j v_{2j}e^{-\lambda_{1j}z} + \gamma_j v_{3j}e^{\lambda_{2j}z} + \delta_j v_{3j}e^{-\lambda_{3j}z}, \qquad (5.2)$$

where $\alpha_j, \beta_j, \gamma_j, \delta_j$ are functions of $z_0, z_1, \ldots, z_{N-2}, z', \hat{e}$—and of $j = \ell(z)$, but are otherwise z independent.

Assume M_0 and M_{N-1} are isotropic media[15] and that **g** represents waves that travel outward at $z = \pm\infty$. Then the $e^{+i\omega t}$ time dependence implicit in (2.1) and (2.2) and the choice of $\surd$ yield $\alpha_0 = \gamma_0 = \beta_{N-1} = \delta_{N-1} = 0$. That expression of an outward-traveling boundary condition completes the definition of **g**.

The coefficients in (5.2) are computed in the next section.

VI. BOUNDARY VALUE PROBLEM

Redefine the phase reference in (5.2) and use the boundary condition $b_{N-1} = d_{N-1} = 0$

$$\tilde{\gamma}_{\hat{e}} = \begin{cases} a_j v_{1j} e^{\lambda_{1j}(z-z_j)} + \cdots + d_j v_{4j} e^{-\lambda_{3j}(z-z_j)}, & z\in(M_0\cup\cdots\cup M_{N-2})\setminus(M_{k+}\cup M_{k-}), \\ a_+ v_{1k} e^{\lambda_{1k}(z-z')} + \cdots + d_+ v_{4k} e^{-\lambda_{3k}(z-z')}, & z\in M_{k+}, \\ a_- v_{1k} e^{\lambda_{1k}(z-z')} + \cdots + d_- v_{4k} e^{-\lambda_{3k}(z-z')}, & z\in M_{k-}, \\ a_{N-1} v_{1,N-1} e^{\lambda_{1,N-1}(z-z_{N-2})} + c_{N-1} v_{3,N-1} e^{\lambda_{3,N-1}(z-z_{N-2})}, & z\in M_{N-1}. \end{cases} \qquad (6.1)$$

(It is true that $a_0 = c_0 = 0$ also.) Each ellipsis in (6.1) represents two eigenmodes. In M_j, for example, the ellipsis represents

$$b_j v_{2j}\exp[-\lambda_{1j}(z-z_j)] + c_j v_{3j}\exp[\lambda_{3j}(z-z_j)].$$

The $\hat{e}$ dependence of coefficients also is suppressed. Thus $4N$ coefficients must be determined for each $\hat{e} = \hat{x},\hat{y},\hat{z}$.

The z continuity of $\mathbf{R}^{-1}_{\ell(z)}\cdot\mathbf{g}(z,z')$ implies z continuity of $\mathbf{R}^{-1}_{\ell(z)}\cdot\tilde{\gamma}_{\hat{e}}(z,z')$, though neither quantity is continuous at $z = z'$. Thus z continuity yields $N-1$ four-vector boundary conditions

$$\mathbf{R}_j^{-1}\cdot(a_j v_{1j} + \cdots + d_j v_{4j})$$
$$= \mathbf{R}^{-1}_{j+1}\cdot[a_{j+1}v_{1,j+1}e^{\lambda_{1,j+1}(z_j - z_{j+1})}$$
$$+ \cdots + d_{j+1}v_{4,j+1}e^{-\lambda_{3,j+1}(z_j - z_{j+1})}],$$
$$\text{for } j\in\{0,\ldots,N-3\}\setminus\{k-1,k\};$$
$$\mathbf{R}^{-1}_{k-1}\cdot(a_{k-1}v_{1,k-1} + \cdots + d_{k-1}v_{4,k-1})$$
$$= \mathbf{R}_k^{-1}\cdot[a_+ v_{1k}e^{\lambda_{1k}(z_{k-1}-z')}$$
$$+ \cdots + d_+ v_{4k}e^{-\lambda_{3k}(z_{k-1}-z')}];$$
$$\mathbf{R}_k^{-1}\cdot[a_- v_{1k}e^{\lambda_{1k}(z_k - z')} + \cdots + d_- v_{4k}e^{-\lambda_{3k}(z_k - z')}]$$
$$= \mathbf{R}^{-1}_{k+1}\cdot[a_{k+1}v_{1,k+1}e^{\lambda_{1,k+1}(z_k - z_{k+1})}$$
$$+ \cdots + d_{k+1}v_{4,k+1}e^{-\lambda_{3,k+1}(z_k - z_{k+1})}];$$

and

$$\mathbf{R}^{-1}_{N-2}\cdot(a_{N-2}v_{1,N-2} + \cdots + d_{N-2}v_{4,N-2})$$
$$= \mathbf{R}^{-1}_{N-1}\cdot(a_{N-1}v_{1,N-1} + c_{N-1}v_{3,N-1}). \qquad (6.2)$$

Equation (6.1) and the z' discontinuity in Theorem 5.1 imply

$$(a_+ - a_-)v_{1k} + \cdots + (d_+ - d_-)v_{4k} = \Upsilon_k\cdot\hat{e}, \qquad (6.3)$$

for $\hat{e} = \hat{x},\hat{y},\hat{z}$. For each $\hat{e}$, (6.2) and (6.3) are N four-vector boundary conditions. The boundary conditions determine the $4N$ coefficients in (6.1), subject to limitations of linear algebra.

VII. GREEN'S MATRIX COMPUTATION

Seven steps at the end of Sec. IV equate the problem of solving Maxwell curl equations with the problem of computing $\tilde{e}_j$. Section V opens with a computation of $\tilde{e}_j$ in terms of a Green's matrix **g**. It remains to compute **g**.

The Green's matrix is computed in three steps.

(1) Obtain $\mathbf{R}_j$ as in (4.5), k_{xj} and k_{yj} as follows (4.6), Σ_j as in (4.7), Υ_j as in (4.8), and eigenmodes as in (5.1).

(2) For each $\hat{e}\in\{\hat{x},\hat{y},\hat{z}\}$, solve a boundary value problem—consisting of (6.2), (6.3), and $a_0 = c_0 = 0$—for the $4N$ coefficients that are not yet determined, whose base letters are a,b,c, and d.

(3) Use (6.1) to compute the columns $\tilde{\gamma}_{\hat{e}}(z,z')$ of **g**.

The second step is the only nontrivial one; it is the subject of this section and Appendix B.

For each $\hat{e}$, carry the a-, b-, c-, and d-type coefficients of (6.2) and (6.3) to the left and leave $\Upsilon_k\cdot\hat{e}$ on the right. The result is a system of $4N$ linear equations with $4N$ unknowns. The system is inhomogeneous due to $\Upsilon_k\cdot\hat{e}$ alone.

Equations (6.2) and (6.3) relate coefficients for each

M_j to, at most, coefficients for M_{j-1} and M_{j+1}. If coefficients are ordered by sequential layers, then the ordered system of equations is described by a banded matrix. Thus **g** can be computed by inverting a $4N\times 4N$ banded matrix that has $8N-4$ nonzero entries. (Figure 1 shows that each boundary contributes eight nonzero entries, except z_0 and z_{N-2} which supply six each.) Some numerical techniques are efficient especially for inverting banded matrices.[16]

The Green's matrix can be computed with an alternative algorithm that inverts matrices no larger than 2×2. (Details are in Appendix B.) The first step is to offer a guess—$a_{N-1}^{(1)}$ and $c_{N-1}^{(1)}$—of the coefficients a_{N-1} and b_{N-1}. The guess can be far from the right answer. Use (6.2), (6.3), and the guess to compute coefficients $a_j^{(1)},b_j^{(1)},c_j^{(1)},d_j^{(1)}$ in successive layers M_{N-2}, $M_{N-3},\ldots,M_0$. The result represents a state $\tilde{\gamma}_{\hat{e}}^{(1)}$ with outward-traveling waves at $z=-\infty$, but with both outward- and inward-traveling waves at $z=+\infty$. In a similar way, use $a_{N-1}^{(2)}\equiv c_{N-1}^{(3)}\equiv 1$ and $a_{N-1}^{(3)}\equiv c_{N-1}^{(2)}\equiv 0$ to construct states $\tilde{\gamma}^{(2)}$ and $\tilde{\gamma}^{(3)}$ that satisfy (6.2) but have *continuity* in place of (6.3). Then determine the linear combination of the coefficients of $\tilde{\gamma}_{\hat{e}}^{(1)}$, $\tilde{\gamma}^{(2)}$, and $\tilde{\gamma}^{(3)}$ that equals the coefficients of the outward-traveling mode $\tilde{\gamma}_{\hat{e}}$.

This section completes a solution of the scattering problem for unflawed media.

VIII. FLAWS

Present a current **J** to a reference medium $\zeta_j^{(1)}\equiv \mathbf{T}_j^{-1}\cdot\epsilon_j\cdot\mathbf{T}_j$ and to a flawed medium $\zeta_j^{(2)}=\zeta_j^{(1)}+\zeta_j^{(\Delta)}$. The reference medium is homogeneous in each layer. The flaw $\zeta_j^{(\Delta)}$ has bounded support.

The current **J** radiates fields that satisfy Maxwell curl equations $\nabla\times\mathbf{H}^{(m)}=i\omega\zeta_j^{(m)}\cdot\mathbf{E}^{(m)}+\mathbf{J}$ and $\nabla\times\mathbf{E}^{(m)}=-i\mu_0\omega\mathbf{H}^{(m)}$, for $m=1,2$. Sections IV–VII apply to media that are homogeneous in each layer; they show $\mathbf{E}^{(1)}$ and $\mathbf{H}^{(1)}$ can be computed from

$$\tilde{e}^{(1)}=\mathbf{R}_{\ell(z)}^{-1}\cdot\int_{-\infty}^{\infty}dz'\,\mathbf{g}\cdot\mathbf{T}_{\ell(z')}\cdot\tilde{\mathbf{J}},$$

but they do not necessarily apply to the flawed medium $\zeta_j^{(2)}$.

Subtract the $m=1$ curl equations from the $m=2$ equations to obtain

$$\nabla\times[\mathbf{H}^{(2)}-\mathbf{H}^{(1)}]=i\omega\zeta_{\ell(z)}^{(1)}\cdot[\mathbf{E}^{(2)}-\mathbf{E}^{(1)}]+i\omega\zeta_{\ell(z)}^{(\Delta)}\cdot\mathbf{E}^{(2)}$$

and

$$\nabla\times[\mathbf{E}^{(2)}-\mathbf{E}^{(1)}]=-i\mu_0\omega[\mathbf{H}^{(2)}-\mathbf{H}^{(1)}].$$

The curl equations for differences have the same form as the $m=1$ curl equations, with $i\omega\zeta_{\ell(z)}^{(\Delta)}\cdot\mathbf{E}^{(2)}$ in place of **J**. It follows that

$$\tilde{e}^{(2)}-\tilde{e}^{(1)}=i\omega\mathbf{R}_{\ell(z)}^{-1}\cdot\int_{-\infty}^{\infty}dz'\,\mathbf{g}\cdot\mathbf{T}_{\ell(z')}\cdot\mathscr{F}[\zeta_{\ell(z')}^{(\Delta)}\cdot\mathbf{E}^{(2)}],$$

where $\mathscr{F}$ is the xy-Fourier transform of (4.1). Let

$$\mathbf{G}(z,z')\equiv i\omega\mathbf{R}_{\ell(z)}^{-1}\cdot\mathbf{g}(z,z')\cdot\mathbf{T}_{\ell(z')};$$

then

$$\tilde{e}^{(2)}=\tilde{e}^{(1)}+\int_{-\infty}^{\infty}dz'\,\mathbf{G}\cdot\mathscr{F}[\zeta_{\ell(z')}^{(\Delta)}\cdot\mathbf{E}^{(2)}]. \tag{8.1}$$

We wish to obtain from (8.1) an integral equation. Toward that goal, *assume* the flaw affects $\hat{x}$ and $\hat{y}$ conductivities only[17]; that is, $\hat{z}\cdot\zeta_j^{(\Delta)}=0$. Let

$$\mathbf{G}_T\equiv\begin{pmatrix}1&0&0&0\\0&1&0&0\end{pmatrix}\cdot\mathbf{G},$$

$$\xi_j\equiv\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}\cdot\zeta_j^{(\Delta)}\cdot\begin{pmatrix}1&0&0\\0&1&0\end{pmatrix}^T,$$

$\varphi^{(m)}\equiv(E_x^{(m)},E_y^{(m)})^T$, and $\tilde{\varphi}^{(m)}\equiv\mathscr{F}[\varphi^{(m)}]$; then

$$\tilde{\varphi}^{(2)}=\tilde{\varphi}^{(1)}+\int_{-\infty}^{\infty}dz'\,\mathbf{G}_T\cdot\mathscr{F}[\xi_{\ell(z')}\cdot\varphi^{(2)}]. \tag{8.2}$$

Convolve $\mathscr{F}[\xi_{\ell(z')}\cdot\varphi^{(2)}]$ to obtain from (8.2) an integral equation for $\tilde{\varphi}^{(2)}$; alternatively, compute $\varphi^{(2)}=\mathscr{F}^{-1}[\tilde{\varphi}^{(2)}]$ using (8.2) and convolution to obtain an integral equation for $\varphi^{(2)}$.

Define new Green's functions $\mathbf{g}_\pm$ that satisfy Theorem 5.1 and have boundary conditions $\mathbf{g}_+(z\uparrow+\infty;z')=0$ and $\mathbf{g}_-(z\downarrow-\infty;z')=0$. The construction in Appendix B shows $\mathbf{g}_+(z,z')=0$, for $z>z'$, and $\mathbf{g}_-(z,z')=0$, for $z<z'$. Define $\mathbf{G}_{T\pm}$ and $\varphi_\pm^{(m)}$ as $\mathbf{G}_T$ and $\varphi^{(m)}$ are defined, but with $\mathbf{g}_\pm$ in place of **g**. Then

$$\tilde{\varphi}_+^{(1)}=\begin{pmatrix}1&0&0&0\\0&1&0&0\end{pmatrix}\cdot\mathbf{R}_{\ell(z)}^{-1}\cdot\int_z^{+\infty}dz'\,\mathbf{g}_+\cdot\mathbf{T}_{\ell(z')}\cdot\tilde{\mathbf{J}},$$

$$\tilde{\varphi}_-^{(1)}=\begin{pmatrix}1&0&0&0\\0&1&0&0\end{pmatrix}\cdot\mathbf{R}_{\ell(z)}^{-1}\cdot\int_{-\infty}^{z}dz'\,\mathbf{g}_-\cdot\mathbf{T}_{\ell(z')}\cdot\tilde{\mathbf{J}},$$

and

$$\tilde{\varphi}_+^{(2)}=\tilde{\varphi}_+^{(1)}+\int_z^{+\infty}dz'\,\mathbf{G}_{T+}\cdot\mathscr{F}[\xi_{\ell(z')}\cdot\varphi_+^{(2)}],$$

$$\tilde{\varphi}_-^{(2)}=\tilde{\varphi}_-^{(1)}+\int_{-\infty}^{z}dz'\,\mathbf{G}_{T-}\cdot\mathscr{F}[\xi_{\ell(z')}\cdot\varphi_-^{(2)}],$$

from which we obtain integral equations for $\varphi_\pm^{(2)}$. The integral equations for $\varphi^{(2)}$ and $\varphi_\pm^{(2)}$ are similar to Fredholm and Volterra equations used in one-dimensional quantum inverse scattering.[18,19]

ACKNOWLEDGMENT

Our work was funded through the Naval Surface Weapons Center, White Oak Laboratory, Contract No. N60921-86-C-0172.

