Finite-dimensional representations of the Lie superalgebra gl(2/2) in a gl(2)⊕gl(2) basis. I. Typical representations

Alfred H. Kamupingene and Nguyen Anh Ky
Department of Physics, University of Sofia, 1126 Sofia, Bulgaria

Tchavdar D. Palev
Institute for Nuclear Research and Nuclear Energy, 1184 Sofia, Bulgaria

(Received 25 May 1988; accepted for publication 28 September 1988)

In a series of two papers all finite-dimensional irreducible representations and some indecomposible representations of the general linear Lie superalgebra gl(2/2) are constructed in a basis suitable for the decomposition gl(2/2) ⊃ gl(2) ⊕ gl(2). In this paper each induced gl(2/2) module W is represented as a direct sum of its irreducible gl(2) ⊕ gl(2) submodules V_i, $1 \leqslant i \leqslant 16$. The basis Γ in W is chosen to consist of the union of all Γ_i, where Γ_i is an appropriate basis in each V_i. Expressions for the transformation of Γ under the action of the generators are written down for all induced and hence, also, for all typical gl(2/2) modules.

I. INTRODUCTION

In this paper and the one that follows[1] we study all finite-dimensional irreducible representations of the general linear Lie superalgebra (LS) gl(2/2). The latter consists of all squared four-dimensional matrices. As a basis in gl(2/2) we choose all Weyl matrices e_{ij}, $i,j = 1,2,3,4$, where $e_{ij} \in \mathrm{gl}(2/2)$ is a matrix with 1 on the ith row and the jth column and 0 elsewhere. Assign to each index i a degree (i), which is 0 for $i = 1,2$ and 1 for $i = 3,4$. The general e_{ij} is even (resp. odd) if $(i) + (j)$ is an even (resp. an odd) number. The multiplication (= the supercommutator) [,] on gl(2/2) is given with the linear extension of the relations

$$[\![e_{ij},e_{kl}]\!] = \delta_{jk} e_{il} - (-1)^{[(i)+(j)][(k)+(l)]} \delta_{il} e_{kj}\,. \tag{1.1}$$

The even subalgebra

$$\mathrm{gl}(2/2)_0 = \text{lin. env.}\{e_{ij} \mid i,j = 1,2 \text{ and } i,j = 3,4\} \tag{1.2}$$

is isomorphic to gl(2) ⊕ gl(2). For definiteness we set

$$\text{left gl}(2) \equiv \mathrm{gl}(2)_l = \text{lin. env.}\{e_{ij} \mid i,j = 1,2\}\,, \tag{1.3}$$

$$\text{right gl}(2) \equiv \mathrm{gl}(2)_r = \text{lin. env.}\{e_{ij} \mid i,j = 3,4\}\,. \tag{1.4}$$

Thus we have

$$\mathrm{gl}(2/2)_0 = \mathrm{gl}(2)_l \oplus \mathrm{gl}(2)_r\,. \tag{1.5}$$

The 16-dimensional algebra gl(2/2) is not simple. It contains as an ideal the 15-dimensional special linear LS sl(2/2),

$$\mathrm{sl}(2/2) = \left\{ a \,\middle|\, a \in \mathrm{gl}(2/2),\ \mathrm{str}(a) \equiv \sum_{i=1}^{n} (-1)^{(i)} a_{ii} = 0 \right\}. \tag{1.6}$$

The center Z of sl(2/2) is spanned on the unit matrix. The factor algebra

$$A(1/1) = \mathrm{sl}(2/2)/Z \tag{1.7}$$

is one of the basic Lie superalgebras (LS's) in the terminology and in the notation of Kac,[2] i.e., sl(2/2) is a central extension of $A(1/1)$. Introduce a new basis in the Cartan subalgebra H of gl(2/2), namely,

$$e_1 = e_{11} - e_{22}, \quad e_2 = e_{33} - e_{44}, \qquad e_3 = e_{11} + e_{22} + e_{33} + e_{44}, \quad e_4 = e_{11} + e_{22} - e_{33} - e_{44}\,. \tag{1.8}$$

Then

$$\mathrm{sl}(2/2) = \text{lin. env.}\{e_1,e_2,e_3,e_{ij} \mid i \neq j = 1,2,3,4\}\,, \tag{1.9}$$

$$A(1/1) = \text{lin. env.}\{e_1,e_2,e_{ij} \mid i \neq j = 1,2,3,4\}\,. \tag{1.10}$$

Considering (1.10), one has to take into account that, whenever e_3 appears in the supercommutation relations, it has to be replaced by zero.

Apart from the algebras $B(0,n)$ (Ref. 3) all other basic LS's have indecomposible finite-dimensional modules. So far it is not known how to construct all such modules. Contrary to this, the finite-dimensional irreducible modules (fidirmods) over any basic LS are fully classified.[4]

The structure of the fidirmods and of other indecomposible modules over the basic LS's has been an object of investigation of several authors (see, for instance, Refs. 5–10). In particular, a Young tableaux technique has been generalized to certain LS's[7]; some first steps towards a generalization of the concept of the Gel'fand–Zetlin basis have also been done.[8,9] Irrespective of the progress, there is still much to be done in order to complete the representation theory of the basic LS's. For physical applications, for instance, it is important to know the matrix elements of the generators within at least one basis of each fidirmod. This problem has been solved so far only for gl(n/1) (Ref. 8) and for some low rank LS's (see some of the papers in Ref. 6) or for certain representations of other LS's.

In the present paper and in Ref. 1 we study all fidirmods of the Lie superalgebra gl(2/2). They lead also to all finite-dimensional irreducible representations of its subalgebra sl(2/2). In our case the generator e_4 [see (1.8)], which does not belong to sl(2/2), is diagonal within each gl(2/2) fidirmod. Therefore, the modules are simultaneously irreducible or simultaneously indecomposable with respect to both gl(2/2) and its subalgebra sl(2/2).

The results of the present investigation overlap to a certain extent with those obtained in Ref. 9 (case $n = 2$) and in Ref. 10 (when $n = m = 2$). There are, however, two essential differences. First, from Refs. 9 and 10 one can derive results only for a class of fidirmods of gl(2/2), whereas here we study all fidirmods. Second, the basis here is different from the one used in Refs. 9 and 10. In the present paper each

 0022-2488/89/030553-18$02.50

induced gl(2/2) module W is represented as a direct sum of its irreducible $gl(2) \oplus gl(2)$ submodules V_i, $i = 1,2,...,6$. The basis Γ in W is such that each basis vector belongs to one of the submodules V_i. In Refs. 9 and 10 the decomposition of the gl(2/2) module W is carried along the chain gl(2/2) $\supset$ gl(2/1). Each basis vector is a vector from one gl(2/1) irreducible submodule of W.

II. INDUCED REPRESENTATIONS OF gl(2/2)

A. Some abbreviations and notation

We list first some of the abbreviations and the notation that will be used throughout the paper:

LS, LS's—Lie superalgebra, Lie superalgebras,

fidirmod(s)—finite-dimensional irreducible module(s),

GZ basis—Gel'fand–Zetlin basis,

lin. env.(X)—the linear envelope of X,

$[\![\, ,]\!]$—product (supercommutator) in the LS,

$[x, y] = xy - yx$,

$V_l \otimes V_r$—tensor product of the linear spaces V_l and V_r or a tensor product between the $gl(2)_l$ module V_l and the $gl(2)_r$ module V_r,

$V_1 \odot V_2$—tensor product between the $gl(2)_l \oplus gl(2)_r$ modules V_1 and V_2,

$[m] = [m_{13},m_{23},m_{33},m_{43}]$,

$[m]_l = [m_{13},m_{23}]$, $[m]_r = [m_{33},m_{43}]$,

$l_{ij} = m_{ij} - i$, for $i = 1,2$ and $l_{ij} = m_{ij} - i + 2$, for $i = 3,4$,

U—the universal enveloping algebra of gl(2/2),

$W([m])$—an induced gl(2/2) module,

$\mathbb{C}$—the complex numbers,

$\mathbb{Z}$—all integers,

$\mathbb{Z}_+$—all non-negative integers.

B. Fidirmods of $gl(2)_l \oplus gl(2)_r$

The fidirmods of gl(2/2) will be induced from the fidirmods of the even subalgebra. Therefore, the finite-dimensional irreducible representations of $gl(2)_l \oplus gl(2)_r$ will be essential for the exposition. Here we briefly recall some of their main properties that will be used often. Each $gl(2)_l \oplus gl(2)_r$ fidirmod shall be realized in a space, which is a tensor product of a $gl(2)_l$ fidirmod V_l and a $gl(2)_r$ fidirmod V_r, namely, $V_l \otimes V_r$. If $g_l \oplus g_r \in gl(2)_l \oplus gl(2)_r$ and $x_l \otimes x_r \in V_l \otimes V_r$, then $V_l \otimes V_r$ is turned into a $gl(2)_l \oplus gl(2)_r$ module in a natural way,

$$(g_l \oplus g_r)(x_l \otimes x_r) = g_l x_l \otimes x_r + x_l \otimes g_r x_r . \qquad (2.1)$$

Hence the transformation of the $gl(2)_l \oplus gl(2)_r$ fidirmod $V_l \otimes V_r$ is completely defined from the transformations of the corresponding gl(2) fidirmods. Throughout the paper we use the Gel'fand and Zetlin notation for the fidirmods (and also for the corresponding representations[11]) of gl(2). Every gl(2) fidirmod is labeled by two numbers m_{12} and m_{22} that are complex in general ($\mathbb{Z}_+$ = all non-negative integers), which satisfy the lexical condition

$$m_{12} - m_{22} \in \mathbb{Z}_+ . \qquad (2.2)$$

As a basis in the gl(2) fidirmod $V([m_{12},m_{22}])$, corresponding to (2.2), we choose the Gel'fand–Zetlin basis[11] (GZ basis)

$$\begin{bmatrix} m_{12} & & m_{22} \\ & m_{11} & \end{bmatrix} . \qquad (2.3)$$

The number m_{11} labels the basis vectors in $V([m_{12},m_{22}])$ and takes all possible values, consistent with the "betweenness" or lexical condition

$$m_{12} - m_{11} \in \mathbb{Z}_+ , \quad m_{11} - m_{22} \in \mathbb{Z}_+ . \qquad (2.4)$$

It is convenient to introduce modified notation for $gl(2)_l$ and $gl(2)_r$, namely,

for $gl(2)_l$,

$$[m]_l = [m_{13},m_{23}] , \quad m_{13} - m_{11} \in \mathbb{Z}_+ , \quad m_{11} - m_{23} \in \mathbb{Z}_+ ;$$

for $gl(2)_r$,

$$[m]_r = [m_{33},m_{43}] , \quad m_{33} - m_{31} \in \mathbb{Z}_+ , \quad m_{31} - m_{43} \in \mathbb{Z}_+ . \qquad (2.5)$$

Denote the corresponding to $gl(2)_l$ and $gl(2)_r$ fidirmods as

$$V([m]_l) \equiv V([m_{13},m_{23}])$$

and

$$V([m]_r) \equiv V([m_{33},m_{43}]) \qquad (2.6)$$

with GZ basis vectors

$$\begin{bmatrix} m_{13} & & m_{23} \\ & m_{11} & \end{bmatrix} \equiv \begin{bmatrix} [m]_l \\ m_{11} \end{bmatrix} \equiv (m)_l$$

and

$$\begin{bmatrix} m_{33} & & m_{43} \\ & m_{31} & \end{bmatrix} \equiv \begin{bmatrix} [m]_r \\ m_{31} \end{bmatrix} \equiv (m)_r , \qquad (2.7)$$

respectively. As a basis in the $gl(2)_l \oplus gl(2)_r$ fidirmod

$$V([m]_l) \otimes V([m]_r) \equiv V_0([m_{13},m_{23},m_{33},m_{43}]) , \qquad (2.8)$$

we take the tensor product basis

$$\begin{bmatrix} m_{13} & & m_{23} \\ & m_{11} & \end{bmatrix} \otimes \begin{bmatrix} m_{33} & & m_{43} \\ & m_{31} & \end{bmatrix} \equiv \begin{bmatrix} [m]_l \\ m_{11} \end{bmatrix} \otimes \begin{bmatrix} [m]_r \\ m_{31} \end{bmatrix} \equiv (m)_l \otimes (m)_r \equiv (m) , \qquad (2.9)$$

and refer to it as a canonical $gl(2)_l \oplus gl(2)_r$ basis in $V_0([m_{13},m_{23},m_{33},m_{43}])$. Every $gl(2)_l \oplus gl(2)_r$ fidirmod $V_0([m_{13},m_{23},m_{33},m_{43}])$ can be represented as a tensor product space $V([m]_l) \otimes V([m]_r)$ of a $gl(2)_l$ fidirmod $V([m]_l)$ and a $gl(2)_r$ fidirmod $V([m]_r)$. We refer to $V([m]_l)$ and $V([m]_r)$ as left and right gl(2) fidirmods [or $gl(2)_l$ and $gl(2)_r$ fidirmods] of $V_0([m_{13},m_{23},m_{33},m_{43}])$, respectively. The action of $gl(2)_l \oplus gl(2)_r$ on $V_0([m_{13},m_{23},m_{33},m_{43}])$ is determined from (2.1) and the transformation of the basis within the left and the right gl(2) modules of $V_0([m_{13},m_{23},m_{33},m_{43}])$. For $V([m]_l)$, for instance, one has[11]

$$e_{11}\begin{bmatrix} m_{13} & m_{23} \\ & m_{11} & \end{bmatrix} = (l_{11}+1)\begin{bmatrix} m_{13} & m_{23} \\ & m_{11} & \end{bmatrix}, \tag{2.10}$$

$$e_{22}\begin{bmatrix} m_{13} & m_{23} \\ & m_{11} & \end{bmatrix} = (l_{13}+l_{23}-l_{11}+2)\begin{bmatrix} m_{13} & m_{23} \\ & m_{11} & \end{bmatrix}, \tag{2.11}$$

$$e_{12}\begin{bmatrix} m_{13} & m_{23} \\ & m_{11} & \end{bmatrix} = |(l_{13}-l_{11})(l_{23}-l_{11})|^{1/2}\begin{bmatrix} m_{13} & m_{23} \\ & m_{11}+1 & \end{bmatrix}, \tag{2.12}$$

$$e_{21}\begin{bmatrix} m_{13} & m_{23} \\ & m_{11} & \end{bmatrix} = |(l_{13}-l_{11}+1)(l_{23}-l_{11}+1)|^{1/2}\begin{bmatrix} m_{13} & m_{23} \\ & m_{11}-1 & \end{bmatrix}, \tag{2.13}$$

where here and throughout the paper

$$l_{ij} = m_{ij} - i\,, \text{ for } i=1,2,\ l_{ij} = m_{ij} - i + 2, \text{ for } i = 3,4. \tag{2.14}$$

The transformations of $V([m]_r)$ under the action of $gl(2)_r$ are similar and can be easily derived from (2.10)–(2.13).