APPENDIX A: DERIVATION OF (4.7)

Use (4.5) to multiply out $\Sigma_j=\mathbf{R}_j\cdot\mathbf{S}_j\cdot\mathbf{R}_j^{-1}$: $\Sigma_{11j}=\Sigma_{12j}=\Sigma_{21j}=\Sigma_{22j}=\Sigma_{33j}=\Sigma_{34j}=\Sigma_{43j}=\Sigma_{44j}=0$ and

$$\begin{aligned}
\Sigma_{13j}&=c_j^2S_{13j}+c_js_j(S_{23j}+S_{14j})+s_j^2S_{24j},\\
\Sigma_{14j}&=c_j^2S_{14j}+c_js_j(S_{24j}-S_{13j})-s_j^2S_{23j},\\
\Sigma_{23j}&=c_j^2S_{23j}+c_js_j(S_{24j}-S_{13j})-s_j^2S_{14j},\\
\Sigma_{24j}&=c_j^2S_{24j}-c_js_j(S_{23j}+S_{14j})+s_j^2S_{13j},\\
\Sigma_{31j}&=c_j^2S_{31j}+c_js_j(S_{32j}+S_{41j})+s_j^2S_{42j},\\
\Sigma_{41j}&=c_j^2S_{41j}+c_js_j(S_{42j}-S_{31j})-s_j^2S_{32j},\\
\Sigma_{32j}&=c_j^2S_{32j}+c_js_j(S_{42j}-S_{31j})-s_j^2S_{41j},\\
\Sigma_{42j}&=c_j^2S_{42j}-c_js_j(S_{32j}+S_{41j})+s_j^2S_{31j},
\end{aligned} \tag{A1}$$

where $c_j \equiv \cos\theta_j$ and $s_j \equiv \sin\theta_j$. Substitute from (4.4) for S_{lkj} in (A1) and multiply to obtain

$$\Sigma_{13j} = -\Sigma_{24j} = (-i/\omega\epsilon_j)[k_x k_y (c_j^2 - s_j^2) + c_j s_j (k_y^2 - k_x^2)],$$

$$\Sigma_{14j} = -i\mu_0\omega + (i/\omega\epsilon_j)(c_j k_x + s_j k_y)^2, \tag{A2}$$

$$\Sigma_{23j} = i\mu_0\omega - (i/\omega\epsilon_j)(c_j k_y - s_j k_x)^2,$$

$$\begin{aligned}\Sigma_{32j} &= [i\omega(c_j^2\zeta_{22j} - 2c_j s_j \zeta_{12j} + s_j^2\zeta_{11j}) - (i/\mu_0\omega)(c_j k_x + s_j k_y)^2],\\ \Sigma_{41j} &= [-i\omega(c_j^2\zeta_{11j} + 2c_j s_j\zeta_{12j} + s_j^2\zeta_{22j}) + (i/\mu_0\omega)(c_j k_y - s_j k_x)^2],\\ \Sigma_{31j} &= \{i\omega[\zeta_{12j}(c_j^2 - s_j^2) + c_j s_j(\zeta_{22j} - \zeta_{11j})] + (i/\mu_0\omega)[k_x k_y(c_j^2 - s_j^2) + c_j s_j(k_y^2 - k_x^2)]\},\\ \Sigma_{42j} &= \{i\omega[\zeta_{12j}(c_j^2 - s_j^2) + c_j s_j(\zeta_{22j} - \zeta_{11j})] - (i/\mu_0\omega)[k_x k_y(c_j^2 - s_j^2) + c_j s_j(k_y^2 - k_x^2)]\}.\end{aligned} \tag{A3}$$

Substitute from $\zeta_j = \mathbf{T}_j^{-1}\cdot\epsilon_j\cdot\mathbf{T}_j$ for ζ_{lkj} in (A3) to compute

$$\begin{aligned}\Sigma_{31j} &= -\Sigma_{42j} = -(\epsilon_j/\mu_0)\Sigma_{13j},\\ \Sigma_{32j} &= i\omega\epsilon_j - (i/\mu_0\omega)(c_j k_x + s_j k_y)^2,\\ \Sigma_{41j} &= -i\omega\epsilon_{xj} + (i/\mu_0\omega)(c_j k_y - s_j k_x)^2.\end{aligned} \tag{A4}$$

The term Σ_{13j} appears in (A2). Substitute in (A2) and (A4) for k_x and k_y in terms of k_{xj} and k_{yj} to obtain the nonzero elements in (4.7).

APPENDIX B: AN ALGORITHM FOR SEC. V

Suppress subscripts $\hat{e} = \hat{x},\hat{y},\hat{z}$. For $n = 1,2,3$, let

$$\tilde{\gamma}^{(n)} = \begin{cases} a_j^{(n)} v_{1j} e^{\lambda_{1j}(z - z_j)} + \cdots + d_j^{(n)} v_{4j} e^{-\lambda_{3j}(z - z_j)}, & z \in (M_0 \cup \cdots \cup M_{N-2}) \setminus (M_{k+} \cup M_{k-}),\\ a_+^{(n)} v_{1k} e^{\lambda_{1k}(z - z')} + \cdots + d_+^{(n)} v_{4k} e^{-\lambda_{3k}(z - z')}, & z \in M_{k+},\\ a_-^{(n)} v_{1k} e^{\lambda_{1k}(z - z')} + \cdots + d_-^{(n)} v_{4k} e^{-\lambda_{3k}(z - z')}, & z \in M_{k-},\\ a_{N-1}^{(n)} v_{1,N-1} e^{\lambda_{1,N-1}(z - z_{N-2})} + c_{N-1}^{(n)} v_{3,N-1} e^{\lambda_{3,N-1}(z - z_{N-2})}, & z \in M_{N-1}. \end{cases} \tag{B1}$$

Define $a_{N-1}^{(2)} \equiv c_{N-1}^{(3)} \equiv 1$ and $a_{N-1}^{(3)} \equiv c_{N-1}^{(2)} \equiv 0$, but assign any values to $a_{N-1}^{(1)}$ and $c_{N-1}^{(1)}$. As a further definition: the ($n = 1,2,3$) coefficients satisfy continuity condition (6.2); the ($n = 1$) coefficients satisfy (6.3); and the ($n = 2,3$) coefficients satisfy a *continuity* condition constructed by setting to zero the right-hand side of (6.3). [Equation (6.3) is the source of the $\hat{e}$ dependence of $\tilde{\gamma}$ and $\tilde{\gamma}^{(1)}$; $\tilde{\gamma}^{(2)}$ and $\tilde{\gamma}^{(3)}$ are $\hat{e}$ independent.] Coefficients of $\tilde{\gamma}^{(n)}$ in M_{N-1} are given by definition; coefficients in other layers will be computed iteratively. The computation yields continuous states $\tilde{\gamma}^{(2)}$ and $\tilde{\gamma}^{(3)}$, and a state $\tilde{\gamma}^{(1)}$ with the same discontinuity as $\tilde{\gamma}$.

This paragraph describes the iterative step. Let $j \in \{1,2,\ldots,N-2,k+,k-\}$ and let M_l be the layer immediately above M_j. Suppose the following are known: $a_j^{(n)}, b_j^{(n)}, c_j^{(n)}, d_j^{(n)}, \mathbf{R}_j, \mathbf{R}_l, \Upsilon_{r(z')}$, and all the eigenvectors, eigenvalues, and boundary coordinates appearing in (6.2) and (6.3). For $n = 1,2,3$, use the appropriate boundary condition from (6.2), (6.3), or the modified (6.3) to obtain an equation of the form $a_l^{(n)} v_{1l} \exp[w_1^{(n)}] + b_l^{(n)} v_{2l} \times \exp[w_2^{(n)}] + c_l^{(n)} v_{3l} \exp[w_3^{(n)}] + d_l^{(n)} v_{4l} \exp[w_4^{(n)}] = (r_1, r_2, r_3, r_4)^T$, where $\mathbf{r}$ and the exponents $w^{(n)}$ are computed in terms of known quantities listed in the previous sentence. Use (5.1) to rewrite the previous equation as

$$\begin{aligned}(\Sigma_{14l}/\lambda_{1l})(\alpha_l - \beta_l) &= r_1,\\ (\Sigma_{24l}/\lambda_{1l})(\alpha_l - \beta_l) + (\gamma_l + \delta_l) &= r_2,\\ (\Sigma_{32l}/\lambda_{3l})(\gamma_l - \delta_l) &= r_3,\\ (\alpha_l + \beta_l) - (\Sigma_{31l}/\lambda_{3l})(\gamma_l - \delta_l) &= r_4,\end{aligned} \tag{B2}$$

where

$$\begin{aligned} a_l^{(n)} &= (\alpha_l/\lambda_{1l}) \exp[-w_1^{(n)}],\\ b_l^{(n)} &= (\beta_l/\lambda_{1l}) \exp[-w_2^{(n)}],\\ c_l^{(n)} &= (\gamma_l/\lambda_{3l}) \exp[-w_3^{(n)}],\\ d_l^{(n)} &= (\delta_l/\lambda_{3l}) \exp[-w_4^{(n)}].\end{aligned}$$

The solution to (B2) is

$$\begin{aligned} a_l^{(n)} &= \frac{1}{2\lambda_{1l}}\left(r_4 + \frac{\Sigma_{31l}}{\Sigma_{32l}} r_3 + \frac{\lambda_{1l}}{\Sigma_{14l}} r_1\right) \exp[-w_1^{(n)}],\\ b_l^{(n)} &= \frac{1}{2\lambda_{1l}}\left(r_4 + \frac{\Sigma_{31l}}{\Sigma_{32l}} r_3 - \frac{\lambda_{1l}}{\Sigma_{14l}} r_1\right) \exp[-w_2^{(n)}],\\ c_l^{(n)} &= \frac{1}{2\lambda_{3l}}\left(r_2 - \frac{\Sigma_{24l}}{\Sigma_{14l}} r_1 + \frac{\lambda_{3l}}{\Sigma_{32l}} r_3\right) \exp[-w_3^{(n)}],\\ d_l^{(n)} &= \frac{1}{2\lambda_{3l}}\left(r_2 - \frac{\Sigma_{24l}}{\Sigma_{14l}} r_1 - \frac{\lambda_{3l}}{\Sigma_{32l}} r_3\right) \exp[-w_4^{(n)}].\end{aligned}$$

In that way, compute iteratively the coefficients $a_l^{(n)}, b_l^{(n)}, c_l^{(n)}, d_l^{(n)}$ for all $l \in \{0,1,\ldots,N-2,k+,k-\}$ and for $n = 1,2,3$.

Both vectors $\tilde{\gamma}^{(1)}$ and $\tilde{\gamma}$ are outward traveling in M_{N-1}, but $\tilde{\gamma}$ is outward traveling in M_0 and $\tilde{\gamma}^{(1)}$ is not. We seek p and $q \ni \tilde{\gamma}$ and $\tilde{\gamma}^{(1)} + p\tilde{\gamma}^{(2)} + q\tilde{\gamma}^{(3)}$ have identical outward-traveling boundary conditions; then $\tilde{\gamma} = \tilde{\gamma}^{(1)} + p\tilde{\gamma}^{(2)} + q\tilde{\gamma}^{(3)}$. Use (B1) and the outward-traveling condition from M_0 to show that p and q solve the 2×2 system

$$\begin{aligned} a_0^{(1)} + p a_0^{(2)} + q a_0^{(3)} &= 0,\\ c_0^{(1)} + p c_0^{(2)} + q c_0^{(3)} &= 0.\end{aligned} \tag{B3}$$

If the system has a unique solution then p, q is it and

$$\begin{aligned} a_j &= a_j^{(1)} + p a_j^{(2)} + q a_j^{(3)},\\ b_j &= b_j^{(1)} + p b_j^{(2)} + q b_j^{(3)},\\ c_j &= c_j^{(1)} + p c_j^{(2)} + q c_j^{(3)},\\ d_j &= d_j^{(1)} + p d_j^{(2)} + q d_j^{(3)}\end{aligned} \tag{B4}$$

solve the problem of computing exactly the coefficients of $\mathbf{g}$ for each M_j.

This appendix applies also to numerical computations with roundoff errors. Compounding of numerical errors is suppressed significantly by techniques described in Appendix D of Ref. 12; note also that scaling improves the numerical stability of solving (B3). The result—(B4)—of rounded numerical computations is not a precisely correct answer, though we expect (B4) to improve on the guess $a_{N-1}^{(1)}, c_{N-1}^{(1)}$. Thus we expect these appended computations to yield an improved guess, for which we can repeat the computations, hoping for further numerical improvement.

[1]D. W. Berreman, "Optics in stratified and anisotropic media: 4×4-matrix formulation," J. Opt. Soc. Am. **62**, 502 (1972).

[2]V. G. Daniele and R. S. Zich, "Radiation by arbitrary sources in anisotropic and stratified media," Radio Sci. **8**, 63 (1973).

[3]C. Altman, A. Schatzberg, and K. Suchy, "Symmetries and scattering relations in plane-stratified anisotropic, gyroscopic, and bianisotropic media," Appl. Phys. B **26**, 147 (1981).

[4]C. M. Krowne, "Determination of the Green's function in the spectral domain using a matrix method: Application to radiators or resonators immersed in a complex anisotropic layered medium," IEEE Trans. Antennas Propag. **AP-34**, 247 (1986).

[5]O. Hashimoto and Y. Shimizu, "Reflecting characteristics of anisotropic rubber sheets and measurement of complex permittivity tensor," IEEE Trans. Microwave Theory Tech. **MTT-34**, 1202 (1986).

[6]A. A. Mostafa, C. M. Krowne, and K. A. Zaki, "Numerical spectral matrix method for propagation in general layered media: Application to isotropic and anisotropic substrates," IEEE Trans. Microwave Theory Tech. **MTT-35**, 1399 (1987).

[7]J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon, New York, 1962), Chap. VIII, Sec. 9.

[8]D. E. Boerner, "The calculation of electromagnetic fields from an arbitrary source in a horizontally layered earth," M.S. thesis, University of Toronto, 1983.

[9]P. E. Wannamaker, G. W. Hohmann, and W. A. SanFilipo, "Electromagnetic modeling of three-dimensional bodies in layered earths using integral equations," Geophysics **49**, 60 (1984).

[10]W.-Y. Pan, "Surface-wave propagation along the boundary between sea water and one-dimensionally anisotropic rock," J. Appl. Phys. **58**, 3963 (1985).

[11]N. K. Das and D. M. Pozar, "A generalized spectral-domain Green's function for multilayer dielectric substrates with application to multilayer transmission lines," IEEE Trans. Microwave Theory Tech. **MTT-35**, 326 (1987).

[12]T. M. Roberts, H. A. Sabbagh, and L. D. Sabbagh, "Electromagnetic interactions with an anisotropic slab," to be published in IEEE Trans.

[13]Infinitely many anisotropic conductors admit plane waves that do not attenuate in their directions of propagation. See T. M. Roberts and H. A. Sabbagh, "A phase velocity effect peculiar to anisotropic conductors," unpublished. Copies are available from T. M. Roberts.

[14]
$$a<z \Rightarrow \partial_z \int_a^z dz'\, K(z,z') = K(z,z'\uparrow z) + \int_a^z dz'\, \partial_z K(z,z')$$

and

$$z<a \Rightarrow \partial_z \int_z^a dz'\, K(z,z') = -K(z,z'\downarrow z) + \int_z^a dz'\, \partial_z K(z,z').$$

[15]In Ref. 13 the nontriviality of identifying the direction of propagation in anisotropic conductors is discussed.

[16]J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK Users' Guide* (SIAM, Philadelphia, 1979), Chap. 2. See also Appendix D in Ref. 12. Appendix D has notes on condition numbers, scaling, and judicious order of summation.

[17]The assumption is for the sake of simplicity. For an integral equation obtained without the assumption, read J. R. Bowler, L. D. Sabbagh, and H. A. Sabbagh, "A theoretical and computational model of eddy-current probes incorporating volume integral and conjugate gradient methods," Sec. II, submitted for publication. Copies are available from H. A. Sabbagh.

[18]K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, New York, 1977), Chap. II, Sec. 1; Chap. XVII, Sec. 1.

[19]R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982), 2nd ed., Chap. 12, Sec. 1.

A new eight-vertex model with an infinite number of commensurate phases[a),b)]

Diptiman Sen
Department of Physics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

(Received 13 January 1988; accepted for publication 8 June 1988)

A symmetric eight-vertex model, containing four even vertices with real weights and four odd vertices with imaginary weights, is found to exhibit an infinite number of commensurate phases. The phase diagram is conjectured to be a complete devil's staircase similar to that of certain one-dimensional systems. Associated naturally with the model are two diffeomorphic one-dimensional maps whose asymptotic trajectories are either stable cycles or intermittently chaotic, depending on the phase.

I. INTRODUCTION

Baxter's solution of a symmetric eight-vertex model was an important development in statistical physics (Baxter[1]). The model includes, as special cases, a large number of Ising-like models in two dimensions that generically exhibit second-order phase transitions (Baxter[2]). The mapping from the vertices of the Baxter model to Ising spins located at the sites of the dual lattice (Wu,[3] Kadanoff and Wegner[4]) depends crucially on the fact that all the eight vertices are even, i.e., the number of incoming or outgoing arrows is even.

In Fig. 1, we show four even vertices and four odd ones. These, together with their reverse vertices (in which each arrow is reversed and the corresponding vertices are numbered 1' through 8'), constitute the 16 possible vertices in two dimensions. The Baxter model assigns real non-negative weights to the vertices 1234 and their reverses, and zero weights to the other eight vertices. The model is symmetric in that $w_i = w_{i'}$. So there are only four weights to be specified.

Now one may consider other possible symmetric eight-vertex models by assigning nonzero weights to a different combination of four vertices from Fig. 1. The total number of such models is $\binom{8}{4} = 70$. However, there are several symmetries present in these models (Fan and Wu[5]). For example, reversing all the horizontal arrows interchanges the vertices thusly: 1↔2, 3↔4, 5↔6, and 7↔8. There are other symmetries one can construct that map even vertices into odd ones and vice versa. For instance, each vertex may have either its right arrow or the other three arrows reversed. This interchanges 1↔5, 2↔6, 3↔7, and 4↔8.

Using all these symmetries reduces the number of independent eight-vertex models to six. These have vertices 1234 (Baxter), 1567, 1256, 1356, 1357, and 1457. The latter five contain both even and odd vertices, which makes it impossible to map them into two-dimensional Ising spin systems. This is probably why they have never been studied before.

Suppose now that we have one of these six models on a periodic lattice with M rows and N columns. The calculation of the partition function Z starts by defining four 2×2 matrices R_{++}, R_{--}, R_{+-}, and R_{-+} (Baxter[1]). The matrix element $(R_{\alpha\alpha'})_{\lambda\lambda'}$ is the weight given to the vertex shown in Fig. 2. Here α (α') = + (−) means that the corresponding arrow points up (down), and λ (λ') = 1 (2) means that the arrow points to the right (left).