C. Induced representations

We now proceed to introduce, following Kac,[4] the gl(2/2) modules, induced from the $gl(2)_l \oplus gl(2)_r$ fidirmods (2.8). To this end we first introduce a new notation

$$[m] = [m_{13},m_{23},m_{33},m_{43}] \tag{2.15}$$

and write for the $gl(2)_l \oplus gl(2)_r$ fidirmods (2.8) also

$$V([m]_l) \otimes V([m]_r) \equiv V_0([m_{13},m_{23},m_{33},m_{43}]) \equiv V_0([m])\,. \tag{2.16}$$

Choose as an ordered basis in the Cartan subalgebra H of gl(2/2) the generators $e_{11},e_{22},e_{33},e_{44}$ and let

$$e^1,e^2,e^3,e^4, \quad e^i(e_{jj}) = \delta^i_j \tag{2.17}$$

be its dual basis from the space H^* of all linear functionals over H. Then $m_{13},m_{23},m_{33},m_{43}$ are the coordinates of the $gl(2)_l \oplus gl(2)_r$ highest weight $\Lambda \in H^*$ of $V_0([m])$ in the dual basis,

$$\Lambda = m_{13}e^1 + m_{23}e^2 + m_{33}e^3 + m_{43}e^4\,. \tag{2.18}$$

Denote by P_+ the linear span of all odd positive root vectors,

$$P_+ = \text{lin. env.}\{e_{13},e_{14},e_{23},e_{24}\} \tag{2.19}$$

and let P be the subalgebra of gl(2/2), which is a direct space of $gl(2)_l \oplus gl(2)_r$ and P_+. Extend the $gl(2)_l \oplus gl(2)_r$ module $V_0([m])$ to a P module, setting

$$P_+ V_0([m]) = 0\,. \tag{2.20}$$

The gl(2/2) module $W([m])$, induced from the $gl(2)_l \oplus gl(2)_r$ module $V_0([m])$, is defined to be the factor space

$$W([m]) = (U \otimes V_0([m]))/I([m])\,, \tag{2.21}$$

which is the tensor product of the gl(2/2) universal enveloping algebra U with $V_0([m])$ and subsequently factorized by the subspace

$$I([m]) = \text{lin. env.}\{up\otimes v - u\otimes pv \,|\, u\in U,\ p\in P\subset U,\ v\in V_0([m])\}\,. \tag{2.22}$$

The linear space $W([m])$ acquires a structure of a gl(2/2) module by virtue of

$$g(u\otimes v) = gu\otimes v\,, \quad g\in gl(2/2)\,, \quad u\otimes v\in W([m])\,, \quad gu\otimes v\in W([m])\,. \tag{2.23}$$

From the Poincaré–Birkhoff–Witt theorem[4] one concludes that U is a linear span of all elements

$$g = (e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}p\,, \quad \theta_1,\ldots,\theta_4 = 0,1\,, \tag{2.24}$$

where $p\in U$ is an arbitrary polynomial of the generators of P. Considering $g\otimes v$ as an element of $W([m])$, from (2.22) we have

$$g\otimes v = (e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}p\otimes v = (e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}\otimes w\,, \quad w = pv\in V_0([m])\,. \tag{2.25}$$

Therefore,

$$W([m]) = \text{lin. env.}\{(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}\otimes v \,|\, v\in V_0([m])\,,\ \theta_1,\ldots,\theta_4 = 0,1\}\,. \tag{2.26}$$

Let T be the subalgebra spanned on all polynomials (in the sense of U) of the odd negative root vectors, i.e.,

$$T = \text{lin. env.}\{(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4} \times |\theta_1,\ldots,\theta_4 = 0,1\}\subset U\,. \tag{2.27}$$

By virtue of (2.26) and (2.27) we see that the gl(2/2) module $W([m])$, considered as a linear space, is a tensor product of T and $V_0([m])$,

$$W([m]) = T\otimes V_0([m])\,. \tag{2.28}$$

As a first basis in $W([m])$ we choose the vectors [see (2.9)]

$$\begin{aligned}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle &= (e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}\otimes (m)_l\otimes (m)_r \\ &\equiv (e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}\otimes (m)\,, \\ (m) &\equiv (m)_l \otimes (m)_r\,.\end{aligned} \tag{2.29}$$

The basis (2.29) will be referred to as an induced basis. With a straightforward computation, that is, making use only of the supercommutation relations (1.1), one derives $(i,j = 1,2,3,4)$

$$\begin{aligned} e_{ij}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle = \{&(-1)^{[(i)+(j)][\theta_1+\theta_2+\theta_3+\theta_4]}(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}e_{ij} \\ &+ \theta_1(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}[-\delta_{i1}\delta_{j3}\theta_3 + \delta_{i1}\delta_{j3}\theta_2 + (-1)^{(i)[\theta_2+\theta_3+\theta_4]}\delta_{j3}e_{i1} \\ &- (-1)^{(j)+[(j)+1][\theta_2+\theta_3+\theta_4]}\delta_{i1}e_{3j}] \\ &+ \theta_2(e_{31})^{\theta_1}(e_{41})^{\theta_3}(e_{42})^{\theta_4}[-\delta_{i2}\delta_{j3}\theta_1 - (-1)^{\theta_1}\delta_{i2}\delta_{j3}\theta_4 \\ &+ (-1)^{[(i)+1]\theta_1+(i)[\theta_3+\theta_4]}\delta_{j3}e_{i2} - (-1)^{(j)[\theta_1+\theta_2]+[(j)+1][\theta_3+\theta_4]}\delta_{i2}e_{3j}] \end{aligned}$$

$$
\begin{aligned}
&+ \theta_3(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{42})^{\theta_4}[- (- 1)^{\theta_1 + \theta_2 + \theta_3}\delta_{i1}\delta_{j4}\theta_4 + (- 1)^{\theta_2}\delta_{i1}\delta_{j4}\theta_1 \\
&+ (- 1)^{[(i) + 1][\theta_1 + \theta_2] + (i)\theta_4}\delta_{j4}e_{i1} - (- 1)^{(j)[\theta_1 + \theta_2 + \theta_3] + [(j) + 1]\theta_4}\delta_{i1}e_{4j}] \\
&+ \theta_4(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}[- (- 1)^{\theta_1 + \theta_2 + \theta_3}\delta_{i2}\delta_{j4}\theta_2 - (- 1)^{\theta_1 + \theta_2}\delta_{i2}\delta_{j4}\theta_3 \\
&+ (- 1)^{[(i) + 1][\theta_1 + \theta_2 + \theta_3]}\delta_{j4}e_{i2} - (- 1)^{(j)[\theta_1 + \theta_2 + \theta_3 + \theta_4]}\delta_{i2}e_{4j}] \\
&- \delta_{i2}\delta_{j3}\theta_1\theta_4(e_{32})^{\theta_2}(e_{41})^{\theta_3 + 1} + (- 1)^{\theta_3}\delta_{i1}\delta_{j4}\theta_1\theta_4(e_{32})^{\theta_2 + 1}(e_{41})^{\theta_3} \\
&- (- 1)^{\theta_1}\delta_{i1}\delta_{j3}\theta_2\theta_3(e_{31})^{\theta_1}(e_{42})^{\theta_4 + 1} + (- 1)^{\theta_1}\delta_{i2}\delta_{j4}\theta_2\theta_3(e_{31})^{\theta_1 + 1}(e_{42})^{\theta_4}\} \otimes (m)_l \otimes (m)_r .
\end{aligned}
\tag{2.30}
$$

From (2.30) it is easy to obtain the transformation of the induced basis under the action of any particular gl(2/2) generator e_{ij}, $i,j = 1,2,3,4$. For later use we write here only the expressions for e_{32} and e_{23}:

$$e_{32}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle = (- 1)^{\theta_1}(1 - \theta_2)|\theta_1,1,\theta_3,\theta_4;(m)\rangle , \tag{2.31}$$

$$
\begin{aligned}
&e_{23}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle \\
&\quad = - \theta_1\theta_2|1,0,\theta_3,\theta_4;(m)\rangle - \theta_1\theta_4|0,\theta_2,\theta_3 + 1,0;(m)\rangle \\
&\quad + \theta_1|0,\theta_2,\theta_3,\theta_4;e_{21}(m)\rangle + (- 1)^{\theta_1}\theta_2|\theta_1,0,\theta_3,\theta_4; \\
&\quad (e_{22} + e_{33})(m)\rangle - (- 1)^{\theta_1}\theta_2\theta_4|\theta_1,0,\theta_3,\theta_4;(m)\rangle \\
&\quad - (- 1)^{\theta_1 + \theta_2 + \theta_3 + \theta_4}\theta_4|\theta_1,\theta_2,\theta_3,0;e_{43}(m)\rangle .
\end{aligned}
\tag{2.32}
$$

D. Typical representations

Proposition 1: Each vector $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle$ from the induced basis (2.29) is a weight vector. The correspondence weight vector $\rightarrow$ weight reads

$$
\begin{aligned}
&|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle \\
&\quad \rightarrow (m_{11} - \theta_1 - \theta_3)e^1 + (m_{13} + m_{23} \\
&\qquad - m_{11} - \theta_2 - \theta_4)e^2 \\
&\qquad + (m_{31} + \theta_1 + \theta_2)e^3 + (m_{33} + m_{43} \\
&\qquad - m_{31} + \theta_3 + \theta_4)e^4 .
\end{aligned}
\tag{2.33}
$$

Proof: From (2.30) one derives that, for any $k = 1,2,3,4$ and $h\in H$,

$$h\,|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle = \lambda(h)|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle , \tag{2.34}$$

where $\lambda\in H^*$. Hence $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle$ is a weight vector with a weight λ. In the basis e^1,e^2,e^3,e^4, λ reads

$$\lambda = \lambda(e_{11})e^1 + \lambda(e_{22})e^2 + \lambda(e_{33})e^3 + \lambda(e_{44})e^4 . \tag{2.35}$$

The coordinates $\lambda(e_{kk})$ of λ are the eigenvalues [see (2.34)] of e_{kk} on $|\theta_1,\theta_2,\theta_3,\theta_4;[m]\rangle$. Using (2.30), with an explicit computation we obtain

$$
\begin{aligned}
&e_{11}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle \\
&\quad = (m_{11} - \theta_1 - \theta_3)|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle ,
\end{aligned}
\tag{2.36}
$$

$$
\begin{aligned}
&e_{22}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle \\
&\quad = (m_{13} + m_{23} - m_{11} - \theta_2 - \theta_4)|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle ,
\end{aligned}
\tag{2.37}
$$

$$
\begin{aligned}
&e_{33}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle \\
&\quad = (m_{31} + \theta_1 + \theta_2)|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle ,
\end{aligned}
\tag{2.38}
$$

$$
\begin{aligned}
&e_{44}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle \\
&\quad = (m_{33} + m_{43} - m_{31} + \theta_3 + \theta_4)|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle .
\end{aligned}
\tag{2.39}
$$

Hence [see (2.34)]

$$
\begin{aligned}
&\lambda(e_{11}) = (m_{11} - \theta_1 - \theta_3) , \\
&\lambda(e_{22}) = (m_{13} + m_{23} - m_{11} - \theta_2 - \theta_4) , \\
&\lambda(e_{33}) = (m_{31} + \theta_1 + \theta_2) , \\
&\lambda(e_{44}) = (m_{33} + m_{43} - m_{31} + \theta_3 + \theta_4) .
\end{aligned}
\tag{2.40}
$$

The last result together with (2.35) gives (2.33).

From (2.33) we conclude that the highest weight vector x_Λ with a weight Λ corresponds to the case $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 0, m_{11} = m_{13}, m_{31} = m_{33}$. Similarly, the lowest weight vector x_V with a weight V corresponds to $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 1, m_{11} = m_{23}, m_{31} = m_{43}$. Up to a multiplicative constant both vectors x_Λ and x_V are uniquely defined from their weights Λ and V. For x_Λ and x_V (2.33) reads

$$
\begin{aligned}
x_\Lambda &= \left|0,0,0,0;\begin{bmatrix} m_{13} & & m_{23} \\ & m_{13} & \end{bmatrix} \otimes \begin{bmatrix} m_{33} & & m_{43} \\ & m_{33} & \end{bmatrix}\right\rangle \\
&\equiv 1 \otimes \begin{bmatrix} m_{13} & & m_{23} \\ & m_{13} & \end{bmatrix} \otimes \begin{bmatrix} m_{33} & & m_{43} \\ & m_{33} & \end{bmatrix} \\
&\rightarrow \Lambda = m_{13}e^1 + m_{23}e^2 + m_{33}e^3 + m_{43}e^4 ,
\end{aligned}
\tag{2.41}
$$

$$
\begin{aligned}
x_V &= \left|1,1,1,1;\begin{bmatrix} m_{13} & & m_{23} \\ & m_{23} & \end{bmatrix} \otimes \begin{bmatrix} m_{33} & & m_{43} \\ & m_{43} & \end{bmatrix}\right\rangle \\
&\equiv e_{31}e_{32}e_{41}e_{42} \otimes \begin{bmatrix} m_{13} & & m_{23} \\ & m_{23} & \end{bmatrix} \otimes \begin{bmatrix} m_{33} & & m_{43} \\ & m_{43} & \end{bmatrix}\Bigg\rangle \\
&\rightarrow V = (m_{23} - 2)e^1 + (m_{13} - 2)e^2 + (m_{43} + 2)e^3 \\
&\qquad + (m_{33} + 2)e^4 .
\end{aligned}
\tag{2.42}
$$

In (2.41) 1 denotes the unity of the gl(2/2) universal enveloping algebra U. Comparing (2.18) with (2.41) we conclude that the $gl(2)_l \oplus gl(2)_r$ module $V_0([m])$ and the gl(2/2) module $W([m])$ have one and the same highest weight.