Given two neighboring rows of vertical arrows $\alpha = (\alpha_1\alpha_2\cdots\alpha_N)$ and $\alpha'_1 = (\alpha'_1\alpha'_2\cdots\alpha'_N)$, the transfer matrix T is defined to be a $2^N\times 2^N$ matrix, whose elements are given by the trace of a product of N matrices,

$$T_{\alpha\alpha'} = \mathrm{Tr}\,(R_{\alpha_1\alpha'_1}R_{\alpha_2\alpha'_2}\cdots R_{\alpha_N\alpha'_N}). \tag{1}$$

The partition function is then

$$Z = \mathrm{Tr}\,T^M. \tag{2}$$

We define the energy per vertex of an infinite lattice to be

$$E = -\lim_{M,N\to\infty} (1/MN)\ln Z. \tag{3}$$

For the Baxter model, each of the four matrices $R_{\alpha\alpha'}$ is nonzero, which makes the calculation of Z rather difficult. After a brief examination of the other five models, we see that the simplest one is likely to be the one denoted 1256, because two of the matrices, R_{+-} and R_{-+}, are then zero. This paper is entirely about that model.

In Secs. II and III, we study two versions of 1256. The first one has only real weights and is very simple, having only

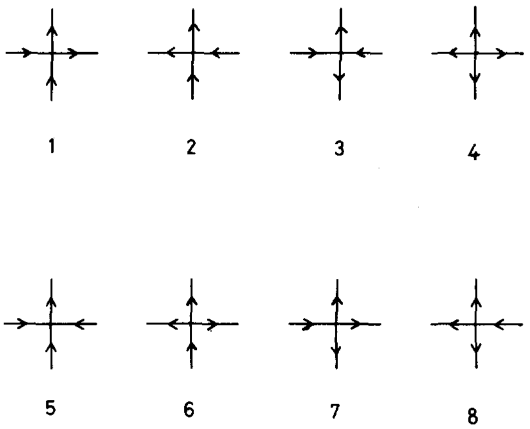


FIG. 1. The eight possible vertices with the top arrow pointing up.

[a)] In memory of Elizabeth Gardner.

[b)] Presented at the conference on "Mathematical Problems in Statistical Mechanics," Heriot-Watt University, U.K., 3–5 August 1987.

 0022-2488/88/122682-06$02.50

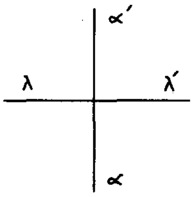


FIG. 2. The four arrow variables α, α', λ, and λ' of a vertex.

two phases. The second one has both real and imaginary weights and contains an infinite number of periodically modulated phases whose stabilities depend on the values of two parameters.

While the Baxter model describes realistic two-dimensional models, our model does not describe any physical model, because some of the weights are imaginary. However, it seems to have properties very similar to certain nonstatistical one-dimensional models described in Sec. V. This provides a motivation for the mathematical analysis of this model.

II. A SIMPLE MODEL OF TYPE 1256

The model has four even and four odd vertices and is described by the matrices

$$R_{++} = \begin{pmatrix} w_1 & w_5 \\ w_6 & w_2 \end{pmatrix}, \quad R_{--} = \begin{pmatrix} w_2 & w_6 \\ w_5 & w_1 \end{pmatrix}. \tag{4}$$

From Fig. 1 we see that the vertices are such that all the vertical arrows in a particular column point the same way—either up or down. With the configuration α of vertical arrows being the same from one row to the next, the transfer matrix is diagonal and is given by the elements

$$T_\alpha = \mathrm{Tr}\, R_{\alpha_1} R_{\alpha_2} \cdots R_{\alpha_N}. \tag{5}$$

We have used the diagonality to write $T_{\alpha\alpha'}$, R_{++}, and R_{--} more simply as T_α, R_+, and R_-.

Among all the configurations of vertical arrows, henceforth called phases, the periodically modulated ones will be found to play a special role. A convenient notation for these is the following. Given a number l lying in the interval $[0,1]$, the corresponding phase $P(l)$ will have the vertical arrow in the n column given by

$$s_n = \text{sign of } ([nl/2 + c] - \tfrac{1}{2}), \tag{6}$$

where $[x]$ is the fractional part of x satisfying $0 \leqslant [x] < 1$. In (6), s_n is either $+$ or $-$ and will be called a spin. The constant c has no effect on the energy and is only mentioned for generality.

If l is rational, the phase is called commensurate and it is modulated in a periodic way. Examples are $P(0) = +$, $P(\frac{1}{2}) = + + - -$, and $P(1) = + -$. If l is irrational, the phase is called incommensurate or quasiperiodic. There are also phases which repeat periodically but are not modulated in the sense of (6). These will not be given a special name.

Returning to the matrices R_+ and R_-, note that they satisfy

$$R_- = \sigma^1 R_+ \sigma^1, \tag{7}$$

where σ^1 is a Pauli matrix. Hence any phase and its reverse, obtained by interchanging $+$ and $-$, contribute equally to the partition function.

For convenience, we define the matrices

$$A = R_+, \quad B = \sigma^1 R_+. \tag{8}$$

Inside the trace in (5), any phase can equally well be written in terms of A's and B's.

For a commensurate phase $P(l)$, let $l = r/s$, where r and s are relatively prime and $s > r \geqslant 1$, except for $l = 0$ or 1 when we define s to be 1. In the enumeration of $P(l)$ in terms of $+$ and $-$, one needs either s or $2s$ spins, depending on whether r is even or odd. In terms of A or B however, one always requires s symbols of which r are B's and $s - r$ are A's. Examples are $P(0) = A$, $P(\frac{1}{2}) = AB$, and $P(1) = B$.

For $P(l)$, denote its matrix representation—product of s A's and B's—as $M(l)$. The matrix $M(l)$ has two eigenvalues. The magnitude of these, raised to the power s^{-1}, will be called $\Lambda(l)$ and $\lambda(l)$, where $\Lambda(l) \geqslant \lambda(l) \geqslant 0$. Note that

$$\Lambda(l)\lambda(l) = |\det A| = |\det B|. \tag{9}$$

The total number of phases is 2^N. This number does not lead to a finite entropy per vertex in the thermodynamic limit $M,N \to \infty$. In this limit, therefore, the energy is given entirely by the phase α (usually unique) for which T_α in (5) is the largest. Since a phase is just defined by the N spins s_n, the system is effectively one dimensional. The exponential of the energy comes from the phase l with the largest possible eigenvalue $\Lambda_{\max}(l)$,

$$e^{-E} = Z^{1/MN} \underset{M\to\infty}{\to} T_\alpha^{1/N} \underset{N\to\infty}{\to} \Lambda_{\max}(l). \tag{10}$$

It is now easy to compute the free energy of the 1256 model. One finds that among all phases, commensurate or otherwise, one of only two always dominates in the thermodynamic limit. There are $P(0) = A$ and $P(1) = B$. From (4), their larger eigenvalues are

$$\begin{aligned} \Lambda(0) &= \tfrac{1}{2}[w_1 + w_2 + ((w_1 - w_2)^2 + 4w_5w_6)^{1/2}], \\ \Lambda(1) &= \tfrac{1}{2}[w_5 + w_6 + ((w_5 - w_6)^2 + 4w_1w_2)^{1/2}]. \end{aligned} \tag{11}$$

All other phases have a $\Lambda(l)$ which lies in between these two values. All the $\Lambda(l)$ are equal if

$$(w_1 - w_2)^2 = (w_5 - w_6)^2, \tag{12}$$

in which case the system is completely disordered.

Away from the surface (12), the system goes into one of the completely ordered phases A or B. On crossing (12), there is a first-order phase transition.

The 1256 model with positive real weights is therefore extremely simple compared to the Baxter model 1234. It is effectively one dimensional, with two ordered phases separated by a first-order transition.

III. A DIFFERENT 1256 MODEL WITH INFINITE COMMENSURATE PHASES

Faced with the fairly trivial model of the last section, we may ask whether a different choice of weights in (4) can lead to a richer phase diagram. After some experimentation, we discovered that the answer is yes if the number w_5 and w_6 for the odd vertices are made imaginary! This would make no sense if the w's are statistical weights (i.e., real and positive). However, we discover that the mathematics of this model is unexpectedly similar to a class of one-dimensional systems.

We first note that since the total number of vertices of types 5, 5′, 6, and 6′ must be even in any configuration that satisfies periodic boundary conditions, the corresponding weight will be real. But it need not be positive and the weights of different configurations may partially cancel each other. For a phase P, let Z_P denote the absolute value of its contribution to the partition function, i.e., the absolute value of the sum of the weights of the various arrangements of the horizontal arrows.

To extract some physical sense from this mode, we *define* the most stable phase to be the one for which $\lim_{M,N\to\infty} Z_P^{1/MN}$ is the largest. The calculation of the energy for any phase $P(l)$ is then similar to the previous case with four real weights—multiply N matrices together (or only s matrices if the phase is commensurate) to get a matrix $M(l)$, calculate the absolute value of its larger eigenvalue, and raise it to the power N^{-1} (or s^{-1}) to obtain $\Lambda(l)$.

Consider now the matrices (4) with w_1, w_2 real but not necessarily positive and w_5, w_6 imaginary. The transfer matrix element (5) only involves a product and a trace. One is therefore free to perform a similarity transformation on R_+ and R_- by a matrix of the form $\exp(i\alpha\sigma^1)$. A transformation by σ^2 or σ^3 is also allowed since they anticommute with σ^1. With these manipulations, R_+ and R_- can be reduced to

$$R_+ = \begin{pmatrix} \cos\phi + y & i\sin\phi \\ i\sin\phi & \cos\phi - y \end{pmatrix}, \quad R_- = \begin{pmatrix} \cos\phi - y & i\sin\phi \\ i\sin\phi & \cos\phi + y \end{pmatrix}, \tag{13}$$

where $0\leqslant\phi\leqslant\pi/2$ and $y\geqslant 0$. An overall normalizing factor has been ignored since it affects the energy by the same additive constant in all phases. The number of essential parameters is therefore two, not four.

We can reduce the range of ϕ by half through the following observation. Defining the matrices A and B as before, it is easily shown that a phase $P(l)$ has the same energy at the point (ϕ,y) as the phase $P' = P(1-l)$, obtained by interchanging A and B, at the point $(\pi/2-\phi,y)$. This important property, which will be called parity, allows one to concentrate on the interval $[0,\pi/4]$ for ϕ.

Parity is a physically meaningful transformation in many cases. For instance, in the ANNNI model (Bak and von Boehm,[6] Fisher and Selke,[7] Villain and Gordon[8]), which also has an infinite number of commensurate phases, parity reverses the sign of the nearest neighbor interaction J_1 (making it antiferromagnetic) but keeps the second nearest neighbor interaction as it is. In the model of Bak and Bruinsma[9] mentioned in Sec. V, parity reverses the sign of the magnetic field.

We now begin to analyze the phase diagram of the model (13). At any point (ϕ,y), for some phases the two eigenvalues of the matrix $M(l)$ will be complex conjugate numbers. Since $\Lambda(l)\lambda(l) = 1-y^2$ by (9), these phases have

$$\Lambda(l) = \lambda(l) = (1-y^2)^{1/2}. \tag{14}$$

These phases will be called unstable.

For the other phases, the eigenvalues will be real and will satisfy

$$\Lambda(l) > (1-y^2)^{1/2} > \lambda(l). \tag{15}$$

The most stable or dominant phase is the one for which $\Lambda(l)$ is the largest since the energy is then lowest. All the other phases satisfying (15) are merely metastable.

For $y\geqslant 1$, it turns out that the most stable phase is A for all $0\leqslant\phi<\pi/4$. For $\phi=\pi/4$, A and B have the same free energy as expected from the symmetry under parity. The transition across $\phi=\pi/4$ is first order. The situation is reminiscent of the previous model with four real weights. Here, however, the energies of all the other phases are greater than, not equal to, those of A and B on the line $\phi=\pi/4$.

The interesting part of the phase diagram lies in the region $0<y<1$. Anticipating a detailed analysis, the form of the phase diagram is shown in Fig. 3.

The most easily derivable features of the diagram are as follows. Along the line $\phi=0$, as well as along $y=1$, the dominant phase is $P(0)=A$. Along $\phi=\pi/4$, it is $P(\frac{1}{2})=AB$. On the line $y=0$, all phases have the same energy, namely zero.

For y small but nonzero, an interesting structure unfolds. If $2\phi/\pi$ is fixed at a rational value, the dominant phase, as y goes to zero, is given by $l=2\phi/\pi$. If $l=r/s$, the free energy is

$$E(l) = -y/[s\sin(\pi/2s)] \tag{16}$$

to first order in y. As y increases from zero, the region of dominance of this phase initially opens up linearly and symmetrically about the line $2\phi/\pi = r/s$. The region is bounded by two lines with slopes

$$\left.\frac{dy}{d\phi}\right|_{y=0} = \pm 2s\cos\frac{\pi}{4s}. \tag{17}$$

Phases of high periodicity are narrow. Beyond a point where y is of order $1/s$, the region no longer opens up but starts contracting (except for the phase $l=0$ which never contracts). The derivation of (17) and the rates of contraction will be discussed later.

In view of (17), it is clear that all commensurate phases must extend at least a part of the way up from $y=0$. We have been unable to discover completely what happens further into the interior of Fig. 3, though we will make a conjecture.

Intuitively, one may expect either of three possibilities.

(a) Some of the commensurate phases disappear as y increases. More precisely, for any nonzero y, all phases with s larger than some number (which increases as y decreases)

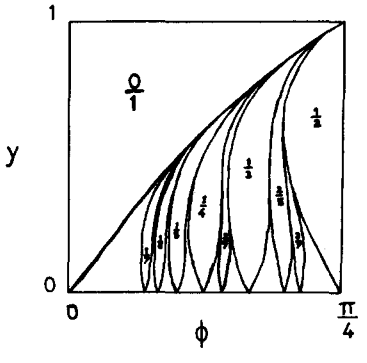


FIG. 3. The ten largest domains of the phase diagram. The rational numbers shown are the modulation numbers $l=r/s$ of the most stable phases, with $s<7$.

are absent. There are then only a finite number of commensurate phases. This is called a harmless staircase.[6] Incommensurate phases may or may not appear in between the commensurate ones.

(b) None of the commensurate phases disappear. This means that for any y, as ϕ increases from 0 to $\pi/4$, between any two phases there will be a third one. This situation has been picturesquely named the devil's staircase (Mandelbrot[10]). Each commensurate phase $l = r/s$ has a finite width in ϕ denoted $\Delta\phi(l,y)$, and it is clear that $\Delta\phi \to 0$ as $s \to \infty$. The devil's staircase is called complete if these phases fill up the phase diagram.

(c) Almost the same as in (b), except that incommensurate phases appear in between the commensurate ones. The devil's staircase is then called incomplete.

Possibility (a) seems unlikely for the following reason. It is easily shown that all phases have the same energy, $-\frac{1}{2}\ln 2$, at the point $\phi = \pi/4$, $y = 1$. As one descends slightly from that point with a slope $dy/d\phi = 1$, all phases continue to have the same energy to first order. This suggests that all commensurate phases survive up to $y = 1$, albeit with widths that go to zero as y goes to 1. This is difficult to reconcile with scenerio (a) in which commensurate phases with high periodicities progressively drop out, as y increases, by becoming unstable to other phases.

We find it more difficult to analytically distinguish between possibilities (b) and (c), though we conjecture that the devil's staircase is complete as in (b).

The following statement is easily proved. If l_i is a fixed irrational number and l_r denotes rational numbers, then as l_r approaches l_i, $\Lambda(l_r) \to \Lambda(l_i)$. Hence commensurate phases can approximate the energy of incommensurate phases arbitrarily accurately. Though this does not prove that the devil's staircase is complete, we see that the only phases one needs to consider in the phase diagram are the commensurate ones.

How then is Fig. 3 derived? To begin, consider the phase $l = 0$. This has

$$\Lambda(0) = \cos\phi + (y^2 - \sin^2\phi)^{1/2}. \qquad (18)$$

A sequence of phases approaching $l = 0$ is the series $A^{s-1}B$, $s = 1,2,3,\ldots$, for which $l = 1/s$. As $s \to \infty$, we find that

$$\Lambda(1/s) = \Lambda(0)(\sin^2\phi/(y^2 - \sin^2\phi))^{1/2s} \qquad (19)$$

ignoring terms which are exponentially small in s. The term $\Lambda(1/s)$ wins or loses with respect to $\Lambda(0)$ if the second factor in (19) is greater than or less than one. The dividing line is therefore

$$y = 2^{1/2}\sin\phi. \qquad (20)$$

This is the boundary of the region $l = 0$ in Fig. 3. The condition (20) does not depend on how the phases l approach zero. While we used the sequence $l = 1/s$ for simplicity, the sequence $2/(4s+3)$ would give the same result.

The above calculation is identical in spirit to the one that is usually done in the one-dimensional models mentioned in Sec. V. The domain of stability of a commensurate phase is always found by asking when it becomes energetically favorable to introduce a single discommensuration (McMillan[11]), or several discommensurations with a periodicity much larger than the phase being studied. A discommensuration is a link where the periodic structure is broken—the spins on either side of the lin are in the desired phase but they do not match across the link. Thus A^sB corresponds to alternating strings of R_+ and R_-, each string being in the phase $l = 0$ for a length of $s + 1$ spins.

As another example, we consider the phase $P(\frac{1}{2}) = AB$ and examine the sequence $(AB)^sA$ which has $l = 2s/(2s+1)$. On comparing it to $\Lambda(AB)$ in the limit $s \to \infty$, one obtains the condition

$$[\cos\phi + [(\sin\phi)/x](\cos 2\phi + y^2)] = \sin 2\phi + x$$

with

$$x = [\sin^2 2\phi - (1 - y^2)^2]^{1/2}. \qquad (21)$$

This is the left boundary of the region $l = \frac{1}{2}$ in Fig. 3. Note that both the curves (20) and (21) go to the point $\phi = \pi/4$, $y = 1$ with slope 1, as expected from a previous discussion.

The equations for the boundaries of phases with larger periodicities become complicated quite rapidly although certain asymptotic statements can be made. The higher phases, with s going up to 7, have been drawn in Fig. 3 in a qualitative way keeping in mind their shapes under the starting and ending points at $y = 0$ and 1.

This naturally brings us to the question, how does the width of a phase $l = r/s$ shrink either as $s \to \infty$ or as $y \to 1$?

We first define a variable $\mu(l,\phi,y)$,

$$\mu(l) = \lambda(l)/\Lambda(l) = (1 - y^2)/\Lambda^2(l). \qquad (22)$$

Note that this goes from 1 to 0 as y goes from 0 to 1. The energy follows from $\mu(l)$ through $E(l) = \frac{1}{2}\ln\mu(l) - \frac{1}{2}\ln(1 - y^2)$. In this eight-vertex model, $\mu(l)$ proves to be a very useful object.

Suppose now that y is kept fixed in Fig. 3 and we approach the right or left boundary, $B_R(l)$ or $B_L(l)$, of a particular phase l by decreasing or increasing ϕ. We successively pass through the regions of phases $l' = r'/s'$ such that $l' \to l$ and $s' \to \infty$. Then the width $\Delta\phi$ of the phase l' can be shown to decrease exponentially as

$$(1/s')\ln \Delta\phi(l',y)\,|_{B_{R,L}(l)} \underset{s\to\infty}{\to} \ln \mu_{R,L}(l,y), \qquad (23)$$

where $\mu_{R,L}(l,y)$ denotes the value of $\mu(l,y)$ on the lines $B_{R,L}(l,y)$. Thus $\mu_{R,L}(l,y)$ depends on both y and l and also on whether we are on the line $B_R(l)$ or $B_L(l)$. This is in contrast to the models mentioned in Sec. V where $\lim_{s'\to\infty}(1/s')\ln\Delta\phi(l' = r'/s')$ depends on y and nothing else.

The simplest example of the dependence of μ on y comes from $B_R(0)$, given by (20), where

$$\mu_R(0,y) = [(2 - y^2)^{1/2} - y]/[(2 - y^2)^{1/2} + y]. \qquad (24)$$

In the next section, we describe an interesting way of computing the energy of any phase which is not unstable.

IV. ONE-DIMENSIONAL MAPS, STABLE CYCLES, AND CHAOS

There is a different way of determining $\mu(l)$ which can easily be implemented on a programmable calculator. This quickly leads into the field of chaos.

The idea is to look at how a column vector changes under successive applications of the matrices A and B. For a phase $l = r/s$, consider the string $M(l)$ of A's and B's which has a periodicity s. If the phase is not unstable, the eigenvalues of $M(l)$ are real and of unequal magnitude, and the eigenvectors must be of the form

$$\begin{pmatrix} \cos\theta \\ i\sin\theta \end{pmatrix}, \tag{25}$$

where θ is an angular variable of period π, satisfying $-\pi/2 < \theta \leqslant \pi/2$. On acting with the successive matrices A or B of the string $M(l)$, θ changes. If $\Lambda(l)$ and $\lambda(l)$ are real and unequal, the sequence of θ's tends to a stable cycle of periods at a rate given by

$$\lim_{n\to\infty} \frac{\theta_{n+s} - \theta_n}{\theta_n - \theta_{n-s}} \to \mu^s(l). \tag{26}$$

If $M(l)$ has complex eigenvalues instead, then the successive iterations of θ generate a chaotic trajectory.

The matrix A acting on θ_n produces θ^A_{n+1}, where

$$\tan\theta^A_{n+1} = \frac{\sin(\theta_n + \phi) - y\sin\theta_n}{\cos(\theta_n + \phi) + y\cos\theta_n}. \tag{27}$$

If B acts on θ_n, we get θ^B_{n+1} which is related to θ^A_{n+1} by

$$\theta^B_{n+1} = \theta^A_{n+1} + \pi/2. \tag{28}$$

The maps A and B are diffeomorphisms (invertible and differentiable) if $y < 1$.

Of all the phases that give stable cycles, the one in which the stable cycle is aproached the fastest is the most stable phase because it has the least energy.

In order to understand the nature of chaos in this model, let us consider a fixed phase l and ask how the asymptotic form of the trajectory generated by iterations of the mapping $M(l)$ changes with the parameters (ϕ, y). For example, take $l = 0$, keep y fixed at some nonzero value, and increase ϕ from 0 to $\pi/4$. For $2^{1/2}\sin\phi < y$, this phase is the most stable. Between $2^{1/2}\sin\phi = y$ and $\sin\phi = y$, the map still has a stable fixed point but this is not the most stable cycle. Finally, for $\sin\phi > y$, the trajectory becomes chaotic. At $\sin\phi = y$, $\mu(0)$ takes the value 1.

The transition to chaos occurs via intermittency (Pomeau and Manneville[12]), rather than through a cascade of period doublings (Feigenbaum[13]). This is true for any phase l when the line $\mu(l) = 1$ is crossed.

The neigborhood of the transition to chaos is marked by very clear signals. On the stable side of the critical line $\phi = \phi_c(l, y)$, there is a single stable cycle $(\theta_1\theta_2\cdots\theta_s)$ to which all points flow except for s points which form an unstable cycle. On approaching $\phi = \phi_c$, the rate of flow slows down as μ goes to 1 from below in proportion to $|\phi - \phi_c|^{1/2}$. The stable cycle and the unstable one come together at the critical line. On the chaotic side, there are no cycles, stable or unstable, of any periodicity (except for a countable set of points in θ) and the iterated trajectory is sensitive to the initial point. The trajectory slows down temporarily when it passes close to the would-be stable cycle which now lies near but not on the real axis for θ. The number of iterations required to cross a fixed small interval near the cycle blows up as $|\phi - \phi_c|^{-1/2}$.

V. PREVIOUSLY STUDIED ONE-DIMENSIONAL MODELS

That the phase diagram of a one-dimensional model may be a devil's staircase has been known for a long time (Aubry,[14] Bak and Bruinsma[9]). Regarding the experimental situation, Aubry[15] and Bak[16] mention several systems that exhibit a large number of commensurate phases as some parameter is changed. Practically, of course, one cannot observe all the steps of a devil's staircase because most of them are extremely narrow and sensitive to small changes in the physical variables.

All the theoretical models contain frustration in that they either have two competing periodicities[14] or they have long range convex antiferromagnetic interactions competing with either a short range ferromagnetic interaction or a uniform magnetic field.[9] In our model, there is frustration in the sense that the even vertices by themselves would prefer phase A, while the odd vertices would prefer phase B.

Aubry[17] solved a very general class of models and proved that they have the devil's staircase property. The ground state phase diagram for these models and Fig. 3 are so similar that one might imagine that they would be identical after some complicated transformation. While this may be true for the diagrams, we have not been able to prove an equivalence directly at the level of the equations that determine the energy. The difficulties in attempting to do this are twofold. First, the earlier models start by defining an energy function which is always real. In our case, the energy is ill-defined in the unstable phases. Second, our model naturally leads to one-dimensional maps that do not seem to exist in the other models. Though the original Aubry model[14] has a two-dimensional map, there does not appear to be any connection between that and the one considered in Sec. IV.

It is an open question whether or not the eight-vertex model lies in the same class as the ones discussed by Aubry.[17]

VI. FINAL REMARKS

We have presented an eight-vertex model that has an infinite number of commensurate phases perhaps forming a complete devil's staircase. It is necessary to derive an exact expression for the minimum energy as a function of (ϕ, y) to determine if this is so. If it is true, one will have a new kind of model with this remarkable property. One may then ask what the fractal dimension d of the staircase is. For a fixed value of $y > 0$, let the total width of all the steps which are each less than a number r wide go as r^{1-d} as $r \to 0$. Due to the exponential decrease in step widths (23), the value of d is zero, just as in Aubry's case.[14,17]

The model gives rise to two diffeomorphic one-dimensional maps whose mathematical properties may be of interest in the context of chaos. The maps depend on certain parameters, collectively denoted by t. An interesting question for further investigation might be the following. Given two diffeomorphic maps A_t and B_t, not necessarily represented by marices as in this paper, but related by the shift (28), when is it true that the sequence of most stable phases forms a devil's staircase of modulation numbers as t is varied? Here stability is defined by the rate of approach to the stable cycle.

ACKNOWLEDGMENTS

I thank E. Gardner for stimulating discussions.

I also thank the U. K. Science and Engineering Research Council for financial support through Grant No. NG 17094.

[1]R. J. Baxter, Ann. Phys. (NY) **70**, 193 (1972).
[2]R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, London, 1982).
[3]F. Y. Wu, Phys. Rev. B **4**, 2312 (1971).
[4]L. P. Kadanoff and F. J. Wegner, Phys. Rev. B **4**, 3989 (1971).
[5]C. Fan and F. Y. Wu, Phys. Rev. B **2**, 723 (1970).
[6]P. Bak and J. von Boehm, Phys. Rev. B **21**, 5297 (1980).
[7]M. E. Fisher and W. Selke, Phys. Rev. Lett. **44**, 1502 (1980).
[8]J. Villain and M. B. Gordon, J. Phys. C **13**, 3117 (1980).
[9]P. Bak and R. Bruinsma, Phys. Rev. Lett. **49**, 249 (1982).
[10]B. B. Mandelbrot, *Fractals Form Chance and Dimension* (Freeman, San Francisco, 1977).
[11]W. L. McMillan, Phys. Rev. B **14**, 1496 (1976).
[12]Y. Pomeau and P. Manneville, Commun. Math. Phys. **74**, 189 (1980).
[13]M. J. Feigenbaum, J. Stat. Phys. **19**, 25 (1978).
[14]S. Aubry, *Solitons and Condensed Matter Physics*, edited by A. Bishop and T. Schneider (Springer, Berlin, 1978), p. 264.
[15]S. Aubry, J. Phys. (Paris) **44**, 147 (1983).
[16]P. Bak, Rep. Prog. Phys. **45**, 587 (1982).
[17]S. Aubry, J. Phys. C **16**, 2497 (1983).

A general representation for the effective dielectric constant of a composite

G. F. Dell'Antonio
Department of Mathematics, University of Rome, I-00185 Rome, Italy
V. Nesi
Courant Institute, New York University, New York, New York 10003

(Received 16 January 1987; accepted for publication 22 June 1988)

In its dependence on the dielectric constants ϵ_k of its homogeneous components, the effective dielectric constant ϵ^* of a composite is a function analytic in a domain Ω. Some relevant results about the effective dielectric constant are collected, including the form of Ω, and then a general representation of ϵ^* as an analytic function is given. In this representation, the dependence on the geometry is separated from the dependence on the ϵ_k''s.

I. INTRODUCTION

Representations for the effective dielectric constant ϵ^* of a composite material have been considered recently.[1,2] They allow us to find *a priori* bounds on ϵ^* when only partial information is available, through moment inequalities[2,3] or combined with asymptotic expansions.[4]

Consider a dielectric composite with N homogeneous components. Denote by ϵ^i the dielectric tensor of the ith component and by ϵ^i_α, $\alpha = 1,2,3$, its eigenvalues. It will be convenient to relabel the eigenvalues and consider ϵ^* as a function of $\epsilon \equiv \{\epsilon_1,...,\epsilon_M\}$; one will have $M < N$ when some of the components have a symmetry (e.g., are isotropic) so that some eigenvalues coincide.

One seeks a representation in the form

$$\epsilon^*(\epsilon) = \int F(\lambda,\epsilon)\mu(d\lambda), \tag{1.1}$$

where μ has support in a subset of R^M and is completely determined by the geometry of the composite, while $F(\lambda,\epsilon)$ are known functions. Typically, $F^{-1}(\lambda,\epsilon)$ is a polynomial in ϵ and λ.

The most natural way to derive a representation of the form (1.1) is to consider $\epsilon^*(\epsilon)$ as the boundary value of a function $\epsilon^*(z)$, $z \in C^M$, analytic in a domain Ω, and to make use of the representation theory for analytic functions. The mathematical properties of Maxwell's equations determine the analyticity domain Ω.

Since ϵ^* is homogeneous of order 1 in the ϵ_i, one can describe the domain Ω using complex projective variables or setting $\epsilon_{i_0} = 1$ for some chosen i_0. The latter procedure spoils the symmetry, but is often convenient for practical purposes.

If $M = 2$ (i.e., two isotropic components), the domain of analyticity, expressed in the variable $\epsilon_1 \cdot \epsilon_2^{-1}$, is the complex plane cut along the negative real axis. In this case, the representation is particularly simple, well known, and well exploited.[1]

When $M > 2$ the domain Ω is of less familiar type. The purpose of this paper is to give a general representation of a function analytic in Ω in terms of measures supported by a distinguished subset of the boundary of Ω. When homogeneous variables are used, the distinguished boundary has real dimension $M - 1$ and is the union of real hyperquadrants.

We shall use the Martinelli–Bochner representation in the form developed by Weyl for analytic polyhedra.[5] In our case, the difficulty comes from the behavior of ϵ^* near the distinguished boundary. To control it, we shall use information that can be derived from Maxwell's equations.[6]

The content of the paper is the following. In Sec. II we assemble notations and recall some useful properties of the effective dielectric constant. In Sec. III we describe the relevant features of Weyl's analysis and give the representation for M arbitrary. In Sec. IV we study the case $M = 3$ and give details on the control of the limits involved. The general case is treated along the same lines, with an increase in bookkeeping difficulties; the details will not be given here. In Sec. V we study some properties of the representation for $M = 3$, and give some simple connections with the functional representation developed in Ref. 6.

We end this introduction by comparing the representation given here with the one proposed recently by Golden.[2] He works with the variables $\zeta_i = z_i/z_M$, $i = 1,...,M-1$, and remarks that Ω contains as a subdomain the set $\mathbf{D}_1 \times ... \times \mathbf{D}_{M-1} \equiv \mathbf{D}$, where $\mathbf{D}_i \equiv \{\zeta \,|\, \mathrm{Im}\, \zeta_i > 0\}$. Golden gives then a representation of type (1.1) which holds for every function analytic in $\mathbf{D}$, with suitable behavior at the boundary.

The representation contains a positive measure supported by a subset Σ' of $\partial \mathbf{D}$; Σ' is again the union of hyperquadrants in R^{M-1}. Positivity is a consequence of the fact that $\mathrm{Im}\, \epsilon^* > 0$ whenever $\mathrm{Im}\, \zeta_i > 0$, $i = 1,...,M-1$.

In the representation given in Ref. 2, as well as in ours, the measure on Σ' carries all the information about the geometry of the composite. The dependence on ϵ appears explicitly in the function $F(\epsilon,\zeta)$.

An important feature of the representation introduced by Golden is that the measure carried by Σ' is positive. One can therefore adapt methods and results from the theory of moments to provide *a priori* estimates on ϵ^*. In the representation derived here some of the measures are signed, so that a straightforward application of the method of moments is not possible.

On the other hand, there is no simple way to characterize those measures μ on Σ' that lead to a function which is not only analytic in $\mathbf{D}$ but has also an analytic extension to Ω. There are, of course, choices of μ for which the extension is

 0022-2488/88/122688-07$02.50

obvious, and in fact the function is analytic in a much larger domain. In the case $M = 3$ some of these cases have been worked out and correspond to simple geometries.

The general situation is unclear, but there are indications that estimates for these particular cases can lead to general *a priori* bounds on ϵ^* (Ref. 3).

II. NOTATION AND PRELIMINARY RESULTS

Let $D \subset R^3$ be a domain filled with a dielectric composite, made of N homogeneous components. The ith component occupies a domain $D_i \subset D$, with characteristic function χ_i, and has dielectric tensor ϵ_i.

For a given boundary condition on ∂D, let $E(x)$ be the solution of Maxwell's equations for electrostatics. The effective dielectric constant is defined as

$$\epsilon^* = \int (E,\epsilon E)(x)dx, \tag{2.1}$$

where $\int$ denotes average and $\epsilon(x) = \Sigma \epsilon_i \chi_i(x)$. By definition, the domains D_i and the boundary condition at ∂D define the geometry of the composite.

Let $\bar{\epsilon}$ be the average of $\epsilon(x)$ for the given geometry

$$\bar{\epsilon} = \int (B,\epsilon B)dx,$$

where B is a vector field that specifies the boundary conditions (in significant cases, B is a constant vector field that represents the electric field for a homogeneous isotropic medium under the same boundary conditions).

Obviously $\bar{\epsilon} = \Sigma c_i \epsilon_i$, where c_i depends only on the geometry. Define

$$\epsilon' \equiv \bar{\epsilon} - \epsilon^*. \tag{2.2}$$

We shall give a representation for ϵ'.