Proposition 2: The induced module $W([m])$ is a gl(2/2) fidirmod if and only if

$$l_{i3} + l_{j3} + 3 \neq 0 , \quad \forall i = 1,2 \text{ and } j = 3,4 . \tag{2.43}$$

Proof: To start with, we recall that the correspondence root vector $\Leftrightarrow$ root in gl(2/2) reads

$$e_{ij} \Leftrightarrow e^i - e^j , \quad i \neq j = 1,2,3,4. \tag{2.44}$$

Consider the vector

$$x = e_{31}e_{32}e_{41}e_{42} \otimes v \in W([m])\,, \qquad (2.45)$$

$$v = \begin{bmatrix} m_{13} & m_{23} \\ & m_{13} \end{bmatrix} \otimes \begin{bmatrix} m_{33} & m_{43} \\ & m_{33} \end{bmatrix}.$$

The subspace of $W([m])$, defined as

$$W_1 = Ux \equiv \{ux | u \in U\} \qquad (2.46)$$

is transformed into itself under the action of gl(2/2) and is, therefore, a gl(2/2) submodule in $W([m])$. If $W([m])$ is irreducible then $W_1 = W([m])$. In this case there should exist an element $u \in U$ shifting x onto the highest weight vector x_Λ [see (2.41)], i.e., $x_\Lambda \equiv \mathbf{1} \otimes v = ux$ and, in particular, it should be possible to find a monomial of the generators ($c \in \mathbb{C}$)

$$u_0 = c(e_{21})^{n_1}(e_{43})^{n_2}(e_{12})^{n_3}(e_{34})^{n_4}(e_{24})^{\varphi_1}(e_{14})^{\varphi_2}(e_{23})^{\varphi_3}$$
$$\times (e_{13})^{\varphi_4}(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}\,,$$
$$\varphi_i,\theta_i = 0,1, \quad n_i \in \mathbb{Z}_+, \quad i = 1,2,3,4\,, \qquad (2.47)$$

such that

$$u_0\, x \equiv u_0 e_{31}e_{32}e_{41}e_{42} \otimes v = \mathbf{1} \otimes v \equiv x_\Lambda\,. \qquad (2.48)$$

Considering (2.48) as an equation, we now proceed to find an explicit expression for u_0, i.e., to determine the values of n_i, φ_i, and θ_i, $i = 1,2,3,4$. First of all, since in $U\,(e_{ij})^2 = 0$ for $j = 1,2$ and $i = 3,4$, $u_0 x = 0$, if at least one $\theta_i = 1$. Hence the candidate for u_0, which fulfills (2.48), has to be of the form

$$u_0 = c(e_{21})^{n_1}(e_{43})^{n_2}(e_{12})^{n_3}(e_{34})^{n_4}(e_{24})^{\varphi_1}$$
$$\times (e_{14})^{\varphi_2}(e_{23})^{\varphi_3}(e_{13})^{\varphi_4}\,. \qquad (2.49)$$

The next restriction on the possible values of $n_1,\ldots,\varphi_4$ in (2.49) comes from the observation that [see (2.48)] $u_0 e_{31}e_{32}e_{41}e_{42} \otimes v$ and $\mathbf{1} \otimes v$ should have one and the same weight. This will be the case if the weight λ of $u_0 e_{31}e_{32}e_{41}e_{42}$, i.e., of

$$(e_{21})^{n_1}(e_{43})^{n_2}(e_{12})^{n_3}(e_{34})^{n_4}(e_{24})^{\varphi_1}(e_{14})^{\varphi_2}$$
$$\times (e_{23})^{\varphi_3}(e_{13})^{\varphi_4}e_{31}e_{32}e_{41}e_{42} \qquad (2.50)$$

is zero. Taking into account that the weight of a product of weight vectors is given with the sum of the weights of the corresponding multiples and using (2.44), we obtain

$$\lambda = (n_3 - n_1 + \varphi_2 + \varphi_4 - 2)e^1$$
$$+ (n_1 - n_3 + \varphi_1 + \varphi_3 - 2)e^2$$
$$+ (n_4 - n_2 - \varphi_3 - \varphi_4 + 2)e^3$$
$$+ (n_2 - n_4 - \varphi_1 - \varphi_2 + 2)\,. \qquad (2.51)$$

The requirement $\lambda = 0$ gives

$$\varphi_1 = \varphi_2 = \varphi_3 = \varphi_4 = 1\,, \quad n_1 = n_3 = n\,, \quad n_2 = n_4 \equiv m\,. \qquad (2.52)$$

Inserting (2.52) in (2.49) and substituting it in (2.48) we end up with the expression ($n,m \in \mathbb{Z}_+$)

$$c(e_{21})^n(e_{43})^m(e_{12})^n(e_{34})^m e_{24}e_{14}e_{23}e_{13}e_{31}e_{32}e_{41}e_{42} \otimes v$$
$$= \mathbf{1} \otimes v\,, \qquad (2.53)$$

which has to be considered as an equation for the unknowns m and n.

By a straightforward computation one obtains that

$$[e_{ij}, e_{24}e_{14}e_{23}e_{13}e_{31}e_{32}e_{41}e_{42}] = 0\,,$$
$$\forall i \neq j = 1,2 \text{ or } i \neq j = 3,4\,. \qquad (2.54)$$

Therefore, using also (2.22), we can write the left-hand side of (2.53) in the form

$$ce_{24}e_{14}e_{23}e_{13}e_{31}e_{32}e_{41}e_{42} \otimes (e_{21})^n(e_{43})^m(e_{12})^n(e_{34})^m v\,. \qquad (2.55)$$

Since v is the highest weight vector of the $\mathrm{gl}(2)_l \oplus \mathrm{gl}(2)_r$ irreducible module $V_0([m])$, it is annihilated by e_{12} and e_{34} [see also (2.12)]. Therefore, (2.55) is a nonzero vector from $W([m])$ only if $m = n = 0$. Thus the monomial u_0 [see (2.47)] has to be of the form

$$u_0 = ce_{24}e_{14}e_{23}e_{13}\,, \qquad (2.56)$$

where c is an arbitrary constant. We underline that the expression (2.56) for u_0 gives only a necessary condition for a solution of Eq. (2.48). In order to see when it is also sufficient, insert (2.56) in (2.48),

$$ce_{24}e_{14}e_{23}e_{13}e_{31}e_{32}e_{41}e_{42} \otimes v = \mathbf{1} \otimes v\,. \qquad (2.57)$$

Using the supercommutation relations (1.1), the definition of $W([m])$, and the relations (2.10)–(2.14) and (2.20), one obtains for the left-hand side of (2.57)

$$ce_{24}e_{14}e_{23}e_{13}e_{31}e_{32}e_{41}e_{42} \otimes v$$
$$= c(l_{13} + l_{33} + 3)(l_{13} + l_{43} + 3)$$
$$\times (l_{23} + l_{33} + 3)(l_{23} + l_{43} + 3) \otimes v. \qquad (2.58)$$

Inserting (2.58) in (2.57) we conclude that u_0, defined with (2.56), gives a solution only if the conditions (2.43) are fulfilled and

$$c = 1/(l_{13} + l_{33} + 3)(l_{13} + l_{43} + 3)$$
$$\times (l_{23} + l_{33} + 3)(l_{23} + l_{43} + 3).$$

In other words, if

$$\exists i = 1,2 \text{ and } j = 3,4 \text{ such that } l_{i3} + l_{j3} + 3 = 0\,, \qquad (2.59)$$

then [see (2.41), (2.45), and (2.46)] $x_\Lambda \equiv \mathbf{1} \otimes v \in W_1 \equiv Ux$. Hence $W_1 \neq W([m])$, i.e., $W([m])$ contains a proper gl(2/2) invariant subspace W_1. Therefore, if (2.59) holds, then $W([m])$ is not an irreducible gl(2/2) module. Moreover, since [see (2.45)]

$$x = e_{31}e_{32}e_{41}e_{42} \otimes v = e_{31}e_{32}e_{41}e_{42}(\mathbf{1} \otimes v)$$
$$= e_{31}e_{32}e_{41}e_{42}x_\Lambda\,, \qquad (2.60)$$

there exists no complement to the W_1 subspace, which is invariant under gl(2/2). Hence, if (2.59) holds, $W([m])$ is an indecomposable gl(2/2) module.

In order to complete the proof, assume that the conditions (2.43) hold and take any two elements $y,z \in W([m])$. From the very construction of $W([m])$ and the irreducibility of the $\mathrm{gl}(2)_l \oplus \mathrm{gl}(2)_r$ module $V_0([m])$ it is clear that

$$\exists u_1 \in U, \text{ such that } y = u_1 x_\Lambda\,. \qquad (2.61)$$

It is not difficult to show that [see (2.45)]

$$\exists u_2 \in U, \text{ such that } x = u_2 z\,. \qquad (2.62)$$

Combining (2.48), (2.56), (2.61), and (2.62), we obtain

$$y = cu_1 e_{24}e_{14}e_{23}e_{13}u_2 z\,. \qquad (2.63)$$

Hence, if the conditions (2.34) are fulfilled, $W([m])$ is a gl(2/2) fidirmod. ■

Proposition 2 could have been derived also from the results of Kac[4] on the typical representations of the subalgebra sl(2/2). We have given the above proof because it deals directly with entities that are more appropriate for gl(2/2) and is, moreover, self-contained. Adopting the terminology of Ref. 4 for the representations of the basic Lie superalgebras, we give the following definition.

Definition: The representation of gl(2/2) is typical if it can be realized in an irreducible induced module $W([m])$. The corresponding typical representation modules will also be called typical.

This definition is consistent with the terminology for the basic LS's, because each typical (in the above definition) gl(2/2) module is also typical (in the sense of Ref. 4) with respect to the subalgebra sl(2/2). From Proposition 2 we now have that the induced gl(2/2) module is typical and hence irreducible if and only if the conditions (2.43) hold.

III. REPRESENTATIONS OF gl(2/2) IN A gl(2)⊕gl(2) BASIS

A. Structure of $V([m])$ with respect to gl(2)⊕gl(2)

Consider the gl(2/2) module $W([m]) = T\otimes V_0([m])$ as a representation space of the even subalgebra $gl(2)_l \oplus gl(2)_r$ and let $e_{ij}\in gl(2)_l\oplus gl(2)_r$. By virtue of the fact that

$$\begin{aligned}
&[\![e_{ij},(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}]\!]\\
&\quad=(-1)^{\theta_2+\theta_3+\theta_4}\theta_1(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4}\\
&\qquad\times[\delta_{j3}e_{i1}-\delta_{i1}e_{3j}]\\
&\qquad+(-1)^{\theta_3+\theta_4}\theta_2(e_{31})^{\theta_1}(e_{41})^{\theta_3}(e_{42})^{\theta_4}\\
&\qquad\times[\delta_{j3}e_{i2}-\delta_{i2}e_{3j}]\\
&\qquad+(-1)^{\theta_4}\theta_3(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{42})^{\theta_4}[\delta_{j4}e_{i1}-\delta_{i1}e_{4j}]\\
&\qquad+\theta_4(e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}[\delta_{j4}e_{i2}-\delta_{i2}e_{4j}], \qquad (3.1)
\end{aligned}$$

the subspace T of U is invariant in relation to the adjoint representation of U, restricted to the even subalgebra,

$$[\![gl(2)_l\oplus gl(2)_r, T]\!]\subset T. \qquad (3.2)$$

We, therefore, can consider T as a $gl(2)_l\oplus gl(2)_r$ module. Since, on the other hand, for every $g\in gl(2)_l\oplus gl(2)_r$ and $t\otimes v\in V_0([m])$

$$g(t\otimes v) = (\mathrm{ad}\, g)t\otimes v + t\otimes gv, \qquad (3.3)$$

the representation of the even subalgebra is realized in the tensor product $T\odot V_0([m])$ of the $gl(2)_l\oplus gl(2)_r$ modules T and $V_0([m])$, i.e., this representation is a tensor product of two $gl(2)_l\oplus gl(2)_r$ representations. We have introduced the notation $\odot$ instead of $\otimes$ in order to underline that the tensor product is between two modules of one and the same algebra, namely, $gl(2)_l\oplus gl(2)_r$. The symbol $\otimes$ will be reserved for a tensor product of modules over different algebras (which is a representation space of the direct sum of these algebras). For instance, if V_l is a $gl(2)_l$ module and V_r is a $gl(2)_r$ module, then we write $V_l\otimes V_r$ for the module over $gl(2)_l\oplus gl(2)_r$. In the sense of linear spaces $\odot$ and $\otimes$ coincide with the usual tensor product.

The $gl(2)_l\oplus gl(2)_r$ module T is reducible. It is a direct sum of six irreducible submodules,

$$\begin{aligned}
T = {}& T_1([0,0,0,0])\oplus T_2([0,-1,1,0])\\
&\oplus T_3([-1,-1,2,0])\oplus T_4([0,-2,1,1])\\
&\oplus T_5([-1,-2,2,1])\oplus T_6([-2,-2,2,2]). \qquad (3.4)
\end{aligned}$$

In the brackets of each T_i we have written the coordinates of the corresponding to T_i highest weight Λ_i in the dual basis e^1,e^2,e^3,e^4 [see (2.17)]. For instance, $\Lambda_2 = -e^2+e^3$. The structure of every T_i as a subset of U is

$$T_1 = \text{lin. env.}\{(e_{31})^0(e_{32})^0(e_{41})^0(e_{42})^0\equiv \mathbf{1}\} = \mathbb{C}, \qquad (3.5)$$

$$T_2 = \text{lin. env.}\{e_{31},e_{32},e_{41},e_{42}\}, \qquad (3.6)$$

$$T_3 = \text{lin. env.}\{e_{31}e_{32},e_{31}e_{42}-e_{32}e_{41},e_{41}e_{42}\}, \qquad (3.7)$$

$$T_4 = \text{lin. env.}\{e_{32}e_{42},e_{31}e_{42}+e_{32}e_{41},e_{31}e_{41}\}, \qquad (3.8)$$

$$T_5 = \text{lin. env.}\{e_{31}e_{32}e_{41},e_{31}e_{32}e_{42},e_{31}e_{41}e_{42},e_{32}e_{41}e_{42}\}, \qquad (3.9)$$

$$T_6 = \text{lin. env.}\{e_{31}e_{32}e_{41}e_{42}\}. \qquad (3.10)$$

The vectors in the brackets $\{\cdots\}$ of (3.5)–(3.10) are linearly independent and define a basis in each T_i. For further use it is convenient to relate this basis with the canonical basis (2.9). The relation, which is consistent with the action of the even subalgebra, reads

in $T_1([0,0,0,0])$,

$$\begin{bmatrix}0, & 0\\ & 0\end{bmatrix}\otimes\begin{bmatrix}0, & 0\\ & 0\end{bmatrix} = (e_{31})^0(e_{32})^0(e_{41})^0(e_{42})^0 = 1; \qquad (3.11)$$

in $T_2([0,-1,1,0])$,

$$\begin{aligned}
&\begin{bmatrix}0, & -1\\ & -1\end{bmatrix}\otimes\begin{bmatrix}1, & 0\\ & 1\end{bmatrix} = e_{31}, \quad \begin{bmatrix}0, & -1\\ & 0\end{bmatrix}\otimes\begin{bmatrix}1, & 0\\ & 1\end{bmatrix} = -e_{32},\\
&\begin{bmatrix}0, & -1\\ & -1\end{bmatrix}\otimes\begin{bmatrix}1, & 0\\ & 0\end{bmatrix} = e_{41}, \quad \begin{bmatrix}0, & -1\\ & 0\end{bmatrix}\otimes\begin{bmatrix}1, & 0\\ & 0\end{bmatrix} = -e_{42};
\end{aligned} \qquad (3.12)$$