For fixed geometry ϵ' is real analytic in $(R^+)^M$. Standard arguments (Refs. 2 and 6) show that ϵ' has a unique extension to a function (still denoted by ϵ') of M complex variables (denoted by z_i, $i = 1,...,M$), analytic in a domain $\Omega \subset C^M$ which is defined as follows: Ω is the simply connected domain which contains the point $z_i = 1$ $\forall i$ and is bounded by the hyperplanes

$$\sum_{ij} \equiv \{z | z_i \cdot z_j^{-1} \in (-\infty,0]\}.$$

Alternatively, one can describe Ω as the set of all those z in C^M for which there exists in C a straight line through the origin which leaves all the z_i's on the same side.

From the definition of ϵ^* and the properties of Maxwell's equations one can derive some further properties of ϵ^*. We shall list here the ones which will be used in the sequel. They can be verified, e.g., using the functional representation given in Ref. 6; in particular, property (d) is due to the fact that the limit described there is related to the spectral measure of a bounded self-adjoint operator.

(a) The function ϵ' is homogeneous of degree 1. Since $z \in \Omega$ implies $z_i \neq 0$ for all i, one can set $z_M = 1$ and regard z' as a function of $z_1,...,z_{M-1}$. In these variables, the boundary $\partial\Omega$ is a collection of points for which $z_{k_0} \geqslant 0$ for an index $k_0 \neq M$, and either $\operatorname{Im} z_i \geqslant 0$, $i \neq k_0$, $i \neq M$, or $\operatorname{Im} z_i \leqslant 0$, $i \neq k_0$, $i \neq M$.

(b) Let $z_i = 1 + \lambda\sigma_i$, $\lambda \in R$. For any choice of $\sigma_i \in C$, ϵ' is infinitesimal in λ of order 2.

(c) Let A be a subset of $\{1,...,M\}$. Set $z_k = w$ if $k \in A$, $z_i = \zeta$ if $i \notin A$. As a function of w and ζ, ϵ' is given by

$$\epsilon'(w,\zeta) = (w-\zeta)^2 \int \mu_A(d\lambda)(\lambda w + (1-\lambda)\zeta)^{-1}, \tag{2.3}$$

where μ is a positive measure supported in $[0,1]$.

To state the next property, consider the subset Σ of $\partial\Omega$ defined as follows:

$$\Sigma \equiv \{z | z \in \partial D,\ z_i/z_j \in R\ \ \forall i,j, \text{ and } z_k/z_m \in R^- \text{ for at least one pair } k,m.\}$$

Note that Σ identifies a subset $\underline{\Sigma}$ of the real projective space RP^M. For $z \in \Sigma$, define A^+ as

$$m \in A^+ \text{ iff } z_m/z_M \in (0,\infty) \tag{2.4}$$

and consider the path $z(\delta)$, $\delta \in R$, given by

$$z_m(\delta) = z_m + i\delta z_M, \quad \text{if } m \in A^+,$$
$$z_m(\delta) = z_m, \quad \text{otherwise}.$$

It is straightforward to verify that $z(\delta) \in \Omega$, if $\delta \neq 0$. Moreover for every $\delta > 0$ the continuous path

$$z_m(\delta,\tau) = z_M\tau + (1-\tau)z_m(\delta), \quad 0 < \tau < 1, \tag{2.5}$$

lies in Ω and is such that $z(\delta,0) = z(\delta)$, $z_m(\delta,1) = z_M$ $\forall m$.

One can now state the following property.

(d) Let $z \in \Sigma$. Then

$$\lim_{\delta \downarrow 0}\{\epsilon'(z(\delta)) - \epsilon'(z(-\delta))\}$$

exists and defines a real-valued measure μ on Σ, which extends to a real-valued measure $\boldsymbol{\mu}$ on $\underline{\Sigma}$. The measure $\boldsymbol{\mu}$ is uniformly bounded on continuous functions. In terms of inhomogeneous coordinates $x_1,...,x_{M-1}$, this means that one can find $c > 0$ such that

$$\left|\int \mu(dx)g(x)\right| \leqslant c(|g|_1 + |g|_\infty), \quad \text{for every } g \in L_1 \cap C.$$

III. THE REPRESENTATION FOR ARBITRARY *M*

Let $\sigma > 0$ and denote by Δ_σ the disk in C that has radius $\sqrt{\sigma}$ and center at $(\sqrt{\sigma} + (\sqrt{\sigma})^{-1}, 0)$. Denote by D_σ the image of Δ_σ under the map $z \to z^2$. One has $D_\sigma < D_{\sigma'}$ if $\sigma < \sigma'$, and $\cup_{\sigma > 0} D_\sigma = C - R^-$.

Let

$$\mathbf{D} \equiv \left\{ z \in C^M, \prod_1^M z_i \neq 0 \right\} \tag{3.1}$$

and consider the functions $Z_\alpha(z)$, analytic in $\mathbf{D}$, defined by

$$Z_i = z_i, \quad i = 1,...,M, \tag{3.2a}$$

$$Z_{(i,j)}(z) = z_i/z_j, \quad i \neq j, \quad i,j = 1,...,M. \tag{3.2b}$$

We define $P(M)$ so that one has $\alpha = 1,...,P(M)$. Set

$$\Omega_\alpha{}^\sigma \equiv \{z \in D \,|\, Z_\alpha(z) \in D_\sigma\} \tag{3.3}$$

and consider the domain

$$W^\sigma \equiv \bigcap_1^{P(M)} \Omega_\alpha{}^\sigma. \tag{3.4}$$

Let Ω^σ be the connected component of W^σ that contains the point $z_k = 1\ \forall k$. The set Ω^σ is open, bounded, and simply connected. By construction it is an analytic polyhedron in the sense of Weyl.[5] Moreover, it is easy to verify that

$$\bigcap_{\sigma>0} \Omega^\sigma = \lim_{\sigma\to 0} \Omega^\sigma = \Omega, \tag{3.5}$$

where Ω is the domain of analyticity defined in Sec. II. We must introduce the sets Ω^σ because Ω is unbounded and Weyl's theory does not apply directly. The properties (a)–(d) will be used to take the limit $\sigma\to\infty$ for the representations in Ω^σ.

The boundary $\partial\Omega^\sigma$ has the form

$$\partial\Omega^\sigma = \bigcup_1^{P(M)} \gamma_\alpha{}^\sigma, \quad \gamma_\alpha{}^\sigma \subset \partial\Omega_\alpha{}^\sigma. \tag{3.6}$$

Let $\{\alpha_1,...,\alpha_N\}$ be a subset of $\{1,..,P(M)\}$, with $\alpha_i \neq \alpha_j$ if $i\neq j$. Define

$$\gamma^\sigma(\alpha_1,...,\alpha_N) \equiv \bigcap_1^N \gamma^\sigma{}_{\alpha_k} \tag{3.7}$$

and call the skeleton of $\partial\Omega$ the oriented set

$$S^\sigma \equiv (-1)^{M(M+1)/2} \bigcup_{\mathrm{P}} \gamma^\sigma{}_\beta \tag{3.8}$$

where the union is taken over the set P of all the subsets of $\{1,...,P(M)\}$ which contain exactly N distinct components. With these notations, the Bochner–Martinelli formula states that, if $f(z)$ is analytic in $\Omega(\sigma)$ and continuous up to the boundary $\partial\Omega^\sigma$, then for every $z\in\Omega^\sigma$ one has

$$f(z) = (-1)^{M(M+1)/2}(2\pi i)^{-M} \times \sum_{\beta\in \mathrm{P}} \int_{\gamma(\beta,\sigma)} f(\lambda) B_\beta(\lambda,z) \prod d\lambda_k, \tag{3.9}$$

where $\lambda = \{\lambda_1,...,\lambda_N\}$ and B_β is defined as

$$B_\beta \equiv \det(Q(\alpha_{h,k})), \quad \beta = \{\alpha_1,...,\alpha_N\},$$
$$k,h = 1,...,N, \quad \alpha_1 < \cdots < \alpha_N, \tag{3.10}$$

$$Q(\alpha,k) \equiv P_{\alpha,k}(\lambda,z)\,[Z_\alpha(\lambda) - Z_\alpha(z)]^{-1}, \tag{3.11}$$

and the rational functions $P_{\alpha,k}$ satisfy

$$Z_\alpha(\lambda) - Z_\alpha(z) = \sum_1^N (\lambda_k - z_k) P_{\alpha,k}(\lambda,z). \tag{3.12}$$

The $P_{\alpha,k}$ are not uniquely determined by (3.12), but the representation obtained is independent of the choice made.

We now study the limit $\sigma\to\infty$ in (3.9).

Let b_R be the ball of radius R at the origin in C^N. Then, set theoretically

$$\lim_{\sigma\to\infty} \gamma^\sigma(\alpha_1,...,\alpha_N) \cap b_\sigma = \{z\in C^N,\ Z_{\alpha_i}\in(-\infty,0],\ i=1,...,N\} \equiv \gamma^\infty(\alpha_1,...,\alpha_N), \tag{3.13}$$

and the limit set is covered twice, coming from opposite directions in Ω, i.e., from $\{z|\operatorname{Im} Z_{\alpha_i} > 0\ \forall i\}$ and from $\{z|-\operatorname{Im} Z_{\alpha_i} > 0\ \forall i\}$.

Therefore, for every function analytic in Ω and with property (d), one has, for all $\beta \equiv \{\alpha_1,...,\alpha_N\}$,

$$\lim_{\sigma\to\infty} \int \gamma_{\sigma(\beta)\cap b_\sigma} B_\beta(\lambda,z) f(\lambda) \prod d\lambda_i = \int_{\gamma^\infty(\beta)} B_\beta(\lambda,z)\mu_\beta(d\lambda) \tag{3.14}$$

for a measure μ_β on $\gamma^\infty(\beta)$.

Notice that the set Σ defined in Sec. II is precisely

$$\Sigma = \bigcup_{\beta\in\mathrm{P}} \gamma^\infty(\beta). \tag{3.15}$$

From (3.14) one sees that one cannot directly use property (d) to perform the limit in the representation (3.7), since the integrand does not belong to L_1. Before taking the limit, one has therefore to apply a subtraction procedure and make use of properties (a)–(c).

This leads to an iteration scheme; the representation for $\epsilon'(\epsilon_1,...,\epsilon_M)$ contains, in addition to new terms, terms that also have the same structure as $\epsilon'(\delta_1,...,\delta_Q)$, where $\{\delta_1,...,\delta_Q\}$ is any subset of $\{\epsilon_1,...,\epsilon_M\}$. The general representation can best be described as follows. Consider in a plane M points $P_1,...,P_M$, no three of which are colinear. We call link (i,j) the nonoriented segment joining P_i with P_j. A graph γ is a collection of links. We use the notation $(i,j)\in\gamma$ to signify that (i,j) is one of the links of γ, and $i\in\gamma$ to signify that one can find j for which (i,j) belongs to γ. We say that k is distinguished for γ (in formulas, $k\underline{\in}\gamma$) if one can find two integers $i_1\neq i_2$ such that $(i_1,k)\in\gamma$, $(i_2,k)\in\gamma$.

Denote by $\Gamma(N)$ the collection of graphs that have N points, $N-1$ links, and no closed loops. Notice that if $N>2$ and $\gamma\in\Gamma(N)$, then γ contains at least one distinguished point. If $N=3$, the distinguished point is unique. We then have the following proposition.

Proposition (3.1) (representation formula): For every N, $3<N<M$, $\gamma\in\Gamma(M)$, $k\underline{\in}\gamma$, $p\in\gamma$, $p\neq k$, one can find real valued measures $\tau^N{}_{\gamma k}$, $\nu^N{}_{\gamma kp}$, $\mu^N{}_{,ij}$ such that

$$\epsilon'(\epsilon_1,...,\epsilon_M) = \sum_{n=2}^M \sum_{\gamma\in\Gamma(N)} \sum_{k\underline{\in}\gamma} \Big[A(\tau^N{}_{\gamma k}), + \sum_{\tilde p} B(\nu^N{}_{\gamma kp})\Big] + \sum_{\gamma\in\Gamma(2)} C(\mu^N{}_{,\gamma}), \tag{3.16}$$

where $\tilde p$ means that $(p,k)\in\gamma$, and moreover

$$A(\tau^N{}_{\gamma k})(\epsilon_1,...,\epsilon_M) \equiv \epsilon_k \prod_\gamma (\epsilon_i - \epsilon_j) \int_0^1 \cdots \int_0^1 \tau^N{}_{\gamma k} \Big(\prod_\gamma d\sigma_{ij}\Big) \times \Big(\prod_\gamma (\sigma_{ij}\epsilon_i + \sigma^*{}_{ij}\epsilon_j)\Big)^{-1},$$

$$B(\nu^N{}_{\gamma kp})(\epsilon_1\cdot\epsilon_M) \equiv \prod_{\gamma pk} (\epsilon_i - \epsilon_j) \int_0^1 \cdot \int_0^1 \nu^N{}_{\gamma kp} \Big(\prod_\gamma^{pk} d\sigma_{ij}\Big) \times \Big(\prod_{\gamma pk} (\sigma_{ij}\epsilon_j + \sigma^*{}_{ij}\epsilon_j)\Big)^{-1},$$

$$C(\mu^N{}_{,ij})(\epsilon_1,...,\epsilon_M) \equiv (\epsilon_i - \epsilon_j)^2 \int_0^1 \frac{\mu^N{}_{,ij}(d\sigma)}{\sigma\epsilon_i + \sigma^*\epsilon_j},$$

where $\sigma^* \equiv 1 - \sigma$.

The symbol Π_γ indicates the product over all links in γ and $\Pi_{\gamma pk}$ stands for the product over all links in γ with the exception of (p,k).

When $M=2$ one has of course

$$\epsilon'(\epsilon_1,\epsilon_2) = (\epsilon_1-\epsilon_2)^2 \int_0^1 \frac{\mu(d\sigma)}{(\sigma\epsilon_1+\sigma^*\epsilon_2)}. \tag{3.17}$$

For $M=3$ the general representation takes the form

$$\epsilon' = \sum{}' \epsilon_i(\epsilon_i-\epsilon_j)(\epsilon_i-\epsilon_k) \times \int_0^1\int_0^1 \frac{\tau_{jk}(d\sigma_j\times d\sigma_k)}{(\sigma_j\epsilon_i+\sigma^*{}_j\epsilon_j)(\sigma_k\epsilon_i+\sigma^*{}_k\epsilon_k)} + \sum{}'(\epsilon_i-\epsilon_j)(\epsilon_i-\epsilon_k)\left\{\int_0^1 \frac{\nu_{ij}(d\sigma_j)}{\sigma_j\epsilon_i+\sigma^*{}_j\epsilon_j} + \int_0^1 \frac{\nu_k(d\sigma_k)}{\sigma_k\epsilon_i+\sigma^*{}_k\epsilon_k}\right\} + \sum_{i<j}(\epsilon_i-\epsilon_j)^2\int_0^1 \frac{\mu_{ij}(d\sigma)}{\sigma\epsilon_i+\sigma^*\epsilon_j}, \tag{3.18}$$

where Σ' denotes the sum over permutations of the indices i,j,k, with $i\neq j\neq k\neq i$.

IV. THE REPRESENTATION FOR $N=3$

One has

$$\epsilon'(\epsilon_1,\epsilon_2,\epsilon_3) = \epsilon_3\epsilon'(\epsilon_1/\epsilon_3,\epsilon_1/\epsilon_3,1). \tag{4.1}$$

We shall construct a representation for the function

$$f(z_1,z_2) \equiv \epsilon'(z_1,z_2,1). \tag{4.2}$$

The properties (a)–(d) of Sec. II take the following form.

(a') One can find constants C_1 and C_2 such that, uniformly in z_1 and z_2,

$$|f(kz_1,z_2)| < (1+|k|)C_1, \quad |f(z_1,kz_2)| < (1+|k|)C_2.$$

(b') For every choice of $\zeta_1,\zeta_2 \in C$, the function $f(1+\delta\zeta_1, 1+\delta\zeta_2)$ is infinitesimal in δ of order 2.

(c') There exist positive measures μ_1,μ_2,μ_3, supported by $(-\infty,0]$ for which

$$f(z,1) = \int \frac{\mu_1(d\lambda)}{\lambda-z}; \quad f(1,z) = \int \frac{\mu_2(d\lambda)}{\lambda-z};$$

$$f(z,z) = \int \frac{\mu_3(d\lambda)}{\lambda-z}.$$

(d') For each i, the discontinuity of the function f across $\gamma_i{}^\infty$ is a measure τ_i supported in $(-\infty,0]\times(-\infty,0]$ and such that there exist C_i for which

$$\left|\int \tau_i(dx)g(x)\right| < C(|g|_1 + |g|_\infty)$$

for every function $g \in C\cap L_1$.