in $T_3([-1,-1,2,0])$,

$$\begin{aligned}
&\begin{bmatrix}-1, & -1\\ & -1\end{bmatrix}\otimes\begin{bmatrix}2, & 0\\ & 2\end{bmatrix} = e_{31}e_{32}, \quad \begin{bmatrix}-1, & -1\\ & -1\end{bmatrix}\otimes\begin{bmatrix}2, & 0\\ & 0\end{bmatrix} = e_{41}e_{42},\\
&\begin{bmatrix}-1, & -1\\ & -1\end{bmatrix}\otimes\begin{bmatrix}2, & 0\\ & 1\end{bmatrix} = |2|^{1/2}e_-, \quad e_- = \tfrac{1}{2}(e_{31}e_{42}-e_{32}e_{41});
\end{aligned} \qquad (3.13)$$

in $T_4([0,-2,1,1])$

$$\begin{bmatrix}0, & -2\\ & -2\end{bmatrix}\otimes\begin{bmatrix}1, & 1\\ & 1\end{bmatrix}=e_{31}e_{41}\,,\quad \begin{bmatrix}0, & -2\\ & 0\end{bmatrix}\otimes\begin{bmatrix}1, & 1\\ & 1\end{bmatrix}=e_{32}e_{42}\,,$$

$$\begin{bmatrix}0, & -2\\ & -1\end{bmatrix}\otimes\begin{bmatrix}1, & 1\\ & 1\end{bmatrix}=-|2|^{1/2}e_+,\quad e_+=\tfrac{1}{2}(e_{31}e_{42}+e_{32}e_{41})\,; \tag{3.14}$$

in $T_5([-1,-2,2,1])$

$$\begin{bmatrix}-1, & -2\\ & -1\end{bmatrix}\otimes\begin{bmatrix}2, & 1\\ & 2\end{bmatrix}=e_{31}e_{32}e_{42}\,,\quad \begin{bmatrix}-1, & -2\\ & -2\end{bmatrix}\otimes\begin{bmatrix}2, & 1\\ & 1\end{bmatrix}=e_{31}e_{41}e_{42}\,,$$

$$\begin{bmatrix}-1, & -2\\ & 2\end{bmatrix}\otimes\begin{bmatrix}2, & 1\\ & 2\end{bmatrix}=-e_{31}e_{32}e_{41}\,,\quad \begin{bmatrix}-1, & -2\\ & -1\end{bmatrix}\otimes\begin{bmatrix}2, & 1\\ & 1\end{bmatrix}=-e_{32}e_{41}e_{42}\,; \tag{3.15}$$

in $T_6([-2,-2,2,2])$

$$\begin{bmatrix}-2, & -2\\ & -2\end{bmatrix}\otimes\begin{bmatrix}2, & 2\\ & 2\end{bmatrix}=e_{31}e_{32}e_{41}e_{42}\,. \tag{3.16}$$

Inserting (3.4) in (2.28), we have

$$W([m])=D_1\oplus D_2\oplus D_3\oplus D_4\oplus D_5\oplus D_6\,, \tag{3.17}$$

where each

$$D_i=T_i\odot V_0([m])\,,\quad i=1,\dots,6\,, \tag{3.18}$$

is an invariant subspace of the even subalgebra. The subspaces D_1 and D_6 are $gl(2)_l\oplus gl(2)_r$ irreducible. We set

$$D_1=T_1([0,0,0,0])\odot V_0([m])$$
$$=V_1([m_{13},m_{23},m_{33},m_{43}])\,, \tag{3.19}$$
$$D_6=T_6([-2,-2,2,2])\odot V_0([m])$$
$$=V_6([m_{13}-2,m_{23}-2,m_{33}+2,m_{43}+2])\,. \tag{3.20}$$

The other four subspaces are reducible. In order to decompose them into $gl(2)_l\oplus gl(2)_r$ fidirmods we proceed as follows.

(1) Each $gl(2)_l\oplus gl(2)_r$ fidirmod $V_0([m_{13},m_{23},m_{33},m_{43}])$ transforms as a tensor $\otimes$ product of a $gl(2)_l$ fidirmod $V_0([m_{13},m_{23}])$ and a $gl(2)_r$ fidirmod $V_0([m_{33},m_{43}])$,

$$V_0([m_{13},m_{23},m_{33},m_{43}])$$
$$=V_0([m_{13},m_{23}])\otimes V_0([m_{33},m_{43}])\,. \tag{3.21}$$

This, in particular, holds for T_i $(i=1,\dots,6)$,

$$T_i([p_{13},p_{23},p_{33},p_{43}])$$
$$=T_i([p_{13},p_{23}])\otimes T_i([p_{33},p_{43}])\,. \tag{3.22}$$

(2) The tensor $\odot$ product of the $gl(2)_l\oplus gl(2)_r$ modules (3.21) and (3.22) transforms as a tensor $\otimes$ product of a $gl(2)_l$ module

$$T_i([p_{13},p_{23}])\odot V_0([m_{13},m_{23}])$$

and a $gl(2)_r$ module

$$T_i([p_{33},p_{43}])\odot V_0([m_{33},m_{43}])\,,\quad i=1,\dots,6\,,$$

i.e.,

$$(T_i([p_{13},p_{23}])\otimes T_i([p_{33},p_{43}]))\odot(V_0([m_{13},m_{23}])$$
$$\otimes V_0([m_{33},m_{43}]))$$
$$=(T_i([p_{13},p_{23}])\odot V_0([m_{13},m_{23}]))$$
$$\otimes(T_i([p_{33},p_{43}])\odot V_0([,m_{33},m_{43}]))\,. \tag{3.23}$$

In general, every $gl(2)_l$ module $T_i([p_{13},p_{23}])\odot V_0([m_{13},m_{23}])$ is reducible. Its decomposition as a direct sum of irreducible modules can be easily carried out.[12] The result is

$$T_i([p_{13},p_{23}])\odot V_0([m_{13},m_{23}])$$
$$=\sum_{k=0}^{n_l}\oplus V_i([m_{13}+p_{13}-k,m_{23}+p_{23}+k])\,. \tag{3.24}$$

Similarly, replacing in (3.23) $13\to 33$ and $23\to 43$, we have

$$T_i([p_{33},p_{43}])\odot V_0([m_{33},m_{43}])$$
$$=\sum_{k=0}^{n_r}\oplus V_i([m_{33}+p_{33}-k,m_{43}+p_{43}+k])\,, \tag{3.25}$$

where

$$n_l=\min(p_{13}-p_{23},m_{13}-m_{23})$$

and

$$n_r=\min(p_{33}-p_{43},m_{33}-m_{43})\,. \tag{3.26}$$

Inserting (3.24) and (3.25) in (2.23) for every $i=1,\dots,6$, and setting

$$m_{12}=m_{13}+p_{13}-k\,,\quad m_{22}=m_{23}+p_{23}+k\,,$$
$$m_{32}=m_{33}+p_{33}-k\,,\quad m_{42}=m_{43}+p_{43}+k\,,$$
$$V_i([m_{12},m_{22}])\otimes V_i([m_{32},m_{42}])$$
$$=V_i([m_{12},m_{22},m_{32},m_{42}])\,, \tag{3.27}$$

we obtain the decomposition of $T_i\odot V_0([m])$ into irreducible $gl(2)_l\oplus gl(2)_r$ modules $V_i([m_{12},m_{22},m_{32},m_{42}])$,

$$D_2=T_2([0,-1,1,0])\odot V_0([m_{12},m_{23},m_{33},m_{43}])$$
$$=\sum_{i,j=0}^{1}{}'\oplus V_2([m_{13}-i,m_{23}+i-1,m_{33}$$
$$-j+1,m_{43}+j])\,, \tag{3.28}$$
$$D_3=T_3([-1,-1,2,0])\odot V_0([m_{13},m_{23},m_{33},m_{43}])$$
$$=\sum_{j=0}^{2}{}'\otimes V_3([m_{13}-1,m_{23}-1,m_{33}-j+2,m_{43}+j])\,, \tag{3.29}$$

$$D_4 = T_4([0,-2,1,1]) \odot V_0([m_{13},m_{23},m_{33},m_{43}]) = \sum_{i=0}^{2}{}' \otimes V_4([m_{13}-i,m_{23}+i-2,m_{33}+1,m_{43}+1]), \tag{3.30}$$

$$D_5 = T_5([-1,-2,2,1]) \odot V_0([m_{13},m_{23},m_{33},m_{43}]) = \sum_{i,j=0}^{1}{}' \oplus V_5([m_{13}-i-1,m_{23}+i-2,m_{33}-j+2,m_{43}+j+1]). \tag{3.31}$$

The prime on the sum in (3.21)–(3.31) is to recall that all nonlexical terms, i.e., those for which the condition (2.2) is not fulfilled for the left or for the right gl(2) fidirmods, have to be deleted from the sum.

From (3.17), (3.19), (3.20), and (3.28)–(3.31) we conclude that the induced gl(2/2) module $W([m])$ decomposes into a direct sum of (no more than) 16 irreducible modules of the even subalgebra $gl(2)_l \oplus gl(2)_r$ (we skip the prime over the sum),

$$\begin{aligned} W([m]) = {} & V_1([m_{13},m_{23},m_{33},m_{43}]) \\ & \oplus \sum_{i,j=0}^{1} V_2([m_{13}-i,m_{23}+i-1,m_{33}-j+1,m_{43}+j]) \\ & \oplus \sum_{j=0}^{2} \oplus V_3([m_{13}-1,m_{23}-1,m_{33}-j+2,m_{43}+j]) \\ & \oplus \sum_{i=0}^{2} \oplus V_4([m_{13}-i,m_{23}+i-2,m_{33}+1,m_{43}+1]) \\ & \oplus \sum_{i,j=0}^{1} \oplus V_5([m_{13}-i-1,m_{23}+i-2,m_{33}-j+2,m_{43}+j+1]) \\ & \oplus V_6([m_{13}-2,m_{23}-2,m_{33}+2,m_{43}+2]). \end{aligned} \tag{3.32}$$

B. A gl(2)⊕gl(2) basis

So far we have introduced one possible basis, the induced basis (2.29), within every induced gl(2/2) module $W([m])$. We have written down also explicit expressions [see (2.30)] for the transformation of this basis under the action of the generators. The induced basis is inconvenient, however, for the description of all irreducible representations of gl(2/2). In certain cases, namely if some of the conditions (2.43) are not fulfilled, the induced module $W([m])$ contains a maximal invariant subspace $I([m])$. In order to obtain the fidirmod, corresponding to the signature $[m]$, one has to go to the factor module $W([m])/I([m])$. A convenient way to do this is to introduce a basis $e_1,...,e_i,...,e_N$ in $W([m])$ such that each e_i is either from $I([m])$ or from a complement subspace to $I([m])$. In such a case to go to the factor space means simply to replace all basis vectors $e_i \in I([m])$ by zero. As it will become clear, the vectors from the induced basis do not have the properties of a basis $e_1,...,e_N$. Therefore, the very determination of $I([m])$ in terms of the induced basis is very difficult. This is one reason to go to a new basis. We choose it in such a way that every basis vector belongs to one and only one of the 16 $gl(2)_l \oplus gl(2)_r$ irreducible submodules $V_i(\cdots)$, written in the right-hand side of (3.32). As we shall see in Ref. 1, the new basis will help us considerably in the study of the nontypical modules. This new basis makes evident the decomposition of the gl(2/2) modules along the chain $gl(2/2) \supset gl(2)_l \oplus gl(2)_r$, which can be of interest in several physical applications.

Let $V_i(m_{12},m_{22},m_{32},m_{42})$ be one of the $gl(2)_l \oplus gl(2)_r$ modules in (3.42). As a basis $\Gamma_i(m_{12},m_{22},m_{32},m_{42})$ in it we choose the canonical basis (2.9), which in this case is denoted as

$$\begin{bmatrix} m_{12} & m_{22} & ; & m_{32} & m_{42} \\ & m'_{11} & & m'_{31} & \end{bmatrix}_i \equiv \begin{bmatrix} m_{12} & m_{22} \\ m'_{11} & \end{bmatrix} \otimes \begin{bmatrix} m_{32} & m_{42} \\ m'_{31} & \end{bmatrix} \equiv \begin{bmatrix} [m]_l \\ m'_{11} \end{bmatrix} \otimes \begin{bmatrix} [m]_r \\ m'_{31} \end{bmatrix} \equiv (m)_i . \tag{3.33}$$

Then

$$\Gamma_i = \cup \Gamma_i(m_{12},m_{22},m_{32},m_{42}), \tag{3.34}$$

where the union is over all signatures $m_{12},m_{22},m_{32},m_{42}$ in D_i [see the decompositions (3.19), (3.20), (3.28)–(3.31)], constitutes a new basis in D_i. As a basis $\Gamma([m])$ in the induced module $W([m])$ we set

$$\Gamma([m]) = \bigcup_{i=1}^{6} \Gamma_i(m_{12},m_{22},m_{32},m_{42}). \tag{3.35}$$

We call this basis a $gl(2)_l \oplus gl(2)_r$ reduced basis or simply reduced basis.