Here

$$\gamma_1{}^\infty \equiv \{z|z_2\in(-\infty,0],\ z_1/z_2\in(-\infty,0]\},$$
$$\gamma_2{}^\infty \equiv \{z|z_1\in(-\infty,0],\ z_1/z_2\in(-\infty,0]\},$$
$$\gamma_3{}^\infty \equiv \{z|z_1\in(-\infty,0],\ z_2\in(-\infty,0]\}.$$

With the notation of Sec. III specialized to the case $M=2$, one has

$$q_{11} = (\lambda_1-z_1)^{-1}, \quad q_{22} = (\lambda_2-z_2)^{-1}, \quad q_{12}=q_{21}=0,$$
$$q_{31} = (\lambda_1-\lambda_2 z_1/z_2)^{-1}, \quad q_{32} = (\lambda_1-\lambda_2 z_2/z_1)^{-1}, \tag{4.3}$$

so that

$$f(z_1,z_2) = \sum_1^3 (4\pi)^{-2}\int_{\gamma(\sigma,i)} f(\lambda)B_i(\lambda,z)d\lambda_1\,d\lambda_2, \tag{4.4}$$

where

$$\gamma(\sigma,i) \equiv \bigcap_{k\neq i} \gamma^\sigma{}_{ak},$$
$$B_1 \equiv (\lambda_1-z_1)^{-1}(\lambda_2 z_1/z_2-\lambda_1)^{-1},$$
$$B_2 \equiv (\lambda_1-z_1)^{-1}(\lambda_1 z_2/z_1-\lambda_2)^{-1}, \tag{4.5}$$
$$B_3 \equiv (\lambda_1-z_1)^{-1}(\lambda_2-z_2)^{-1}.$$

The B_i in (4.5) are not in L_1; therefore we cannot take the limit $\sigma\to\infty$ inside the integral in (4.4). We shall therefore analyze the integrals in (4.4) in more detail before taking the limit. We give details only for B_3.

Notice first that the contour $\gamma(\sigma,3)$ can be deformed to become a subset of the Cartesian product $\Delta^\sigma\times\Delta^\sigma$, where $\Delta^\sigma = \Delta^\sigma{}_> \cup \Delta^\sigma{}_<$ and

$$\Delta^\sigma{}_> \equiv \{z|\ |z|=\sigma, \operatorname{Im} z > 1/|\sigma|\},$$
$$\Delta^\sigma{}_< \equiv \{z|-\sigma<\operatorname{Re} z<0, |\operatorname{Im} z| = 1/\sigma\} \cup \{z|\ |z| = 1/\sigma, \operatorname{Re} z>0\}. \tag{4.6}$$

We shall use the symbol γ for the modified contour, neglecting indices. Let

$$\mathbf{I}(z_1,z_2) \equiv \int\int_\gamma f(\boldsymbol\lambda)\frac{d\lambda_1\,d\lambda_2}{(\lambda_1-z_1)(\lambda_2-z_2)}, \quad \boldsymbol\lambda = \{\lambda_1,\lambda_2\}. \tag{4.7}$$

One has

$$\mathbf{I}(z_1,z_2) - \mathbf{I}(z_1,1) - \mathbf{I}(1,z_2) + \mathbf{I}(1,1) = \int\int_\gamma (z_2-1)(z_1-1)f(\boldsymbol\lambda) \times \frac{d\lambda_1\,d\lambda_2}{(\lambda_1-z_1)(\lambda_2-z_2)(\lambda_1-1)(\lambda_2-1)}. \tag{4.8}$$

Recall now that, when $\sigma\to\infty$, a part of γ covers twice the surface $\{z|\operatorname{Im} z_1 = \operatorname{Im} z_2 = 0, \operatorname{Re} z_1<0, \operatorname{Re} z_2<0\}$ from opposite directions in Ω. It follows from (d') that the contribution from these parts of γ amounts, in the limit $\sigma\to\infty$, to

$$(z_1-1)(z_2-1) \times \int\int \frac{\tau(d\lambda_1\,d\lambda_2)}{(\lambda_1-z_1)(\lambda_2-z_2)(\lambda_1-1)(\lambda_2-1)}. \tag{4.9}$$

We now study the contribution from the remaining part of γ, which we denote by γ^*, $\gamma^* = \bigcup_1^3 \gamma_i$,

$$\gamma_1 \equiv (\Delta_> \times \Delta_<)\cap\gamma, \quad \gamma_2 \equiv (\Delta_< \times\Delta_>)\cap\gamma,$$
$$\gamma_3 \equiv (\Delta_>\times\Delta_>)\cap\gamma. \tag{4.10}$$

We first study

$$\int_{\gamma_1} f(\boldsymbol\lambda)\frac{d\lambda_1\,d\lambda_2}{(\lambda_1-z_1)(\lambda_2-z_2)(\lambda_1-1)(\lambda_2-1)} \equiv \mathbf{J}_1{}^\sigma. \tag{4.11}$$

Since $|f(k\lambda_1,\lambda_2)|$ diverges linearly when $k\to\infty$, we cannot take the limit inside the integral in (4.11). Notice, however, that, uniformly in z_1,z_2 over compact sets, J_1 differs by terms that vanish when $\sigma\to\infty$ from the function

$$\mathbf{J}^{\sigma}{}_1 \equiv \int_{\gamma_1} f(\lambda) \frac{d\lambda_1\, d\lambda_2}{\lambda_2(\lambda_1 - z_1)(\lambda_2 - z_2)(\lambda_1 - 1)}. \quad (4.12)$$

Define $W^{\sigma}(\lambda_1)$ by

$$W^{\sigma}(\lambda_1) \equiv \{\lambda_2 | (\lambda_1, \lambda_2) \in \gamma_1\}. \quad (4.13)$$

Independently of λ_1, the curve $W^{\sigma}(\lambda_1)$ has length smaller than $2\pi\sigma$. Since $|f(\lambda, \sigma, e^{i\theta})| < c\sigma$ uniformly in λ, the sequence of functions $\mathbf{J}_1{}^{\sigma}$ converges, when $\sigma \to 0$, uniformly over compact sets. Therefore the limit function, which we denote by $\mathbf{H}_1(z)$, is analytic in $C - R^-$. Moreover $\mathbf{H}_1(z)/|z|$ is uniformly bounded, since this property has been established for $\mathbf{J}_1(z)$ uniformly in σ. Therefore $\mathbf{H}_1$ can be written

$$\mathbf{H}_1(z) = (z-1) \int \frac{\nu_1(d\lambda)}{\lambda - z} \quad (4.14)$$

for some bounded measure ν_1, supported by $(-\infty, 0]$.

The integral over $\gamma_2{}^{\sigma}$ can be studied similarly and converges, when $\sigma \to \infty$, to

$$\mathbf{H}_2(z) = (z-1) \int \frac{\nu_2(d\lambda)}{\lambda - z} \quad (4.15)$$

for a bounded measure ν_2 supported by $(-\infty, 0]$. Finally the integral over $\gamma_3{}^{\sigma}$ converges to a constant c.

In conclusion, one can find measures $\tau_3(d\lambda_1\, d\lambda_2)$ on $(-\infty, 0] \times (-\infty, 0]$, and $\nu_{31}(d\lambda)$, $\nu_{32}(d\lambda)$ on $(-\infty, 0]$, and a constant c such that

$$\lim_{\sigma \to \infty} \int_{\gamma(\sigma,3)} (z_1 - 1)(z_2 - 1) \times \frac{d\lambda_1\, d\lambda_2}{(\lambda_1 - z_1)(\lambda_2 - z_2)(\lambda_1 - 1)(\lambda_2 - 1)}$$
$$= (z_1 - 1)(z_2 - 1) \left\{ \int\int \frac{\tau_3(d\lambda_1\, d\lambda_2)}{(\lambda_1 - z_1)(\lambda_2 - z_2)} + \int \frac{\nu_{31}(d\lambda)}{\lambda - z_1} + \int \frac{\nu_{32}(d\lambda)}{(\lambda - z)} + c \right\}. \quad (4.16)$$

Similar conclusions are obtained for the two other integrals in (4.6). From (4.4)–(4.8) and (4.16) one concludes that there are functions $\mathbf{G}^{\sigma}{}_i$ analytic in the region bounded by Δ^{σ} such that

$$f(z_1, z_2) + \Psi^{\sigma}(z_1, z_2) = \mathbf{G}_1{}^{\sigma}(z_1) + \mathbf{G}_2{}^{\sigma}(z_2) + \mathbf{G}_3{}^{\sigma}(z_1/z_2), \quad (4.17)$$

where Ψ^{σ} converges, when $\sigma \to \infty$, to a function with a known representation. Therefore also the right-hand side converges and it is straightforward to show that the limit $\mathbf{G}$ has the form

$$\mathbf{G}(z_1, z_2) = \mathbf{G}_1(z_1) + \mathbf{G}_2(z_2) + \mathbf{G}_3(z_1/z_2), \quad (4.18)$$

where the functions $\mathbf{G}_i$ are analytic in $\mathbf{C} \backslash R-$.

Combining this remark with (4.16), one concludes

$$f(z_1, z_2) = \left\{ (z_2 - 1)(z_1 - 1) \int\int \frac{\tau_{12}(d\lambda_1\, d\lambda_2)}{(\lambda_1 - z_1)(\lambda_2 - z_2)} + \text{cyclic} \right\} + \left\{ (z_1 - 1)(z_2 - 1) \left[\int \frac{\nu_{31}(d\lambda)}{\lambda - z_1} + \int \frac{\nu_{32}(d\lambda)}{\lambda - z_2} \right] + \text{cyclic} \right\} + \left\{ (z_1 - 1) \int \frac{\mu_1(d\lambda)}{\lambda - z_1} + \text{cyclic} \right\}, \quad (4.19)$$

where by "cyclic" we have indicated the terms which are obtained by making a cyclic permutation of the indices, and setting $z_3 = z_1/z_2$. The monomials of order 2 in $(z_1 - 1)$, $(z_2 - 1)$, $(z_1/z_2 - 1)$, $(z_2/z_1 - 1)$, which could otherwise be present, are ruled out by the fact that $f(z_1, z_2)$ is regular at $z_1 = 0$ and at $z_2 = 0$, and moreover $|f(\lambda_1 z_1, \lambda_2 z_2)| < c \sup\{\lambda_1, \lambda_2\}$. If one takes (4.1) and (4.2) into account and makes the change in integration variables

$$\lambda \to \sigma \equiv (1 + \lambda)/(1 - \lambda), \quad 0 \leqslant \sigma \leqslant 1,$$

the representation (3.21) follows from (4.19).

It is worth mentioning that the representation (4.19) is not unique. Indeed, since we make no statements on the properties of the measures involved, and in particular whether they give finite weight to sets of lower dimension, the contribution from the measures ν_{ij} can be included in the term with measure τ_{ij}, substituting this measure with $\tau_{ij} + \nu_{ij}(d\sigma_1)\delta(1 - \sigma_2) d\sigma_2$.

Therefore we can assume

$$\nu_{ij} = 0, \quad \text{for all } i, j. \quad (4.20)$$

For the same reason, one can assume

$$\int \frac{\tau_{ij}(d\lambda_1\, d\lambda_2)}{(\lambda_1 - z)(\lambda_2 - z)} = 0, \quad \text{all } z, i, j. \quad (4.21)$$

Indeed, one has

$$(\lambda_1 - z)^{-1}(\lambda_2 - z)^{-1} = ((\lambda_1 - z)^{-1} - (\lambda_2 - z)^{-1})(\lambda_2 - \lambda_1)^{-1}$$

so that (4.21) can be obtained through a redefinition of μ_j, $j = 1,2,3$. We do not know at present whether conditions (4.20) and (4.21) ensure uniqueness of the representation.

V. COMPARISON WITH SOME RESULTS FROM POTENTIAL THEORY

In this section we shall discuss briefly the relation between the analytic representation for ϵ^* derived here and the "functional" representation derived in Ref. 5. We call that representation functional since it expresses ϵ^* in terms of suitable bounded self-adjoint operators on the Hilbert space of square integrable vector fields on D. When the boundaries of the regions D_i are sufficiently regular, these operators can be expressed in terms of integral kernels related to potential theory.

We shall briefly recall the functional representation and then write it in a form which is more suitable for comparison with the analytic representation. We restrict ourselves to the case of a composite made of three isotropic homogeneous components.

Let χ_i be the characteristic function of the domain D_i occupied by the ith component, with dielectric tensor $\epsilon_i I$. Let A be the orthogonal (Hodge) projection from the space H of square integrable vector fields on D onto the space of square integrable vector fields. In many cases, A takes a simple form related to potential theory. For example, if D is a lattice cell and periodic boundary conditions are imposed, then A acts on (periodic) vector fields F as

$$(AF)_i(x) = \frac{\partial}{\partial x_i}(G * \operatorname{div} F)(x),$$

where G is the Green's function for the Laplacian with periodic boundary conditions.

If ϵ_0 is such that $|\epsilon_i/\epsilon_0| < 1$, $i = 1,2,3$, define

$$Q_{\epsilon_0}(x,\epsilon) := \sum_{k=1}^{3}\left(1 - \frac{\epsilon_k}{\epsilon_0}\right)\chi_k(x), \quad \epsilon \equiv \{\epsilon_1,\epsilon_2,\epsilon_3\}.$$

From now on, we shall not indicate explicitly the dependence on ϵ and ϵ_0. It is easy to verify[5] that the representation given below does not depend on ϵ_0; this will also appear from the analysis given here.

Let $\mathbf{B}(x)$ be the electric field solution of Maxwell's equation with the same boundary conditions for a homogeneous isotropic medium. The functional representation derived in Ref. 5 is then, setting $\epsilon' = \epsilon'' - \epsilon^*$, where ϵ'' denotes the average of ϵ,

$$\epsilon' = \epsilon_0\langle B, Q^{1/2}(I - Q^{1/2}AQ^{1/2})^{-1}(Q^{1/2}AQ^{1/2})Q^{1/2}B\rangle, \tag{5.1}$$

and we have used the notation $\langle\cdot,\cdot\rangle$ for the scalar product in H. We shall denote by $\|\cdot\|$ the operator norm in H. Regarding Q as a multiplication operator, one has $\|Q\| < 1$, so that both $(I - Q^{1/2}AQ^{1/2})^{-1}$ and $(I - AQA)^{-1}$ can be expanded in a convergent power series. It is then easy to verify that

$$\epsilon' = \epsilon_0\langle B, QA(I - AQ)^{-1}AQB\rangle. \tag{5.2}$$

The function (5.2) is analytic in the ϵ_i near the point $\epsilon_i = 1$, $i = 1,2,3$. Its power series expansion has coefficients that can be written in terms of the moments of the spectral measure of the operators $A\chi_i A$ in the states $A\chi_j B$ and of the kernels which represent $A\chi_i A$ in the representation in which $A\chi_i A$ is diagonal. If the χ_i are characteristic functions of domains which have a sufficiently regular boundary, all these quantities can be written explicitly in terms of Green's functions. On the other hand, the coefficients of the expansion of (4.19) in powers of the ϵ_i have coefficients that are related to the moments of the measures τ_{ij} and μ_i. There is therefore a relation between the measures in (4.19) and quantities which refer to potential theory. These relations should allow *a priori* estimates on the measures, and should be particularly useful if one wants to determine which class of measures can appear in (4.19) if this formula has to represent the effective dielectric constant of a composite for some geometry. This is also instrumental in proving convergence of the measures τ_{ij} and μ_i under homogenization.

The expression (5.2) lends itself to a resolvent expansion which is particularly useful if two of the dielectric constants of the components differ by a small amount or if two of the components appear only as an almost homogeneous and isotropic subcomposite. Indeed, let f be an arbitrary function of the ϵ_i's. One has

$$A(I - AQA)^{-1}A \equiv \epsilon_0 A[f(I - A\chi_1 A) + \epsilon_1 A\chi_1 A - A\zeta A]^{-1}A, \tag{5.3}$$

where $\zeta \equiv (f - \epsilon_2)\chi_2 + (f - \epsilon_3)\chi_3$. In (5.3) we have chosen to single out the role of the components 2 and 3.

Suppose now that ϵ_2 and ϵ_3 are such that

$$\delta \equiv |\epsilon_2 - \epsilon_3| \ll \min\{|\epsilon_1|, (|\epsilon_2| + \epsilon_3|)/2\}. \tag{5.4}$$

Then one can choose f in such a way that

$$\|A\zeta A\| \ll \inf|(1 - \sigma)f + \epsilon_1\sigma|, \quad 0 < \sigma < 1, \tag{5.5}$$

so that one can write the convergent resolvent expansion

$$A(I - AQA)^{-1}A = \epsilon_0 A \sum [f(I - A\chi_1 A) + \epsilon_1 A\chi_1 A]^{-1} \times \{A\zeta A[f(I - A\chi_1 A) + \epsilon_1 A\chi_1 A]^{-1}\}^k.$$

One can take, e.g., $f(\epsilon) = (\epsilon_2 + \epsilon_3)/2$, but other choices could be more convenient for specific geometries.

The representation (5.2) now becomes

$$\epsilon' = (\epsilon_0)^2\Big\langle B, QA[(I - A_1)f + \epsilon_1 A_1] \times \sum_{k=0}^{1}[\mathfrak{z}((I - A_1)f + \epsilon_1 A_1)]^k AQB\Big\rangle, \tag{5.6}$$

where $A_1 \equiv A\chi_1 A$ and $\mathfrak{z} \equiv A\zeta A$.