Our next and final task in this paper will be to write the transformation of the induced modules in terms of the reduced basis $\Gamma([m])$. To this end we first proceed to express each vector

$$|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle = (e_{31})^{\theta_1}(e_{32})^{\theta_2}(e_{41})^{\theta_3}(e_{42})^{\theta_4} \odot (m)_l \otimes (m)_r \tag{3.36}$$

from the induced basis (2.29) in terms of the reduced basis (3.43). Every induced basis vector $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle$ except $|1,0,0,1;(m)\rangle$ and $|0,1,1,0;(m)\rangle$ belongs to some $D_i \subset W([m])$ [see (3.17)]. If we wish to underline that $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle \in D_i$ we add a subscript i to it, writing $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i$,

$$|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i \in D_i\,. \tag{3.37}$$

Every vector $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i$ can be decomposed in the basis Γ_i,

$$|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i = \sum_{(m)_i\in\Gamma_i} ((m)_i|\theta_1,\theta_2,\theta_3,\theta_4;(m))_i (m)_i\,, \tag{3.38}$$

where $((m)_i|\theta_1,\theta_2,\theta_3,\theta_4;(m))_i$, $(m)_i\in\Gamma_i$, are the coordinates of $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i$ in the basis Γ_i of D_i. In order to compute these coordinates, take any vector $|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i$. According to (2.9) and (3.11)–(3.16) and taking into account also (3.23), we can write each vector (3.37) in the form

$$\begin{aligned}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i &= \left(\begin{bmatrix}p_{13} & p_{23}\\ & p_{11}\end{bmatrix}_i \otimes \begin{bmatrix}p_{33} & p_{43}\\ & p_{31}\end{bmatrix}_i\right) \odot \left(\begin{bmatrix}m_{13} & m_{23}\\ & m_{11}\end{bmatrix} \otimes \begin{bmatrix}m_{33} & m_{43}\\ & m_{31}\end{bmatrix}\right)\\ &= \left(\begin{bmatrix}p_{13} & p_{23}\\ & p_{11}\end{bmatrix}_i \odot \begin{bmatrix}m_{13}, & m_{23}\\ & m_{11}\end{bmatrix}\right) \otimes \left(\begin{bmatrix}p_{33} & p_{43}\\ & p_{31}\end{bmatrix}_i \odot \begin{bmatrix}m_{33} & m_{43}\\ & m_{31}\end{bmatrix}\right),\end{aligned} \tag{3.39}$$

where we have inserted a subscript i in order to underline that

$$\begin{bmatrix}p_{13} & p_{23}\\ & p_{11}\end{bmatrix}_i \otimes \begin{bmatrix}p_{33} & p_{43}\\ & p_{31}\end{bmatrix}_i \in\Gamma_i\,. \tag{3.40}$$

The relations between $\theta_1,\ldots,\theta_4$ and all p_{ij} in the right-hand side of (3.39) can be easily written down [see (3.11)–(3.16)]. From (2.33) we derive

$$\begin{aligned}&-p_{13}-p_{23} = p_{33}+p_{43} = \theta_1+\theta_2+\theta_3+\theta_4\,,\\ &p_{11} = -\theta_1-\theta_3\,,\quad p_{31} = \theta_1+\theta_2\,.\end{aligned} \tag{3.41}$$

For $|1,0,0,1;(m)\rangle$ and $|0,1,1,0;(m)\rangle$ we obtain

$$\begin{aligned}|0,1,1,0;(m)\rangle = &-|2|^{-1/2}\left(\begin{bmatrix}-1 & -1\\ & -1\end{bmatrix}_3 \odot \begin{bmatrix}m_{13} & m_{23}\\ & m_{11}\end{bmatrix}\right) \otimes \left(\begin{bmatrix}2 & 0\\ & 1\end{bmatrix}_3 \odot \begin{bmatrix}m_{33} & m_{43}\\ & m_{31}\end{bmatrix}\right)\\ &-|2|^{-1/2}\left(\begin{bmatrix}0 & -2\\ & -1\end{bmatrix}_4 \odot \begin{bmatrix}m_{13} & m_{23}\\ & m_{11}\end{bmatrix}\right) \otimes \left(\begin{bmatrix}1 & 1\\ & 1\end{bmatrix}_4 \odot \begin{bmatrix}m_{33} & m_{43}\\ & m_{31}\end{bmatrix}\right),\end{aligned} \tag{3.42}$$

$$\begin{aligned}|1,0,0,1;(m)\rangle = &\,|2|^{-1/2}\left(\begin{bmatrix}-1 & -1\\ & -1\end{bmatrix}_3 \odot \begin{bmatrix}m_{13} & m_{23}\\ & m_{11}\end{bmatrix}\right) \otimes \left(\begin{bmatrix}2 & 0\\ & 1\end{bmatrix}_3 \odot \begin{bmatrix}m_{33} & m_{43}\\ & m_{31}\end{bmatrix}\right)\\ &-|2|^{-1/2}\left(\begin{bmatrix}0 & -2\\ & -1\end{bmatrix}_4 \odot \begin{bmatrix}m_{13} & m_{23}\\ & m_{11}\end{bmatrix}\right) \otimes \left(\begin{bmatrix}1 & 1\\ & 1\end{bmatrix}_4 \odot \begin{bmatrix}m_{33} & m_{43}\\ & m_{31}\end{bmatrix}\right).\end{aligned} \tag{3.43}$$

According to (3.24)

$$\begin{bmatrix}p_{13} & p_{23}\\ & p_{11}\end{bmatrix}_i \odot \begin{bmatrix}m_{13} & m_{23}\\ & m_{11}\end{bmatrix} \in \sum_{k=0}^{n_l} V_i([m_{13}+p_{13}-k, m_{23}+p_{23}+k])\,. \tag{3.44}$$

Therefore, denoting by

$$\begin{bmatrix}m_{12} & m_{22}\\ & m'_{11}\end{bmatrix},\quad m_{12} = m_{13}+p_{13}-k\,,\quad m_{22} = m_{23}+p_{23}+k \tag{3.45}$$

the GZ basis in $V_i([m_{13}+p_{13}-k, m_{23}+p_{23}+k])$, we can write

$$\begin{bmatrix}p_{13} & p_{23}\\ & p_{11}\end{bmatrix}_i \odot \begin{bmatrix}m_{13} & m_{23}\\ & m_{11}\end{bmatrix} = \sum_{(m)_l} \left[\begin{matrix}m_{12} & m_{22}\\ & m'_{11}\end{matrix}\middle|\begin{matrix}p_{13} & p_{23}\\ & p_{11}\end{matrix};\begin{matrix}m_{13} & m_{23}\\ & m_{11}\end{matrix}\right] \begin{bmatrix}m_{12} & m_{22}\\ & m'_{11}\end{bmatrix}, \tag{3.46}$$

where the sum in (3.46) is over all GZ basis vectors $(m)_l$ from the subspace, defined by the right-hand side of (3.44). By definition

$$\left[\begin{matrix}m_{12} & m_{22}\\ & m'_{11}\end{matrix}\middle|\begin{matrix}p_{13} & p_{23}\\ & p_{11}\end{matrix};\begin{matrix}m_{13} & m_{23}\\ & m_{11}\end{matrix}\right]$$

are the gl(2) Clebsch–Gordan coefficients. In a similar way we have for the right gl(2),

$$\begin{bmatrix}p_{33} & p_{43}\\ & p_{31}\end{bmatrix}_i \odot \begin{bmatrix}m_{33} & m_{43}\\ & m_{31}\end{bmatrix} = \sum_{(m)_r} \left[\begin{matrix}m_{32} & m_{42}\\ & m'_{31}\end{matrix}\middle|\begin{matrix}p_{33} & p_{43}\\ & p_{31}\end{matrix};\begin{matrix}m_{33} & m_{43}\\ & m_{31}\end{matrix}\right] \times \begin{bmatrix}m_{32} & m_{42}\\ & m'_{31}\end{bmatrix} \tag{3.47}$$

and the sum is over all possible GZ basis vectors $(m)_r$.

Inserting (3.46) and (3.47) in (3.39) and taking into account (3.33) we finally obtain

$$\begin{aligned}|\theta_1,\theta_2,\theta_3,\theta_4;(m)\rangle_i = \sum_{(m)_i} &\left[\begin{matrix}m_{12} & m_{22}\\ & m'_{11}\end{matrix}\middle|\begin{matrix}p_{13} & p_{23}\\ & p_{11}\end{matrix};\begin{matrix}m_{13} & m_{23}\\ & m_{11}\end{matrix}\right]\\ \times&\left[\begin{matrix}m_{32} & m_{42}\\ & m'_{31}\end{matrix}\middle|\begin{matrix}p_{33} & p_{43}\\ & p_{31}\end{matrix};\begin{matrix}m_{33} & m_{43}\\ & m_{31}\end{matrix}\right]\\ \times&\begin{bmatrix}m_{12} & m_{22}; & m_{32} & m_{42}\\ & m'_{11} & & m'_{31}\end{bmatrix}_i.\end{aligned} \tag{3.48}$$

In terms of the reduced basis the vectors $|0,1,1,0;(m)\rangle$ and $|1,0,0,1;(m)\rangle$ [see (3.42) and (3.43)] read

$$\begin{aligned}|0,1,1,0;(m)\rangle = &-|2|^{-1/2}\\ &\times\sum_{(m)_3} \left[\begin{matrix}m_{12} & m_{22}\\ & m'_{11}\end{matrix}\middle|\begin{matrix}-1 & -1\\ & -1\end{matrix};\begin{matrix}m_{13} & m_{23}\\ & m_{11}\end{matrix}\right]\\ &\times\left[\begin{matrix}m_{32} & m_{42}\\ & m'_{31}\end{matrix}\middle|\begin{matrix}2 & 0\\ & 1\end{matrix};\begin{matrix}m_{33} & m_{43}\\ & m_{31}\end{matrix}\right]\end{aligned}$$

$$\times \begin{bmatrix} m_{12} & m_{22} & ; & m_{32} & m_{42} \\ & m'_{11} & & m'_{31} & \end{bmatrix}_3$$

$$- |2|^{-1/2} \sum_{(m)_4} \left[\begin{matrix} m_{12} & m_{22} \\ & m'_{11} \end{matrix}\middle| \begin{matrix} 0 & -2 \\ & -1 \end{matrix} ; \begin{matrix} m_{13} & m_{23} \\ & m_{11} \end{matrix}\right]$$

$$\times \left[\begin{matrix} m_{32} & m_{42} \\ & m'_{31} \end{matrix}\middle| \begin{matrix} 1 & 1 \\ & 1 \end{matrix} ; \begin{matrix} m_{33} & m_{43} \\ & m_{31} \end{matrix}\right]$$

$$\times \begin{bmatrix} m_{12} & m_{22} & ; & m_{32} & m_{42} \\ & m'_{11} & & m'_{31} & \end{bmatrix}_4 , \tag{3.49}$$

$|1,0,0,1;(m)\rangle$

$$= |2|^{-1/2} \sum_{(m)_3} \left[\begin{matrix} m_{12} & m_{22} \\ & m'_{11} \end{matrix}\middle| \begin{matrix} -1 & -1 \\ & -1 \end{matrix} ; \begin{matrix} m_{13} & m_{23} \\ & m_{11} \end{matrix}\right]$$

$$\times \left[\begin{matrix} m_{32} & m_{42} \\ & m'_{31} \end{matrix}\middle| \begin{matrix} 2 & 0 \\ & 1 \end{matrix} ; \begin{matrix} m_{33} & m_{43} \\ & m_{31} \end{matrix}\right]$$

$$\times \begin{bmatrix} m_{12} & m_{22} & ; & m_{32} & m_{42} \\ & m'_{11} & & m'_{31} & \end{bmatrix}_3$$

$$- |2|^{-1/2} \sum_{(m)_4} \left[\begin{matrix} m_{12} & m_{22} \\ & m'_{11} \end{matrix}\middle| \begin{matrix} 0 & -2 \\ & -1 \end{matrix} ; \begin{matrix} m_{13} & m_{23} \\ & m_{11} \end{matrix}\right]$$

$$\times \left[\begin{matrix} m_{32} & m_{42} \\ & m'_{31} \end{matrix}\middle| \begin{matrix} 1 & 1 \\ & 1 \end{matrix} ; \begin{matrix} m_{33} & m_{43} \\ & m_{31} \end{matrix}\right]$$

$$\times \begin{bmatrix} m_{12} & m_{22} & ; & m_{32} & m_{42} \\ & m'_{11} & & m'_{31} & \end{bmatrix}_4 . \tag{3.50}$$

The gl(2) Clebsch–Gordan coefficients (CG coefficients) in the GZ basis are available from the literature. They are related to the CG coefficients $C^{n;p,m}_{n_1;p_1,m_1}$ of the algebra so(3),

$$\left[\begin{matrix} n_{12} & n_{22} \\ & n_{11} \end{matrix}\middle| \begin{matrix} p_{12} & p_{22} \\ & p_{11} \end{matrix} ; \begin{matrix} m_{12} & m_{22} \\ & m_{11} \end{matrix}\right] = C^{n;p,m}_{n_1;p_1,m_1} , \tag{3.51}$$

where

$$j = \tfrac{1}{2}(j_{12} - j_{22}) , \quad j_1 = j_{11} - \tfrac{1}{2}(j_{12} + j_{22}) , \quad j = n,p,m. \tag{3.52}$$

As is well known

$$C^{n;p,m}_{n_1;p_1,m_1} = 0 , \quad \text{if} \quad n \neq m + p, m + p - 1, \ldots, |m - p| \quad \text{or } n_1 \neq m_1 + p_1 . \tag{3.53}$$

In the calculations we have used the expressions for the so(3) CG coefficient, given in Ref. 13. Having established the relation between the induced basis (2.29) and the reduced basis (3.33), all we have to do is to write the transformations (2.30) in terms of the new basis. We skip the rather long intermediate computations of this standard mathematical problem and write down directly the final results. Before that we modify further the notation for the reduced basis vectors (3.33), adding an extra row on it. We set

$$\begin{bmatrix} m_{12} & m_{22} & ; & m_{32} & m_{42} \\ & m_{11} & & m_{31} & \end{bmatrix}_i = \begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{12}, & m_{22}, & m_{32}, & m_{42} \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix}_i \equiv (m)_i . \tag{3.54}$$

The first row $m_{13},m_{23},m_{33},m_{43}$ in $(m)_i$ gives the coordinates of the gl(2/2) highest weight Λ [see (2.18)]. This row is one and the same for all reduced patterns in the induced module $W([m])$ and characterize the module itself. The second row indicates that $(m)_i$ is a vector from the $gl(2)_l \oplus gl(2)_r$ fidirmod $V_i([m_{12},m_{22},m_{32},m_{42}])$ in the decomposition (3.32); the numbers $m_{12},m_{22},m_{32},m_{42}$ are the coordinates of the highest weight of V_i. The numbers m_{11} and m_{31} label the basis vectors in V_i. We have added the zeros in $(m)_i$ only to fill in the empty space. In order to simplify the notation, observe that among all direct summands in (3.33) only the subspaces

$$V_3([m_{12} - 1, m_{22} - 1, m_{32} + 1, m_{42} + 1])$$

and

$$V_4([m_{12} - 1, m_{22} - 1, m_{32} + 1, m_{42} + 1]) \tag{3.55}$$

have one and the same highest weight. Any other $gl(2)_l \oplus gl(2)_r$ fidirmod $V_i([m_{12},m_{22},m_{32},m_{42}])$ is uniquely characterized by its highest weight $m_{12}e^1 + m_{22}e^2 + m_{32}e^3 + m_{42}e^4$. Therefore, in all these cases we skip the subscript i in V_i and also in the basis vectors $(m)_i$, writing simply $V([m_{12},m_{22},m_{32},m_{42}])$ and (m). The subscript will be kept only for the submodules (3.55) and their bases.