Using the fact that $B \in \operatorname{Ker} A$, one can verify that for every bounded operator F one has

$$\langle B, QAFAQB\rangle = [1/(\epsilon_0)^2]\{(\epsilon_1 - \epsilon_2)(\epsilon_1 - \epsilon_3) \times \langle B, \chi_1 AFA\chi_1 B\rangle + \text{cyclic}\}. \tag{5.7}$$

From (5.6) and (5.7) one sees that ϵ' is indeed independent of ϵ_0 and, under assumption (5.5), can be written as

$$\epsilon'(\epsilon) = \frac{1}{2}\sum_{j \neq i \neq k}(\epsilon_i - \epsilon_j)(\epsilon_i - \epsilon_k) \times \langle B, \chi_i[(I - A_1)f + \epsilon_1 A_1]^{-1} \times \sum_{k=0}^{\infty}[\mathfrak{z}((I - A_1)f + \epsilon_1 A_1)^{-1}]^k \chi_i B\rangle. \tag{5.8}$$

The right-hand side of (5.8) can be written as a function of the kernels of the (integral) operators $A_i := A\chi_i A$, $i = 1,2,3$. The term of order zero in δ is, setting $\epsilon_2 = \epsilon_3 = \epsilon$,

$$(\epsilon_1 - \epsilon)^2 \int_0 [\epsilon_1\lambda + \epsilon(1 - \lambda)]^{-1}\mu_1(d\lambda).$$

Comparison with (5.2) under conditions (4.20) and (4.21) shows that μ_1 is the spectral measure of A_1 in the state $\chi_1 B$. To next order one has, setting $\epsilon_3 = \epsilon_2 + \delta$ and choosing $f(\epsilon) = \epsilon_2$,

$$\begin{aligned}
&\delta(\epsilon_1 - \epsilon_2)^2\langle B, \chi_1[\epsilon_2(I - A_1) + \epsilon_1 A_1]^{-1}\chi_3 \\
&\quad \times [\hat{\epsilon}_2(I - A_1) + \epsilon_1 A_1]^{-1}\chi_1 B\rangle \\
&\quad + \delta(\epsilon_1 - \epsilon_2)\{\langle B, \chi_2[\epsilon_2(I - A_1) + \epsilon_1 A_1]^{-1}\chi_2 B\rangle \\
&\quad - \langle B, \chi_3[\epsilon_2(I - A_1) + \epsilon_1 A_1]^{-1}\chi_3 B\rangle\}
\end{aligned}$$

$$= (\epsilon_1 - \epsilon_2)^2\delta \int_0^1 \int_0^1 \{[\epsilon_2(1-\lambda) + \epsilon_1\lambda]$$
$$\times [\epsilon_2(1-\sigma) + \epsilon_1\sigma]\}^{-1}\nu_3(d\lambda\, d\sigma) + \delta(\epsilon_1 - \epsilon_2)$$
$$\times \int [\epsilon_2(1-\lambda) + \epsilon_1\lambda]^{-1}[\alpha_3(d\lambda) - \alpha_2(d\lambda)], \tag{5.9}$$

with $\nu_3(d\lambda\, d\sigma) = \varphi_3(\lambda)\varphi_3(\sigma)\rho_3(d\lambda\, d\sigma)$, where ρ_3 is the spectral kernel of $A\chi_3A$ in the representation in which $A\chi_1A$ is diagonal and φ_3 is the representative of χ_3B in the same representation.

Proceeding as above, one obtains to each order of δ a sum of terms of the form

$$\delta^{n'} \int_0^1 \cdots \int_0^1 \{[\epsilon_2(1-\lambda_1) + \epsilon_1\lambda_1]^{-1} \cdots$$
$$[\epsilon_2(1-\lambda_M) + \epsilon_1\lambda_M]^{-1}\}\nu_3(d\lambda_1 \cdots d\lambda_M),$$

with $M = n'$ or $n' + 1$, where $\nu_3(d\lambda_1\, d\lambda_2 ... d\lambda_M)$ can be described explicitly in terms of the kernel of $A\chi_3A$ and of χ_iB, $i = 1,2,3$, in the representation in which $A\chi_1A$ is diagonal.

On the other hand, from (4.19)–(4.21), the term of first order in δ is

$$\epsilon_1(\epsilon_1 - \epsilon_2)^2 \int\int \frac{\tau_{23}(d\sigma_2\, d\sigma_3)\sigma_3^*}{(\sigma_2\epsilon_3 + \sigma_3\epsilon_2)(\sigma_3\epsilon_1 + \sigma_3^*\epsilon_2)^2}$$
$$+ (\epsilon_2 - \epsilon_1) \int \frac{\tau_{23}(d\sigma_1\, d\sigma_2)}{(\sigma_2\epsilon_1 + \sigma_1^*\epsilon_1)}$$
$$- (\epsilon_2 - \epsilon_1) \int \frac{\tau_{12}(d\sigma_1\, d\sigma_2)}{(\sigma_1\epsilon_2 + \sigma_1^*\epsilon_1)}$$
$$+ 2(\epsilon_2 - \epsilon_1) \int \frac{\mu_1(d\sigma)}{(\sigma\epsilon_2 + \sigma^*\epsilon_1)}. \tag{5.10}$$

Comparison of (5.9) with (5.10) gives the relation of the first moments of the measures τ_{ij} and μ_i with the expectation values of the operators $A\chi_iA$ in the states $A\chi_jB$. Comparison of the higher-order terms in the expansion in powers of δ gives the relation of all moments of the measures τ_{ij} and μ_i in terms of the spectral measures and of the kernels of the operators $A\chi_iA$.

Since we have found so far no systematic way of expressing this relation, we shall refrain from giving here the details.

ACKNOWLEDGMENTS

The work of G.F.D. was supported in part by MPI, while the work of V.N. was carried out on a C.N.R. fellowship.

[1]D. Bergman, "Bounds for the complex dielectric constant of a two component composite material," Phys. Rev. B **23**, 3058 (1981).
[2]K. Golden, Ph.D. thesis, Courant Institute, 1984; K. Golden and G. Papanicolaou, "Bounds for effective parameters of heterogeneous media," Commun. Math. Phys. **90**, 473 (1983).
[3]K. Golden and G. Milton, "The values of holomorphic functions with positive real part in a polydisk," University of Rome preprint, 1986.
[4]G. Milton, "Multicomponent composites and new types of continued fractions I, II," Commun. Math. Phys. **111**, 281, 329 (1988).
[5]G. F. Dell'Antonio, R. Figari, and E. Orlandi, "An approach through orthogonal projections...," Ann. Inst. H. Poincaré **44**, 1 (1986).
[6]B. A. Fuks, *Theory of Several Complex Variables* (Am. Math. Soc., Providence, RI, 1972).

A simple model of asymmetry in Wien dissociation of a weak electrolyte

Douglas K. McIlroy
6 Brandy Court, Calamvale 4116, Australia

(Received 14 January 1987; accepted for publication 17 August 1988)

A simple model of Wien dissociation of a weak electrolyte is proposed in which the relative increase in the dissociation constant, $K(\mathbf{X})/K(\mathbf{0})$, where $\mathbf{X}$ is the uniform applied electric field, is *asymmetric* with respect to the direction of $\mathbf{X}$. The model assumes that one ionic species of the electrolyte is mobile but that the other is *fixed* in a crystal or liquid crystal lattice. The asymmetry of $K(\mathbf{X})/K(\mathbf{0})$ is shown to arise because of the resulting asymmetry in the boundary condition on the distribution function describing pairs of associated ions of the electrolyte at the distance of closest approach of the ions. Solutions for the association rate constant A and $K(\mathbf{X})/K(\mathbf{0})$ of the model electrolyte are obtained by solving the partial differential equation governing the streamfunctions that describe the two ionic states of the electrolyte.

I. INTRODUCTION

In the mathematical theory of Wien dissociation of a weakly dissociated electrolyte[1] it is assumed that in the ionized state the ions of the electrolyte are mobile point charges occupying a space of infinite extent. A consequence of this is that the relative increase in the dissociation constant, $K(\mathbf{X})/K(\mathbf{0})$, due to the application of a uniform electric field $\mathbf{X}$ is *symmetric* with respect to the direction of $\mathbf{X}$. In his analysis of the partial differential equation governing the distribution function describing associated ion pairs of a weak electrolyte which lead to his well-known (symmetrical) result

$$\frac{K(\mathbf{X})}{K(\mathbf{0})}=\frac{I_1[(8\epsilon)^{1/2}]}{(2\epsilon)^{1/2}}, \tag{1.1}$$

where I_1 is the modified Bessel function of the first kind of order 1 and ϵ is the field-dependent parameter,[2] Onsager[3] remarked "a generalization *of this equation* to crystals of lesser symmetry would be desirable but leads to a differential equation that is much more formidable than the equation for a *point-charge electrolyte*." (Italics indicate words inserted by the author.)

In Sec. II we propose a simple model of such an asymmetry in the Wien dissociation of a weak electrolyte for which, when compared to Eq. (1.1), the value of $K(\mathbf{X})/K(\mathbf{0})$ is enhanced in one direction of the applied electric field (the forward-field direction) but is reduced in the opposite direction (the reverse-field direction). In Sec. III this model is analyzed by solving the partial differential equation governing the streamfunctions that describe the two ionic states of the model, viz., the dissociated state and the associated state. Expressions for the association rate constant A and $K(\mathbf{X})/K(\mathbf{0})$ are deduced for the model weak electrolyte and a strong asymmetry is demonstrated to exist between the values of $K(\mathbf{X})/K(\mathbf{0})$ for forward and reverse applied electric fields.

II. THE MODEL

We suppose that the model weak electrolyte is composed of mobile point ions (i ions) of charge e_i and oppositely charged point ions (j ions) of charge e_j that are *fixed* in a crystal or liquid crystal lattice. The orientation of the lattice is itself fixed with respect to the direction of the uniform applied electric field $\mathbf{X}$, e.g., it could be part of the internal structure of a polarized membrane.

Taking the origin O of polar coordinates at the position of a j ion we call the polar coordinates of the i ion of an i,j pair r,θ, where θ is the angle between $\mathbf{X}$ and $\mathbf{r}$. We further suppose that the molecular groups to which the j ions are attached in the lattice, hereafter referred to as the *dielectric bases*, block the free passage of i ions in the region $\pi/2<\theta\leqslant\pi$ at least in the region close to the j ions, when $\mathbf{X}$ is in what we shall call the *forward* applied field direction. The arrangement for such a field is depicted in Fig. 1. With the applied electric field in the opposite direction, which we shall refer to as a *reverse* applied field, the dielectric bases block the free passage of i ions in the region $0\leqslant\theta<\pi/2$ at least in the region close to the j ions. Figure 2 shows the arrangement for the reverse field case.

If the dielectric bases were, e.g., spheres of radius b, previous calculations[4] indicate that as far as the effect that the total electrical potential $\Phi(r,\theta)$ of an i,j pair has on the value of A and $K(\mathbf{X})/K(\mathbf{0})$ is concerned, the effect of these bases is negligible provided $b\ll1$ (in units normalized with respect to the characteristic length $2q$ of the weak electrolyte, where $q=-e_ie_j/2DkT>0$, D is the dielectric constant of the medium, k is Boltzmann's constant, and T is absolute temperature) being $O(e^{-1/b})$. Thus we shall as-

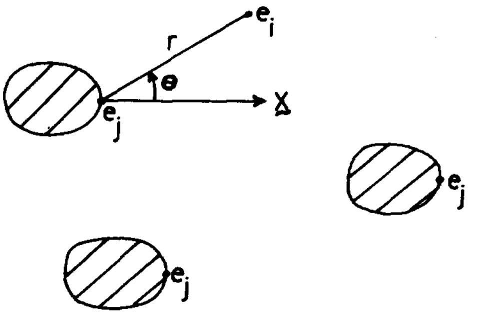


FIG. 1. The forward-field case. Cross-hatched areas represent the dielectric bases.

 0022-2488/88/122695-05$02.50

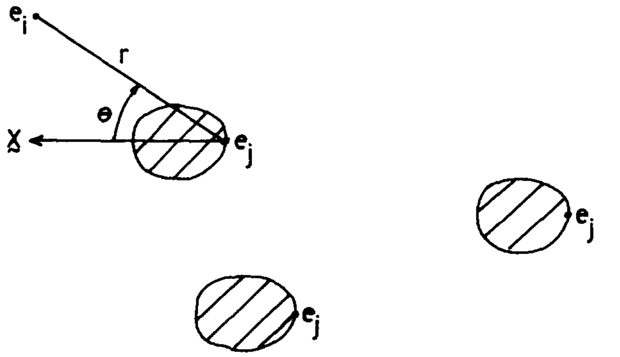


FIG. 2. The reverse-field case.

sume for simplicity that the average radius b of the dielectric bases is such that $b \ll 1$ and we therefore have with very good accuracy in these calculations that

$$\Phi(r,\theta) = -1/r - 2\epsilon r \cos\theta, \tag{2.1}$$

where r and Φ have been nondimensionalized on division by the characteristic length and characteristic potential kT, respectively, and $\epsilon = 2\beta q$, $\beta = e_i|\mathbf{X}|/2kT$. As in the previous analysis[2] we neglect the effect of the ionic atmospheres and to be specific we take $e_i > 0$, $e_j < 0$ so that β and $\epsilon > 0$.

The main effect of the dielectric bases is on the boundary condition at the distance of closest approach of the i and j ions for the distribution function $f^{(2)}(r,\theta)$ describing associated ions. At all points of the surfaces of these bases an equilibrium will exist. Thus on these surfaces

$$f^{(2)}(r,\theta) = e^{-\Phi} - 1, \tag{2.2}$$

since the normalized value of the distribution function describing dissociated ions, $f^{(1)}(r,\theta)$, is unity.[2] The only charge singularity on the surface of the dielectric bases is at the origin so for forward applied fields

$$f^{(2)}(r,\theta) \sim e^{1/r} \quad \text{as } r \to 0, \quad 0 \leqslant \theta \leqslant \pi/2, \tag{2.3}$$

and by comparison

$$f^{(2)}(r,\theta) = 0 \tag{2.4}$$

effectively, at all other points on these surfaces. Equation (2.3) is more conveniently written as

$$\lim_{r\to 0} e^{-1/r} f^{(2)}(r,\theta) = 1, \quad 0 \leqslant \theta \leqslant \pi/2, \tag{2.5}$$

and Eq. (2.4) may be rewritten as

$$f^{(2)}(r,\theta) = 0, \quad \pi/2 < \theta \leqslant \pi, \tag{2.6}$$

on the surfaces of the dielectric bases. For reverse applied electric fields the boundary conditions are

$$\lim_{r\to 0} e^{-1/r} f^{(2)}(r,\theta) = 1, \quad \pi/2 \leqslant \theta \leqslant \pi, \tag{2.7}$$

and Eq. (2.4) becomes

$$f^{(2)}(r,\theta) = 0, \quad 0 \leqslant \theta < \pi/2, \tag{2.8}$$

on the surfaces of the dielectric bases. The condition for the total dissociation of an ion pair

$$\lim_{r\to\infty} f^{(2)}(r,\theta) = 0, \tag{2.9}$$

for all θ, applies to both forward and reverse applied fields and completes the boundary conditions for this model. Clearly the properties of the weak electrolyte are symmetrical about the direction of **X** so that we may further assume that Eqs. (2.5)–(2.9) hold for all azimuthal angles.

III. MATHEMATICAL RESULTS

In analyzing our model electrolyte we have the following cases.

A. Forward applied electric fields

The distribution function for dissociated ions is[2]

$$f^{(1)}(r,\theta) = 1, \tag{3.1}$$

and i ions associate with the central j charge only along trajectories that meet the origin at angles $\theta \leqslant \pi/2$. Using Eq. (3.1) we have[2]

$$-\frac{\partial g^{(1)}}{\partial \theta} + \sin\theta(1 - 2\epsilon r^2 \cos\theta) = 0, \tag{3.2}$$

$$\frac{\partial g^{(1)}}{\partial r} + 2\epsilon r \sin^2\theta = 0, \tag{3.3}$$

where $g^{(1)}(r,\theta)$ is the streamfunction describing dissociated ions. Equations (3.2) and (3.3) have the solution

$$g^{(1)}(r,\theta) = -\cos\theta - \epsilon r^2 \sin^2\theta + \alpha, \tag{3.4}$$

where α is a constant. This solution holds in the *region of association*, which is shown in the cross section in Fig. 3, and on its boundary. Outside this region no association of i ions with j ions occurs because of the presence of the dielectric bases. No association flux crosses the boundary of the region of association, which is formed by the streamlines that meet the origin at $\theta = \pi/2$. These streamlines must correspond to $g^{(1)}(r,\theta) = 0$ and therefore $g^{(1)}(0,\pi/2) = 0$ and Eq. (3.4) then gives $\alpha = 0$. Thus in the region of association and on its boundary,

$$g^{(1)}(r,\theta) = -\cos\theta - \epsilon r^2 \sin^2\theta. \tag{3.5}$$

Next, in Fig. 3, choose any point Q not coincident with O on the boundary of the region of association and any point $P(r,0)$ with $r > 0$ on the ray $\theta = 0$. Referring to Eq. (2.11) of Paper I (Ref. 2), if we choose for S a closed surface of revolution generated by any axial curve OQP, then it follows from

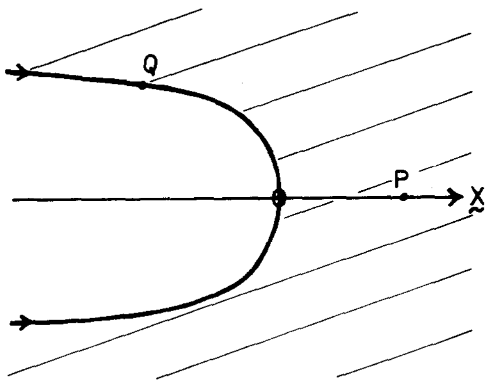


FIG. 3. The cross-hatched area depicts a cross section of the region of association for the forward-field case.

the definition[2] of a streamfunction and Eq. (3.5) that the association rate constant is given by

$$A = -2\pi g^{(1)}(r,0) = 2\pi. \tag{3.6}$$

For associated ions, dissociation of i,j ion pairs occurs along trajectories that emanate from the origin at angles $\theta \leqslant \pi/2$. A cross section of the *region of dissociation* is shown in Fig. 4. In this region and on its boundary

$$f^{(2)}(r,\theta) = \frac{1}{r}\exp\Big(\frac{1}{r} + \epsilon r(\cos\theta - 1)\Big) \times \int_0^1 I_0\Big[(8\epsilon s)^{1/2}\cos\frac{\theta}{2}\Big] e^{-s/r}\,ds \tag{3.7}$$

(where I_0 is the modified Bessel function of the first kind of order zero) because this function satisfies the boundary conditions (2.5) and (2.9) and the partial differential equation obtained by eliminating the streamfunction describing associated ion pairs, $g^{(2)}(r,\theta)$, from Eqs. (3.8) and (3.9) below.[2] This differential equation is also satisfied by the solution $f^{(2)}(r,\theta) = 0$, which satisfies the boundary conditions (2.6) and (2.9) and is therefore the solution outside the region of dissociation. Clearly no dissociation of ion pairs can occur outside the region of dissociation.