The transformation of the reduced basis under the action of the even generators is relatively simple and follows from (2.10)–(2.13). For completeness we write the expressions here also. Let $(m)_{\pm ij}$ be a pattern (3.54), obtained from (m) by replacing $m_{ij} \to m_{ij} \pm 1$. Then one has [see (2.14) for a definition of l_{ij}],

$$e_{11}(m) = m_{11}(m) , \quad e_{22}(m) = (m_{12} + m_{22} - m_{11})(m) , \tag{3.56}$$

$$e_{33}(m) = m_{31}(m) , \quad e_{44}(m) = (m_{32} + m_{42} - m_{31})(m) , \tag{3.57}$$

$$e_{12}(m) = |(l_{12} - l_{11})(l_{22} - l_{11})|^{1/2}(m)_{11} , \tag{3.58}$$

$$e_{21}(m) = |(l_{12} - l_{11} + 1)(l_{22} - l_{11} + 1)|^{1/2}(m)_{-11} , \tag{3.59}$$

$$e_{34}(m) = |(l_{32} - l_{31})(l_{42} - l_{31})|^{1/2}(m)_{31} , \tag{3.60}$$

$$e_{43}(m) = |(l_{32} - l_{31} + 1)(l_{42} - l_{31} + 1)|^{1/2}(m)_{-31} . \tag{3.61}$$

The transformations of $W([m])$ under the action of the odd generators are rather long and, in addition, we did not succeed in presenting them in a compact form. Therefore, we only give here the results for e_{32} and e_{23}.

Transformations under the action of e_{23},

$$e_{23}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}, m_{33}, m_{43} \\ m_{11}, 0, m_{31}, 0 \end{bmatrix} = 0, \tag{3.62}$$

$$e_{23}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}, m_{33}+1, m_{43} \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -(l_{13}+l_{33}+3)\left|\frac{(l_{13}-l_{11})(l_{43}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}, m_{33}, m_{43} \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}, \tag{3.63}$$

$$e_{23}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}, m_{33}, m_{43}+1 \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -(l_{13}+l_{43}+3)\left|\frac{(l_{13}-l_{11})(l_{33}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}, m_{33}, m_{43} \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}, \tag{3.64}$$

$$e_{23}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}-1, m_{33}+1, m_{43} \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -(l_{23}+l_{33}+3)\left|\frac{(l_{23}-l_{11})(l_{43}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}, m_{33}, m_{43} \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}, \tag{3.65}$$

$$e_{23}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}-1, m_{33}, m_{43}+1 \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -(l_{23}+l_{43}+3)\left|\frac{(l_{23}-l_{11})(l_{33}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}, m_{33}, m_{43} \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}, \tag{3.66}$$

$$e_{23}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}+2, m_{43} \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -(l_{13}+l_{33}+3)\left|\frac{(l_{13}-l_{11})(l_{43}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}-1, m_{33}+1, m_{43} \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}$$

$$+ (l_{23}+l_{33}+3)\left|\frac{(l_{23}-l_{11})(l_{43}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}, m_{33}+1, m_{43} \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}, \tag{3.67}$$

$$e_{23}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}, m_{43}+2 \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -(l_{13}+l_{43}+3)\left|\frac{(l_{13}-l_{11})(l_{33}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}-1, m_{33}, m_{43}+1 \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}$$

$$+ (l_{23}+l_{43}+3)\left|\frac{(l_{23}-l_{11})(l_{33}-l_{31}+1)}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}, m_{33}, m_{43}+1 \\ m_{11}, 0, m_{31}-1, 0 \end{bmatrix}, \tag{3.68}$$

$$
\begin{aligned}
e_{23}&\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}-1, & m_{33}+1, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix}_3 \\
&= -(l_{13}+l_{43}+2)\left|\frac{(l_{13}-l_{11})(l_{33}-l_{31}+2)(l_{33}-l_{43}-1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}+1)}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-1, & m_{33}+1, & m_{43} \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix} \\
&\quad + (l_{23}+l_{43}+2)\left|\frac{(l_{23}-l_{11})(l_{33}-l_{31}+2)(l_{33}-l_{43}-1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}+1)}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}, & m_{33}+1, & m_{43} \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix} \\
&\quad - (l_{13}+l_{33}+2)\left|\frac{(l_{13}-l_{11})(l_{43}-l_{31}+2)(l_{33}-l_{43}+1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}-1)}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-1, & m_{33}, & m_{43}+1 \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix} \\
&\quad + (l_{23}+l_{33}+2)\left|\frac{(l_{23}-l_{11})(l_{43}-l_{31}+2)(l_{33}-l_{43}+1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}-1)}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}, & m_{33}, & m_{43}+1 \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix},
\end{aligned} \tag{3.69}
$$

$$
\begin{aligned}
e_{23}&\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}, & m_{33}+1, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix} \\
&= -(l_{13}+l_{43}+3)\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{31}+2)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}, & m_{33}+1, & m_{43} \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix} \\
&\quad + (l_{13}+l_{33}+3)\left|\frac{(l_{13}-l_{11}-1)(l_{43}-l_{31}+2)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}, & m_{33}, & m_{43}+1 \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix},
\end{aligned} \tag{3.70}
$$

$$
\begin{aligned}
e_{23}&\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-2, & m_{33}+1, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix} \\
&= -(l_{23}+l_{43}+3)\left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{31}+2)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-1, & m_{33}+1, & m_{43} \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix} \\
&\quad + (l_{23}+l_{33}+3)\left|\frac{(l_{23}-l_{11}-1)(l_{43}-l_{31}+2)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-1, & m_{33}, & m_{43}+1 \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix},
\end{aligned} \tag{3.71}
$$

$$
\begin{aligned}
e_{23}&\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}-1, & m_{33}+1, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix}_4 \\
&= -(l_{13}+l_{43}+4)\left|\frac{(l_{13}-l_{11})(l_{13}-l_{23}-1)(l_{33}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-1, & m_{33}+1, & m_{43} \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix} \\
&\quad - (l_{23}+l_{43}+4)\left|\frac{(l_{23}-l_{11})(l_{13}-l_{23}+1)(l_{33}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}, & m_{33}+1, & m_{43} \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix} \\
&\quad + (l_{13}+l_{33}+4)\left|\frac{(l_{13}-l_{11})(l_{13}-l_{23}-1)(l_{43}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-1, & m_{33}, & m_{43}+1 \\ m_{11}, & 0, & m_{31}-1, & 0 \end{bmatrix}
\end{aligned}
$$

$$+ (l_{23}+l_{33}+4)\left|\frac{(l_{23}-l_{11})(l_{13}-l_{23}+1)(l_{43}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23} & ,m_{33} & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}, \tag{3.72}$$

$$e_{23}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-1 & ,m_{33}+2 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= -(l_{23}+l_{33}+3)\left|\frac{(l_{23}-l_{11})(l_{43}-l_{31}+2)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23} & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}$$

$$-(l_{13}+l_{43}+3)\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+2 & ,m_{43} \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}$$

$$+(l_{13}+l_{33}+2)\left|\frac{(l_{13}-l_{11}-1)(l_{13}-l_{23}+1)(l_{43}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}_4$$

$$+(l_{13}+l_{33}+4)\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{43}-1)(l_{43}-l_{31}+2)}{2(l_{13}-l_{23})(l_{33}-l_{43}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}_3, \tag{3.73}$$

$$e_{23}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= -(l_{23}+l_{43}+3)\left|\frac{(l_{23}-l_{11})(l_{33}-l_{31}+2)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23} & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}$$

$$+(l_{13}+l_{33}+3)\left|\frac{(l_{13}-l_{11}-1)(l_{43}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33} & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}$$

$$+(l_{13}+l_{43}+2)\left|\frac{(l_{13}-l_{11}-1)(l_{13}-l_{23}+1)(l_{33}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}_4$$

$$-(l_{13}+l_{43}+4)\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{43}+1)(l_{33}-l_{31}+2)}{2(l_{13}-l_{23})(l_{33}-l_{43}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}_3, \tag{3.74}$$

$$e_{23}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= +(l_{13}+l_{33}+3)\left|\frac{(l_{13}-l_{11})(l_{43}-l_{31}+2)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13} & ,m_{23}-2 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}$$

$$-(l_{23}+l_{43}+3)\left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+2 & ,m_{43} \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}$$

$$-(l_{23}+l_{33}+2)\left|\frac{(l_{23}-l_{11}-1)(l_{13}-l_{23}-1)(l_{43}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1 & ,0 \end{bmatrix}_4$$

$$+ (l_{23}+l_{33}+4)\left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{43}-1)(l_{43}-l_{31}+2)}{2(l_{13}-l_{23})(l_{33}-l_{43}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}_3 , \tag{3.75}$$

$$e_{23}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-2 & ,m_{33}+1 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= (l_{13}+l_{43}+3)\left|\frac{(l_{13}-l_{11})(l_{33}-l_{31}+2)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23} & ,m_{33} & ,m_{43} \\ m_{13}, & m_{23}-2 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11}, & 0 & ,m_{31}-1, & 0 \end{bmatrix}$$

$$+ (l_{23}+l_{33}+3)\left|\frac{(l_{23}-l_{11}-1)(l_{43}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33} & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}$$

$$- (l_{23}+l_{43}+2)\left|\frac{(l_{23}-l_{11}-1)(l_{13}-l_{23}-1)(l_{33}-l_{31}+2)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}_4$$

$$- (l_{23}+l_{43}+4)\left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{43}+1)(l_{33}-l_{31}+2)}{2(l_{13}-l_{23})(l_{33}-l_{43}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}_3 , \tag{3.76}$$

$$e_{23}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= (l_{23}+l_{43}+3)\left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-1 & ,m_{33}+2 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}$$

$$- (l_{23}+l_{33}+3)\left|\frac{(l_{23}-l_{11}-1)(l_{43}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}$$

$$- (l_{13}+l_{43}+3)\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}$$

$$- (l_{13}+l_{33}+3)\left|\frac{(l_{13}-l_{11}-1)(l_{43}-l_{31}+3)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-2 & ,m_{33}+1 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31}-1, & 0 \end{bmatrix}. \tag{3.77}$$

Transformations under the action of e_{32},

$$e_{32}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{11}, & 0 & ,m_{31}, & 0 \end{bmatrix} = -\left|\frac{(l_{13}-l_{11})(l_{43}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23}, & m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}, & m_{33}+1 & ,m_{43} \\ m_{11} & ,0 & ,m_{31}+1, & 0 \end{bmatrix}$$

$$-\left|\frac{(l_{13}-l_{11})(l_{33}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23}, & m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}, & m_{33} & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31}+1, & 0 \end{bmatrix}$$

$$-\left|\frac{(l_{23}-l_{11})(l_{43}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23} & ,m_{33} & ,m_{43} \\ m_{13}, & m_{23}-1 & ,m_{33}+1 & ,m_{43} \\ m_{11}, & 0 & ,m_{31}+1, & 0 \end{bmatrix}$$

$$-\left|\frac{(l_{23}-l_{11})(l_{33}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23} & ,m_{33} & ,m_{43} \\ m_{13}, & m_{23}-1 & ,m_{33} & ,m_{43}+1 \\ m_{11}, & 0 & ,m_{31}+1, & 0 \end{bmatrix}, \tag{3.78}$$

$$
\begin{aligned}
e_{32}\begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23} & , m_{33}+1 & , m_{43} \\ m_{11} & , 0 & , m_{31} & , 0 \end{bmatrix}
&= \left|\frac{(l_{23}-l_{11})(l_{43}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+2 & , m_{43} \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix} \\
&\quad - \left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{31}+1)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-2 & , m_{23} & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix} \\
&\quad + \left|\frac{(l_{23}-l_{11})(l_{33}-l_{31}+1)(l_{33}-l_{43}-1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix}_3 \\
&\quad - \left|\frac{(l_{23}-l_{11})(l_{13}-l_{23}+1)(l_{33}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix}_4 ,
\end{aligned}
\tag{3.79}
$$

$$
\begin{aligned}
e_{32}\begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23} & , m_{33} & , m_{43}+1 \\ m_{11} & , 0 & , m_{31} & , 0 \end{bmatrix}
&= \left|\frac{(l_{23}-l_{11})(l_{33}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33} & , m_{43}+2 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix} \\
&\quad + \left|\frac{(l_{13}-l_{11}-1)(l_{43}-l_{31}+1)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-2 & , m_{23} & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix} \\
&\quad + \left|\frac{(l_{23}-l_{11})(l_{43}-l_{31}+1)(l_{33}-l_{43}+1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}-1)}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix}_3 \\
&\quad + \left|\frac{(l_{23}-l_{11})(l_{13}-l_{23}+1)(l_{43}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix}_4 ,
\end{aligned}
\tag{3.80}
$$

$$
\begin{aligned}
e_{32}\begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13} & , m_{23}-1 & , m_{33}+1 & , m_{43} \\ m_{11} & , 0 & , m_{31} & , 0 \end{bmatrix}
&= -\left|\frac{(l_{13}-l_{11})(l_{43}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+2 & , m_{43} \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix} \\
&\quad - \left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{31}+1)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13} & , m_{23}-2 & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix} \\
&\quad - \left|\frac{(l_{13}-l_{11})(l_{33}-l_{31}+1)(l_{33}-l_{43}-1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix}_3 \\
&\quad - \left|\frac{(l_{13}-l_{11})(l_{13}-l_{23}-1)(l_{33}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13} & , m_{23} & , m_{33} & , m_{43} \\ m_{13}-1 & , m_{23}-1 & , m_{33}+1 & , m_{43}+1 \\ m_{11} & , 0 & , m_{31}+1 & , 0 \end{bmatrix}_4 ,
\end{aligned}
\tag{3.81}
$$

$$e_{32}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}-1, m_{33}, m_{43}+1 \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -\left|\frac{(l_{13}-l_{11})(l_{33}-l_{31})}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}, m_{43}+2 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}$$

$$+\left|\frac{(l_{23}-l_{11}-1)(l_{43}-l_{31}+1)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}, m_{23}-2, m_{33}+1, m_{43}+1 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}$$

$$-\left|\frac{(l_{13}-l_{11})(l_{43}-l_{31}+1)(l_{33}-l_{43}+1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}-1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}+1, m_{43}+1 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}_3$$

$$+\left|\frac{(l_{13}-l_{11})(l_{13}-l_{23}-1)(l_{43}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}+1, m_{43}+1 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}_4, \tag{3.82}$$

$$e_{32}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}+2, m_{43} \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= -\left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-2, m_{33}+2, m_{43}+1 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}$$

$$-\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-2, m_{23}-1, m_{33}+2, m_{43}+1 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}, \tag{3.83}$$

$$e_{32}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}, m_{43}+2 \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}$$

$$= \left|\frac{(l_{23}-l_{11}-1)(l_{43}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-2, m_{33}+1, m_{43}+2 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}$$