From paper I,

$$r^2\sin\theta\frac{\partial f^{(2)}}{\partial r} - \frac{\partial g^{(2)}}{\partial\theta} + \sin\theta(1 - 2\epsilon r^2\cos\theta)f^{(2)} = 0, \tag{3.8}$$

$$\sin\theta\frac{\partial f^{(2)}}{\partial\theta} + \frac{\partial g^{(2)}}{\partial r} + 2\epsilon r\sin^2\theta f^{(2)} = 0, \tag{3.9}$$

and from Eq. (3.7)

$$\frac{\partial f^{(2)}}{\partial r} = \Big(-\frac{1}{r} - \frac{1}{r^2} + \epsilon(\cos\theta - 1)\Big)f^{(2)} + \frac{G(r,\theta)}{r^2F(r,\theta)}f^{(2)}, \tag{3.10}$$

where

$$G(r,\theta) = \int_0^1 sI_0\Big[(8\epsilon s)^{1/2}\cos\frac{\theta}{2}\Big] e^{-s/r}\,ds \tag{3.11}$$

and

$$F(r,\theta) = \int_0^1 I_0\Big[(8\epsilon s)^{1/2}\cos\frac{\theta}{2}\Big] e^{-s/r}\,ds. \tag{3.12}$$

Also

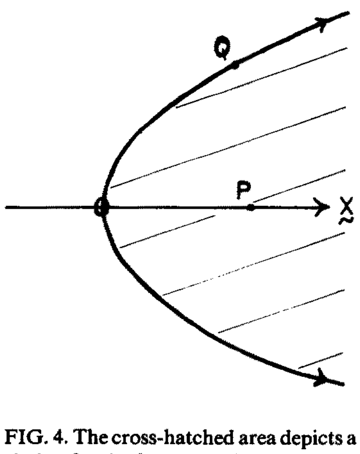


FIG. 4. The cross-hatched area depicts a cross section of the region of dissociation for the forward-field case.

$$\frac{\partial f^{(2)}}{\partial\theta} = \Big(-\epsilon r\sin\theta + \frac{1}{2}\sin\frac{\theta}{2}\frac{H(r,\theta)}{F(r,\theta)}\Big)f^{(2)}, \tag{3.13}$$

where

$$H(r,\theta) = -(8\epsilon)^{1/2}\int_0^1 s^{1/2}I_1\Big[(8\epsilon s)^{1/2}\cos\frac{\theta}{2}\Big] e^{-s/r}\,ds. \tag{3.14}$$

Thus from Eqs. (3.9) and (3.13)

$$\frac{\partial g^{(2)}}{\partial r} = -\frac{1}{2r}\sin\frac{\theta}{2}\sin\theta\exp\Big(\frac{1}{r} + \epsilon r(\cos\theta - 1)\Big) \times\Big(H(r,\theta) + 4\epsilon r\cos\frac{\theta}{2}F(r,\theta)\Big) \tag{3.15}$$

and on integrating by parts we can show from Eq. (3.14) that

$$H(r,\theta) + 4\epsilon r\cos(\theta/2)F(r,\theta) = (8\epsilon)^{1/2}re^{-1/r}I_1[(8\epsilon)^{1/2}\cos\theta/2] \tag{3.16}$$

and using this equation, (3.15) becomes

$$\frac{\partial g^{(2)}}{\partial r} = -\frac{1}{2}(8\epsilon)^{1/2}\sin\frac{\theta}{2}\sin\theta \times\exp(\epsilon r(\cos\theta - 1))I_1\Big[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\Big]. \tag{3.17}$$

From Eqs. (3.8) and (3.10) it follows that

$$\frac{\partial g^{(2)}}{\partial\theta} = \frac{1}{r}\sin\theta\exp\Big(\frac{1}{r} + \epsilon r(\cos\theta - 1)\Big) \times\{G(r,\theta) - (r + \epsilon r^2(1 + \cos\theta))F(r,\theta)\}. \tag{3.18}$$

From Eq. (3.11), on integrating by parts,

$$G(r,\theta) = -re^{-1/r}I_0[(8\epsilon)^{1/2}\cos(\theta/2)] + rF(r,\theta) - (r/2)\cos(\theta/2)H(r,\theta), \tag{3.19}$$

and, using Eq. (3.16), this becomes

$$G(r,\theta) - (r + \epsilon r^2(\cos\theta + 1))F(r,\theta) = -re^{-1/r}I_0[(8\epsilon)^{1/2}\cos(\theta/2)] - [(8\epsilon)^{1/2}/2]r^2\cos(\theta/2) \times e^{-1/r}I_1[(8\epsilon)^{1/2}\cos(\theta/2)]. \tag{3.20}$$

Finally, using Eq. (3.20), Eq. (3.18) becomes

$$\frac{\partial g^{(2)}}{\partial\theta} = -\sin\theta\exp(\epsilon r(\cos\theta - 1))\Big\{I_0\Big[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\Big] + \frac{(8\epsilon)^{1/2}}{2}r\cos\frac{\theta}{2}I_1\Big[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\Big]\Big\}. \tag{3.21}$$

On integrating Eq. (3.17) with respect to r we obtain

$$g^{(2)}(r,\theta) = \Big(\frac{2}{\epsilon}\Big)^{1/2}\cos\frac{\theta}{2}I_1\Big[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\Big] \times\exp(\epsilon r(\cos\theta - 1)) + C(\theta), \tag{3.22}$$

where $C(\theta)$ is a function of θ only, and on integrating Eq. (3.21) with respect to θ and using the identity $I_0(u) = I_1'(u) + (1/u)I_1(u)$, we obtain

$$g^{(2)}(r,\theta) = \Big(\frac{2}{\epsilon}\Big)^{1/2}\cos\frac{\theta}{2}I_1\Big[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\Big] \times\exp(\epsilon r(\cos\theta - 1)) + E(r), \tag{3.23}$$

where $E(r)$ is a function of r only. On comparing Eq. (3.22) and (3.23) we must have $C(\theta) = E(r) = \gamma$, where γ is a constant. Thus

$$g^{(2)}(r,\theta) = \left(\frac{2}{\epsilon}\right)^{1/2} \cos\frac{\theta}{2} I_1\left[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\right] \times \exp(\epsilon r(\cos\theta - 1)) + \gamma. \quad (3.24)$$

No dissociation flux crosses the boundary of the region of dissociation formed by the streamlines that emanate from the origin at $\theta = \pi/2$. These streamlines must correspond to $g^{(2)}(r,\theta) = 0$ and therefore $g^{(2)}(0,\pi/2) = 0$ and Eq. (3.24) gives

$$\gamma = -(1/\epsilon^{1/2}) I_1[(4\epsilon)^{1/2}] \quad (3.25)$$

and so in the region of dissociation and on its boundary

$$g^{(2)}(r,\theta) = \left(\frac{2}{\epsilon}\right)^{1/2} \cos\frac{\theta}{2} I_1\left[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\right] \times \exp(\epsilon r(\cos\theta - 1)) - \frac{1}{\epsilon^{1/2}} I_1[(4\epsilon)^{1/2}]. \quad (3.26)$$

If, in Fig. 4, we select any point Q not coincident with O on the boundary of the region of dissociation and any point $P(r,0)$ with $r>0$ on the ray $\theta = 0$, then, referring to Eq. (2.13) of Paper I, if we choose for S a closed surface of revolution generated by any axial curve OQP, it follows from the definition[2] of a streamfunction and Eqs. (3.6) and (3.26) that

$$\frac{K(\mathbf{X})}{K(\mathbf{0})} = \frac{2\pi}{A} g^{(2)}(r,0) = \left(\frac{2}{\epsilon}\right)^{1/2} I_1[(8\epsilon)^{1/2}] - \frac{1}{\epsilon^{1/2}} I_1[(4\epsilon)^{1/2}]. \quad (3.27)$$

B. Reverse applied electric fields

For dissociated ions, Eq. (3.1) is again valid and i ions associate with the central j ions only along trajectories that meet the origin at angles $\theta \geqslant \pi/2$. Thus the solution for the streamfunction describing dissociated ions in the region of association and on its boundary is again given by Eq. (3.4). The region of association is shown in cross section in Fig. 5, the boundary being formed by the streamlines that meet the origin at $\theta = \pi/2$. Thus on the boundary $g^{(1)}(r,\theta)$ is constant such that

$$g^{(1)}(r,\theta) = g^{(1)}(0,\pi/2). \quad (3.28)$$

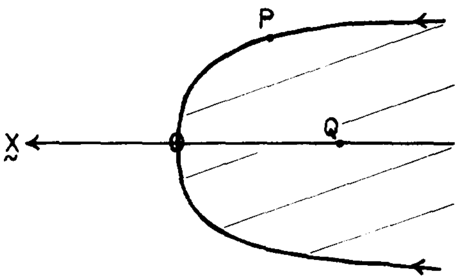


FIG. 5. The cross-hatched area depicts a cross section of the region of association for the reverse-field case.

Also, by symmetry the ray $\theta = \pi$ is a streamline and

$$g^{(1)}(r,\pi) = g^{(1)}(0,\pi) = 0. \quad (3.29)$$

Therefore, from Eq. (3.4), $\alpha = -1$; thus in the region of association and on its boundary

$$g^{(1)}(r,\theta) = -\cos\theta - \epsilon r^2 \sin^2\theta - 1. \quad (3.30)$$

Next, in Fig. 5, choose any point Q not coincident with O on the ray $\theta = \pi$ and any point $P(r,\theta)$ with $r>0$ on the boundary of the region of association. Referring to Eq. (2.11) of Paper I, if we select for S a closed surface of revolution generated by any axial curve OPQ then it follows from the definition[2] of a streamfunction and Eqs. (3.28) and (3.30) that

$$A = -2\pi g^{(1)}(0,\pi/2) = 2\pi. \quad (3.31)$$

For associated ions dissociation of i,j pairs occurs along trajectories that emanate from the origin at angles $\theta \geqslant \pi/2$. The region of dissociation is shown in cross section in Fig. 6. In this region and on its boundary Eq. (3.7) is again valid because it satisfies both boundary conditions (2.7) and (2.9) and the governing differential equation for $f^{(2)}(r,\theta)$. Outside this region the solution $f^{(2)}(r,\theta) = 0$ also again holds and no dissociation of ion pairs can, of course, occur outside the region of dissociation. Clearly, the solution given by Eq. (3.24) is again valid in the region of dissociation and on its boundary. Since the latter is formed by the streamlines that emanate from the origin at $\theta = \pi/2$, it follows that on the boundary $g^{(2)}(r,\theta)$ is a constant such that

$$g^{(2)}(r,\theta) = g^{(2)}(0,\pi/2). \quad (3.32)$$

Also by symmetry the ray $\theta = \pi$ is a streamline and so

$$g^{(2)}(r,\pi) = g^{(2)}(0,\pi) = 0. \quad (3.33)$$

Then Eq. (3.24) gives $\gamma = 0$, and so in the region of dissociation and on its boundary

$$g^{(2)}(r,\theta) = \left(\frac{2}{\epsilon}\right)^{1/2} \cos\frac{\theta}{2} I_1\left[(8\epsilon)^{1/2}\cos\frac{\theta}{2}\right] \times \exp(\epsilon r(\cos\theta - 1)). \quad (3.34)$$

If in Fig. 6 we choose any point Q not coincident with O on the ray $\theta = \pi$ and any point $P(r,\theta)$ on the boundary of the

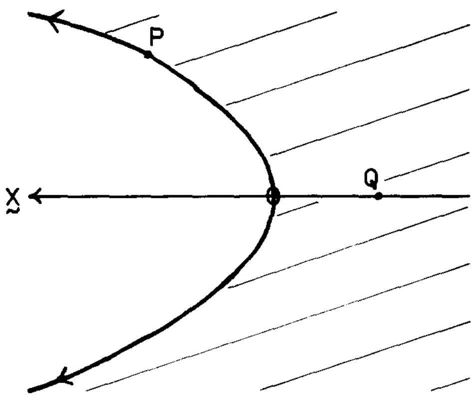


FIG. 6. The cross-hatched area depicts a cross section of the region of dissociation for the reverse-field case.

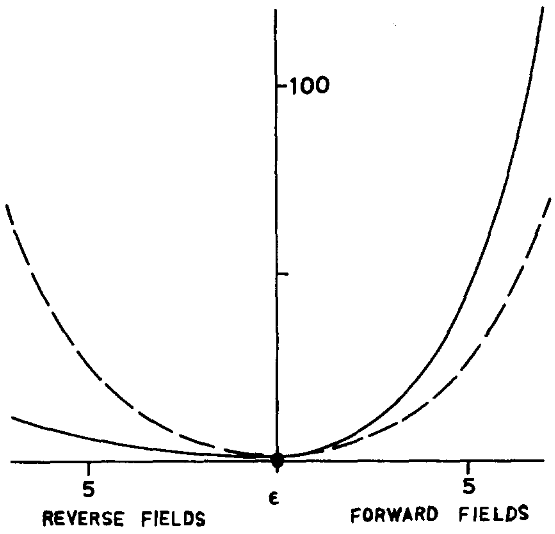


FIG. 7. — represents $K(\mathbf{X})K(\mathbf{0})$ for the asymmetric Wien effect [Eqs. (3.27) and (3.35)]. – – – represents Onsager's $K(\mathbf{X})/K(\mathbf{0})$ [Eq. (1.1)].

region of dissociation, then referring to Eq. (2.13) of Paper I, if we select for S a closed surface of revolution generated by any axial curve OQP, it follows from the definition[2] of a streamfunction and Eqs. (3.31), (3.32), and (3.34) that

$$\frac{K(\mathbf{X})}{K(\mathbf{0})} = \frac{2\pi}{A} g^{(2)}\left(0,\frac{\pi}{2}\right) = \frac{1}{\epsilon^{1/2}} I_1[(4\epsilon)^{1/2}]. \quad (3.35)$$

In Fig. 7 we show the graph of $K(\mathbf{X})/K(\mathbf{0})$ vs ϵ drawn from Eqs. (3.27) and (3.35). The asymmetry in $K(\mathbf{X})/K(\mathbf{0})$ is seen to be very marked: when compard to Eq. (1.1), $K(\mathbf{X})/K(\mathbf{0})$ is considerably enhanced for forward applied fields and is much reduced for reverse applied fields.

A possible application of this model lies in the control mechanism for membrane permeability in nerve.[5] However, the nerve membrane is a thin membrane of thickness $\delta \simeq 1$ whereas we have tacitly assumed that $\delta \gg 1$ in the foregoing analysis. If we denote the streamfunction describing associated ions in the ordinary (symmetric) Wien effect by $\bar{g}^{(2)}(r,\theta)$ [see, e.g., Eq. (5.62) of Paper I], then we may write, from Eqs. (3.27) and (3.35),

$$K(\mathbf{X})/K(\mathbf{0}) = \bar{g}^{(2)}(0,0) - \bar{g}^{(2)}(0,\pi/2), \quad (3.36)$$

for forward applied fields and

$$K(\mathbf{X})/K(\mathbf{0}) = \bar{g}^{(2)}(0,\pi/2) \quad (3.37)$$

for reverse applied fields. In applying our model of the asymmetric Wien effect to the nervous membrane, Eqs. (3.36) and (3.37) apply rather than Eqs. (3.27) and (3.35), and to complete the solution to this problem it would therefore be necessary to compute $\bar{g}^{(2)}(0,0)$ and $g^{(2)}(0,\pi/2)$ for the case $\delta \simeq 1$. The latter problem has not yet been solved.

[1]L. Onsager, J. Chem. Phys. **2**, 599 (1934).
[2]D. K. McIlroy and D. P. Mason, J. Math. Phys. **28**, 2999 (1987).
[3]L. Onsager, Ph.D. thesis, Yale University, 1935.
[4]D. K. McIlroy and D. P. Mason, J. Math. Phys. **25**, 2791 (1984).
[5]D. K. McIlroy, J. Math. Biosci. 7, 313 (1970).