$$+\left|\frac{(l_{13}-l_{11}-1)(l_{43}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43}-1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-2, m_{23}-1, m_{33}+1, m_{43}+2 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}, \tag{3.84}$$

$$e_{32}\begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-1, m_{33}+1, m_{43}+1 \\ m_{11}, 0, m_{31}, 0 \end{bmatrix}_3$$

$$= \left|\frac{(l_{23}-l_{11}-1)(l_{43}-l_{31}+1)(l_{33}-l_{43}-1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-2, m_{33}+2, m_{43}+1 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}$$

$$+\left|\frac{(l_{13}-l_{11}-1)(l_{43}-l_{31}+1)(l_{33}-l_{43}-1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}+1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-2, m_{23}-1, m_{33}+2, m_{43}+1 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}$$

$$-\left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{31}+1)(l_{33}-l_{43}+1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}-1)}\right|^{1/2} \begin{bmatrix} m_{13}, m_{23}, m_{33}, m_{43} \\ m_{13}-1, m_{23}-2, m_{33}+1, m_{43}+2 \\ m_{11}, 0, m_{31}+1, 0 \end{bmatrix}$$

$$-\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{31}+1)(l_{33}-l_{43}+1)}{2(l_{13}-l_{23})(l_{33}-l_{43})(l_{33}-l_{43}-1)}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}-1, & m_{33}+1, & m_{43}+2 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}, \tag{3.85}$$

$$e_{32}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}, & m_{33}+1, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix}$$

$$= -\left|\frac{(l_{23}-l_{11})(l_{43}-l_{31}+1)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}-1, & m_{33}+2, & m_{43}+1 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}$$

$$-\left|\frac{(l_{23}-l_{11})(l_{33}-l_{31}+1)}{(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}-1, & m_{33}+1, & m_{43}+2 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}, \tag{3.86}$$

$$e_{32}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}, & m_{23}-2, & m_{33}+1, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix}$$

$$= \left|\frac{(l_{13}-l_{11})(l_{43}-l_{31}+1)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}-2, & m_{33}+2, & m_{43}+1 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}$$

$$+\left|\frac{(l_{13}-l_{11})(l_{33}-l_{31}+1)}{(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}-2, & m_{33}+1, & m_{43}+2 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}, \tag{3.87}$$

$$e_{32}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}-1, & m_{33}+1, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix}_4$$

$$= -\left|\frac{(l_{23}-l_{11}-1)(l_{13}-l_{23}-1)(l_{43}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}-2, & m_{33}+2, & m_{43}+1 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}$$

$$+\left|\frac{(l_{13}-l_{11}-1)(l_{13}-l_{23}+1)(l_{43}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}-1, & m_{33}+2, & m_{43}+1 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}$$

$$-\left|\frac{(l_{23}-l_{11}-1)(l_{13}-l_{23}-1)(l_{33}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}+1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-1, & m_{23}-2, & m_{33}+1, & m_{43}+2 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}$$

$$+\left|\frac{(l_{13}-l_{11}-1)(l_{13}-l_{23}+1)(l_{33}-l_{31}+1)}{2(l_{13}-l_{23})(l_{13}-l_{23}-1)(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}-1, & m_{33}+1, & m_{43}+2 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}, \tag{3.88}$$

$$e_{32}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}-1, & m_{33}+2, & m_{43}+1 \\ m_{11}, & 0, & m_{31}, & 0 \end{bmatrix}$$

$$= \left|\frac{(l_{23}-l_{11}-1)(l_{33}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13}, & m_{23}, & m_{33}, & m_{43} \\ m_{13}-2, & m_{23}-2, & m_{33}+2, & m_{43}+2 \\ m_{11}, & 0, & m_{31}+1, & 0 \end{bmatrix}, \tag{3.89}$$

$$e_{32}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-1 & ,m_{33}+1 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= -\left|\frac{(l_{23}-l_{11}-1)(l_{43}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31}+1 & ,0 \end{bmatrix}, \tag{3.90}$$

$$e_{32}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+1 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= -\left|\frac{(l_{13}-l_{11}-1)(l_{33}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31}+1 & ,0 \end{bmatrix}, \tag{3.91}$$

$$e_{32}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-1 & ,m_{23}-2 & ,m_{33}+1 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix}$$

$$= \left|\frac{(l_{13}-l_{11}-1)(l_{43}-l_{31}+2)}{(l_{13}-l_{23})(l_{33}-l_{43})}\right|^{1/2}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31}+1 & ,0 \end{bmatrix}, \tag{3.92}$$

$$e_{32}\begin{bmatrix} m_{13} & ,m_{23} & ,m_{33} & ,m_{43} \\ m_{13}-2 & ,m_{23}-2 & ,m_{33}+2 & ,m_{43}+2 \\ m_{11} & ,0 & ,m_{31} & ,0 \end{bmatrix} = 0. \tag{3.93}$$

The transformations (3.62)–(3.93) together with the expressions (3.56)–(3.61) for the even generators and the supercommutation relations (1.1) uniquely determine all other odd generators. In this sense, the above written relations define the representation of the Lie superalgebra gl(2/2).

If the signature $[m_{13},m_{23},m_{33},m_{43}]$, $m_{13}-m_{23}\in\mathbb{Z}_+$, $m_{33}-m_{43}\in\mathbb{Z}_+$, is such that $l_{i3}+l_{j3}+3\neq 0$, $\forall i=1,2$ and $j=3,4$, then the induced gl(2/2) module $W([m])$ is irreducible (Proposition 2) and hence typical. Letting the signature take all possible values, compatible with (2.43), one obtains all typical representations. If the condition (2.43) is not fulfilled, then the induced module is indecomposible. The nontypical irreducible representation is realized in a properly chosen factor space of $W([m])$. This case will be considered in Ref. 1.

[1]N. I. Ilieva and T. D. Palev, "Finite-dimensional representations of the Lie superalgebra gl(2/2) in a gl(2) ⊕ gl(2) basis. II. Nontypical representations" (to be published).

[2]V. G. Kac, Adv. Math. **26**, 8 (1977).

[3]D. Ž. Djoković and G. Hochschild, Ill. J. Math. **20**, 134 (1976).

[4]V. G. Kac, *Lecture Notes in Mathematics*, Vol. 626 (Springer, Berlin, 1978), p. 597.

[5]V. Rittenberg, *Lecture Notes in Physics*, Vol. 79 (Springer, Berlin, 1978); M. Scheunert, *Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979); D. A. Leites, Sov. Prob. Math. **25**, 3 (1984) (in Russian), and references therein.

[6]L. Ross, Trans. Am. Math. Soc. **120**, 17 (1965); A Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975); G. Hochschild, Ill. J. Math. **20**, 107 (1976); D. Ž. Djoković, J. Pure Appl. Algebra **9**, 25 (1976); F. A. Berezin, Sov. J. Nucl. Phys. **29**, 857 (1979); **30**, 605 (1979); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977); M. Marcu, *ibid.* **21**, 1277 (1980); T. D. Palev, *ibid.* **21**, 1293 (1980); T. D. Palev and O. Ts. Stoytchev, C. R. Acad. Bulg. Sci. **35**, 733 (1982); J.-P. Hurni and B. Morel, J. Math. Phys. **23**, 2236 (1982); **24**, 157 (1983); B. Gruber, T. S. Santhanam, and R. Wilson, *ibid.* **25**, 1253 (1984); J. Thierry-Mieg, Phys. Lett. B **138**, 393 (1984); J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).

[7]P. H. Dondi and P. D. Jarvis, Z. Phys. C **4**, 201 (1980); J. Phys. A **14**, 547 (1981); A. H. Balantekin and I. Bars, J. Math. Phys. **22**, 1149, 1810 (1981); **23**, 1239 (1982); A. B. Balantekin, *ibid.* **23**, 486 (1982); I. Bars, B. Morel, and H. Ruegg, *ibid.* **24**, 2253 (1983); F. Delduc and M. Gourdin, *ibid.* **25**, 1651 (1984); **26**, 1865 (1985); I. Bars, Physica D **15**, 42 (1985).

[8]T. D. Palev, J. Math. Phys. **26**, 1640 (1985); **27**, 1994 (1986); **28**, 272, 1987; **28**, 2280 (1987); Funct. Anal. Appl. **21**, 85 (1987) (in Russian).

[9]T. D. Palev, "Essentially typical representations of the Lie superalgebras sl(n/1) and sl(n/2)," in Proceedings of the XVI International Colloquium on Group Theoretical Methods in Physics, Varna (Bulgaria), 1987 (to be published).

[10]T. D. Palev, "Essentially typical representations of the Lie superalgebras gl(n,m) in a Gel'fand–Zetlin basis," Preprint JINR P5-88-169 (1988) (in Russian).

[11]I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950) (in Russian); see also G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).

[12]A useful prescription for decomposition of a tensor product of two gl(n) fidirmods is given in J. D. Louck, Am. J. Phys. **38**, 18 (1970).

[13]L. C. Biedenharn and J. D. Louck, "Angular momentum in quantum physics," in *Encyclopedia of Mathematics and its Applications*, edited by G.-C. Rota (Addison-Wesley, Reading, MA, 1981), Vol. 8 (see Table I in the Appendix).

Lie algebra automorphisms, the Weyl group, and tables of shift vectors

Peter Bouwknegt
Institute for Theoretical Physics, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

(Received 25 May 1988; accepted for publication 26 October 1988)

The canonical lift of the Weyl group in the automorphism group of a simple Lie algebra is discussed. A constructive algorithm is given to compute the shift vector for every Weyl group conjugacy class, and tables for the simply laced Lie algebras of rank $\leqslant 8$ are provided. Applications in physics, such as the equivalence of certain orbifold and twisted WZW string models, as well as the implication for the moduli space of 2-D conformal field theories are discussed.

I. INTRODUCTION

String theory has been a motivation for a lot of mathematical research. As a result, our understanding of (infinite-dimensional) algebras has rapidly increased. A nice example is the Frenkel–Kač–Segal vertex-operator construction, providing a realization of level-1 highest weight modules for simply laced Kač–Moody algebras.[1,2] This construction is based on the vertex operators used in string theories some time ago (see, e.g., Ref. 3 and references therein). After this construction it was realized that it implies the equivalence of two *a priori* different closed string models, namely a closed bosonic string propagating on the maximal torus of some Lie algebra g and a closed bosonic string moving on the group manifold of g. These are known as the torus compactified string and the level-1 Wess–Zumino–Witten (WZW) model,[4] respectively. Since then there have been a lot of extensions of these results to fermionic strings, higher level WZW models, etc. From many points of view a very interesting extension of the torus and WZW models are the orbifold and twisted WZW models, respectively.

In an orbifold compactification, one considers the propagation of a string on the space obtained by identifying, in flat space, points in the orbit of a point group. In particular, one can consider maximal tori of Lie algebras g, where one divides out subgroups of the automorphism group $\mathrm{Aut}(\Lambda_R)$ of the root lattice Λ_R of g. If one divides out the finite Abelian subgroup generated by a single element $w \in \mathrm{Aut}(\Lambda_R)$, one obtains what is called an Abelian orbifold. It has been shown that to every such orbifold there corresponds a "twisted" realization of a level-1 Kač–Moody highest weight module, equivalent to the torus realization iff $w \in \mathscr{W}(g)$, the Weyl group of g, and corresponding to a twisted affine Kač–Moody algebra realization if $w \in \mathrm{Aut}(\Lambda_R)/\mathscr{W}(g)$ (Refs. 5 and 6, see also Refs. 7 and 8).

On the other hand, one can obtain "twisted" realizations of Kač–Moody algebras by twisting the boundary conditions of a WZW model with an automorphism σ of the corresponding Lie algebra g. These twisted WZW models are conveniently described by a so-called shift vector γ in the dual Cartan subalgebra h^* of g.

Because the number of conjugacy classes in $\mathrm{Aut}(g)$ is infinite and $\mathrm{Aut}(\Lambda_R)$ is a finite group, there cannot be a 1–1 correspondence between the orbifold and twisted WZW models. It has been shown, however, that every orbifold model based on a $w \in \mathrm{Aut}(\Lambda_R)$ corresponds to a twisted WZW model,[9] obtained by lifting w into $\mathrm{Aut}(g)$, though the method employed did not immediately give the corresponding shift vector explicitly.

There have been several papers dealing with the determination of the shift vectors.[7,10,11] The methods that were used can be described as determination by exhaustion. All automorphisms $\sigma \in \mathrm{Aut}(g)$ of a certain finite order were determined, and some of their properties were compared. The one matching the properties of the element $w \in \mathrm{Aut}(\Lambda_R)$ was then singled out.

Though this "brute force" method works well it does not give any insight in the underlying mathematical structure, and moreover, for high rank groups and high-order automorphisms, it will simply take too much computer time.

In this paper we will discuss an algorithm that allows one to compute the shift vectors directly, using properties of the automorphisms in regular subalgebras. We provide tables for the shift vectors corresponding to all $w \in \mathscr{W}(g)$ for all simply laced simple Lie algebras of rank $\leqslant 8$. These tables are very interesting from a mathematical point of view because they provide explicit forms of certain generalized theta function identities. From a physical point of view these identities correspond to the two equivalent ways of computing the string partition function—by means of the orbifold and the twisted WZW model. Another physical application is that the equivalence provides information on the moduli space of two-dimensional conformal field theories (CFT's), of integer central charge c. One branch of this moduli space consists of torus compactified scalar field theories,[12] and another branch consists of all possible orbifold versions of these scalar field theories.[13] The equivalence discussed above gives the intersection points of these two branches (see Ref. 14 for $c = 1$, and Ref. 15 for $c \geqslant 2$).

The paper is organized as follows. In Secs. II and III we review the classification of conjugacy classes in the Weyl group $\mathscr{W}(g)$ and the automorphism group $\mathrm{Aut}(g)$ and introduce some notation and useful concepts. In Sec. IV we discuss the lift of $\mathscr{W}(g)$ to a subgroup $\widetilde{\mathscr{W}}$ of $\mathrm{Aut}(g)$, and explain the algorithm to obtain the shift vectors. Tables for rank $\leqslant 8$ simply laced Lie algebras are computed. Section V deals with the applications. We briefly review the definition of the orbifold and twisted WZW models and discuss the partition functions as well as their modular invariance. We point out some properties of the $\sigma \in \mathrm{Aut}(g)$ corresponding to a lifted Weyl group element and make some remarks on the

 0022-2488/89/030571-14$02.50

string models and implications for the moduli space of CFT's of integer c. In the Appendix we point out an apparent relation between conjugacy clases in $\mathscr{W}(g)$ and equivalence classes of A_1 subalgebras of g, and speculate on its relevance in the construction of extended conformal algebras.

II. THE WEYL GROUP

In the following let g be a simple, finite-dimensional Lie algebras over the complex numbers $\mathbf{C}$, and let h be its Cartan subalgebra (CSA). To every root $\beta\in\Delta$ of g one can associate a reflection r_β: $h^*\to h^*$ by

$$r_\beta(\lambda) = \lambda - [2(\beta,\lambda)/(\beta,\beta)]\beta, \quad \lambda\in h^* . \tag{2.1}$$

The Weyl group $\mathscr{W}(g)$ of g is the finite group generated by all reflections r_β, $\beta\in\Delta$. Of course, $\mathscr{W}$ is already generated by the reflections r_i in the simple roots α_i, $i = 1,...,l$, of g.

By dualizing, one can also consider the action of $\mathscr{W}$ on h. Explicitly,

$$r_\beta(\mathbf{h}) = \mathbf{h} - \langle\beta,\mathbf{h}\rangle\beta^\vee, \quad \mathbf{h}\in h . \tag{2.2}$$

By definition, every Weyl group element $w\in\mathscr{W}$ can be decomposed as

$$w = r_{\beta_1} r_{\beta_2}\cdots r_{\beta_k} , \quad \beta_i\in\Delta . \tag{2.3}$$

The expression (2.3) is called reduced if k is the least possible value for this w. This value of k is called the length $\bar{l}(w)$ of w. One can show that (2.3) is reduced iff $\beta_1\cdots\beta_k$ are linearly independent. Moreover, $\bar{l}(w)$ is the number of eigenvalues of w distinct from 1, so, in particular, $\bar{l}(w)\leqslant l$.

An element $w\in\mathscr{W}$ is called *nondegenerate* if $\det(1-w)\neq 0$ [i.e., if $\bar{l}(w) = l$]. Among the nondegenerate elements there are a few which play an important role, the so-called *primitive* elements, defined by the condition $\det(1-w) = \det(a)$, where a is the Cartan matrix of g.

Every $w\in\mathscr{W}$ is expressible in the form[16]

$$w = w_1 w_2 = \Big(\prod_{i\in A} r_{\beta_i}\Big)\Big(\prod_{j\in B} r_{\beta_j}\Big), \tag{2.4}$$

where the roots are divided into two disjoint subsets A and B such that each of them contains only roots that are mutually orthogonal. [Notice that $\bar{l}(w) = \bar{l}(w_1) + \bar{l}(w_2)$ and that w_1 and w_2 are both involutions.]

One associates a Dynkin-like graph Γ to each decomposition (2.4) as follows. A vertex i stands for a root β_i and the number of links between two vertices is given by

$$4(\beta_i,\beta_j)^2/(\beta_i,\beta_i)(\beta_j,\beta_j) . \tag{2.5}$$

An arrow points to the smallest of the two roots.

It is clear that two elements of $\mathscr{W}$ that are conjugated give rise to the same graph Γ. Conversely it is *not* true that Γ uniquely determines the conjugacy class. The correspondence is, however, so close that it is convenient to label the conjugacy classes of $\mathscr{W}$ by the Dynkin-like diagrams Γ. The exceptions as well as the determination of the admissible graphs Γ for given $\mathscr{W}(g)$ can be found in Ref. 16. It turns out that all the admissible graphs of $\mathscr{W}(g)$ correspond to conjugacy classes of primitive elements in regular subalgebras of g. Because the regular subalgebras of g are easily found from the extended Dynkin diagram of g (Ref. 17) this property is very useful in the application we are about to discuss. The number of conjugacy classes of primitive elements for A_n, B_n, C_n, D_n, G_2, F_4, E_6, E_7, and E_8 are, respectively, 1, 1, 1, $[n/2]$, 1, 2, 3, 5, and 9. Remarkably, these numbers also show up as the number of orbits of the extended Dynkin diagram of g under its isometry group,[5] and also as the number of possible values of the defect c_w, to be discussed in Sec. V.

The graphs corresponding to the primitive elements can be found in Refs. 11 and 16. Table I lists some properties of the corresponding Weyl group conjugacy classes. The characteristic polynomial $P_w(t)\equiv\det_{h^*}(1-tw)$ for any Weyl group element can be found from Table I.

Because the Weyl group elements satisfy

$$(w\lambda_1,w\lambda_2) = (\lambda_1,\lambda_2), \quad \forall\lambda_1,\lambda_2\in h^* , \tag{2.6}$$

i.e., they are unitary with respect to the Cartan–Killing form on h^*, the eigenvalues are of the form ϵ^{m_i}, where $\epsilon = \exp(2\pi i/N)$ and N is the order of the Weyl group element. The set $\{m_i\}$ for the primitive elements is given in Table I. We have an orthogonal decomposition of h^* (and h)

$$h^* = \bigoplus_{j=0}^{N-1} h^*_{(j)} \tag{2.7}$$

in terms of eigenspaces $h^*_{(j)}\equiv\{\lambda\in h^*|w\lambda = \epsilon^j\lambda\}$, and we denote by $d_j = \dim h^*_{(j)}$ the multiplicity of the eigenvalue ϵ^j.

There is a particular primitive conjugacy class that always plays a preferred role. This is the conjugacy class of the so-called *Coxeter* element

$$w_C = \prod_{i=1}^{l} r_i , \tag{2.8}$$

which is a product over the Weyl reflections in the simple roots of g. The Coxeter element is the element of highest order in $\mathscr{W}(g)$, the order h is called the Coxeter number of g. The graph Γ of w_C is just the Dynkin diagram of g, and the set $\{m_i\}$ corresponding to the eigenvalues of w_C are in this case called the *exponents* of g.

III. AUTOMORPHISMS OF SIMPLE LIE ALGEBRAS

In this section we will review the classification of automorphisms of finite order of simple, finite-dimensional Lie algebras, as obtained by Kač.[18] Suppose we are given a simple, finite-dimensional Lie algebra g over the complex numbers $\mathbf{C}$. An automorphism of g is an invertible linear transformation σ: $g\to g$, satisfying

$$\begin{aligned} &\sigma[x,y] = [\sigma x,\sigma y] ,\\ &(\sigma x,\sigma y) = (x,y) , \quad \forall x,y\in g , \end{aligned} \tag{3.1}$$

where $(x,y)\sim\mathrm{Tr}(\mathrm{ad}\,x\,\mathrm{ad}\,y)$ is the Cartan–Killing form on g. The set of automorphisms of finite order (i.e., $\sigma^N = 1$ for some $N\in\mathbf{N}$) from a group $\mathrm{Aut}(g)$. It is easy to give examples of automorphisms. Let $x\in g$ and define $\sigma_x\in\mathrm{Aut}(g)$ by

$$\sigma_x = \exp(\mathrm{ad}\,x) = \mathrm{Ad}(\exp x) , \tag{3.2}$$

then (3.1) is a consequence of the Jacobi identity. The set of automorphisms that can be written as σ_x for some $x\in g$ is denoted by $\mathrm{Aut}_0(g)$ and are called inner or invariant automorphisms. In fact, the inner automorphisms almost exhaust the set of automorphisms in the sense that $\mathrm{Aut}_0(g)$ is a

TABLE I. Properties of primitive Weyl group elements. The eigenvalues of w are given as ϵ^{m_i}, $\epsilon = \exp(2\pi i/N)$, where N is the order of w.

Γ	Ord w	$\det(1 - tw)$	m_i
A_n	$n+1$	$t^n + t^{n-1} + \cdots + 1$	$1,2,\ldots,n$
B_n, C_n	$2n$	$t^n + 1$	$1,3,5,\ldots,2n-1$
D_n	$2(n-1)$	$(t^{n-1}+1)(t+1)$	$1,3,5,\ldots,2n-3,n-1$
$D_n(a_1)$	$2\,\mathrm{lcm}(2,n-2)$	$(t^{n-2}+1)(t^2+1)$	$\vdots$
$D_n(a_2)$	$3\,\mathrm{lcm}(3,n-3)$	$(t^{n-3}+1)(t^3+1)$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$D_n(a_{(1/2)n-1})$	n	$(t^{n/2}+1)^2$	$1,1,3,3,\ldots,n-1,n-1$
G_2	6	$t^2 - t + 1$	1,5
F_4	12	$t^4 - t^2 + 1$	1,5,7,11
$F_4(a_1)$	6	$(t^2 - t + 1)^2$	1,1,5,5
E_6	12	$(t^4 - t^2 + 1)(t^2 + t + 1)$	1,4,5,7,8,11
$E_6(a_1)$	9	$(t^6 + t^3 + 1)$	1,2,4,5,7,8
$E_6(a_2)$	6	$(t^2 - t + 1)^2(t^2 + t + 1)$	1,1,2,4,5,5
E_7	18	$(t^6 - t^3 + 1)(t + 1)$	1,5,7,9,11,13,17
$E_7(a_1)$	14	$(t^6 - t^5 + t^4 - t^3 + t^2 - t + 1)(t + 1)$	1,3,5,7,9,11,13
$E_7(a_2)$	12	$(t^4 - t^2 + 1)(t^2 - t + 1)(t + 1)$	1,2,5,6,7,10,11
$E_7(a_3)$	30	$(t^4 - t^3 + t^2 - t + 1)(t^2 - t + 1)(t + 1)$	3,5,9,15,21,25,27
$E_7(a_4)$	6	$(t^2 - t + 1)^3(t + 1)$	1,1,1,3,5,5,5
E_8	30	$t^8 + t^7 - t^5 - t^4 - t^3 + t + 1$	1,7,11,13,17,19,23,29
$E_8(a_1)$	24	$t^8 - t^4 + 1$	1,5,7,11,13,17,19,23
$E_8(a_2)$	20	$t^8 - t^6 + t^4 - t^2 + 1$	1,3,7,9,11,13,17,19
$E_8(a_3)$	12	$(t^4 - t^2 + 1)^2$	1,1,5,5,7,7,11,11
$E_8(a_4)$	18	$(t^6 - t^3 + 1)(t^2 - t + 1)$	1,3,5,7,11,13,15,17
$E_8(a_5)$	15	$t^8 - t^7 + t^5 - t^4 + t^3 - t + 1$	1,2,4,7,8,11,13,14
$E_8(a_6)$	10	$(t^4 - t^3 + t^2 - t + 1)^2$	1,1,3,3,7,7,9,9
$E_8(a_7)$	12	$(t^4 - t^2 + 1)(t^2 - t + 1)^2$	1,2,2,5,7,10,10,11
$E_8(a_8)$	6	$(t^2 - t + 1)^4$	1,1,1,1,5,5,5,5

normal subgroup of $\mathrm{Aut}(g)$ and $\mathrm{Aut}(g)/\mathrm{Aut}_0(g)$ is a finite Abelian group, isomorphic to the symmetry group of the Dynkin diagram of g. These will be called outer automorphisms.

The classification of conjugacy classes in $\mathrm{Aut}(g)$ is most easily described in the context of affine Kač-Moody algebras.[19] Let $g^{(k)}$ be an affine Kač–Moody algebra, $g^{(k)}_{\mathrm{hor}}$ its finite-dimensional horizontal Lie algebra. (For the untwisted affine Kač–Moody algebras $g^{(1)}$, we have $g^{(1)}_{\mathrm{hor}} = g$.) Let l be the rank of $g^{(k)}_{\mathrm{hor}}$. To every $g^{(k)}$ and $l+1$-tuple $s = (s_0,\ldots,s_l)$ of relatively prime non-negative integers (with the convention that 0 is relatively prime to all other integers), one can associate an automorphism $\sigma_{s,k}$ of g of order

$$N = k \sum_{i=0}^{l} a_i^{\vee} s_i \,, \tag{3.3}$$

where $\{a_i^{\vee}\ i = 0,\ldots,l\}$ is a left zero eigenvector of the Cartan matrix of $g^{(k)}$, normalized such that $a_0^{\vee} = 1$. The automorphism is uniquely defined through

$$\begin{aligned} &\sigma_{s,k}(H_i) = H_i \,, \\ &\sigma_{s,k}(E_{\alpha_i}) = \epsilon^{s_i} E_{\alpha_i} \,, \quad \epsilon = \exp(2\pi i/N)\,, \quad i = 0,\ldots,l\,, \end{aligned} \tag{3.4}$$

where H_i and E_{α_i} are, respectively, a basis for the CSA and dual simple root system (i.e., step operators) of $g^{(k)}_{\mathrm{hor}}$.

The classification theorem of Kač[18] can be formulated as follows.

(i) Every $\sigma \in \mathrm{Aut}(g)$ is conjugated to an automorphism $\sigma_{s,k}$, defined by (3.4), where $s = (s_0,\ldots,s_l)$ is an $(l+1)$-tuple of relatively prime non-negative integers, $k \in \{1,2,3\}$ such that $g^{(k)}$ is an affine Kač–Moody algebra, and $l = \mathrm{rank}\,(g^{(k)}_{\mathrm{hor}})$.

(ii) $\sigma_{s,k}$ and $\sigma_{s',k'}$ are conjugated if and only if $k = k'$ and s and s' are related by an automorphism of the Dynkin diagram of $g^{(k)}$.

(iii) k is the least positive integer for which $(\sigma_{s,k})^k \in \mathrm{Aut}_0(g)$.

For simplicity we will henceforth restrict ourselves to inner automorphisms $\sigma_s \equiv \sigma_{s,k=1} \in \mathrm{Aut}_0(g)$, though most of the issues we will consider can straightforwardly be generalized. To every automorphism $\sigma_s \in \mathrm{Aut}_0(g)$ of order N we associate *a shift vector* $\gamma_s \in h^*$ defined by the inner products

$$(\gamma_s,\alpha_i) = s_i/N\,, \quad i = 1,\ldots,l\,, \tag{3.5}$$

or equivalently

$$\gamma_s = \frac{1}{N} \sum_{i=1}^{l} \frac{2}{(\alpha_i,\alpha_i)}\, s_i \Lambda_i \,, \tag{3.6}$$

in terms of the fundamental weights Λ_i of g. The reason why γ_s is called a shift vector will become clear in Sec. V, where we discuss the $g^{(1)}$ level-1 character in a specialization given by σ_s. By using the shift vector, the automorphism on every step operator E_α, $\alpha \in \Delta$, can be given as

$$\sigma_s(E_\alpha) = \exp[2\pi i(\gamma_s,\alpha)]\,E_\alpha \,. \tag{3.7}$$

Every automorphism $\sigma \in \mathrm{Aut}_0(g)$ of order N gives an orthogonal decomposition

$$g = \bigoplus_{j=0}^{N-1} g_{(j)} \,, \tag{3.8}$$

into eigenspaces $g_{(j)}$ belonging to the eigenvalue ϵ^j, where $\epsilon = \exp(2\pi i/N)$. Let $d_j = \dim g_{(j)}$ denote the multiplicity of the eigenvalue, ϵ^j. If σ_s is the Kač automorphism conjugated to σ and γ_s is its corresponding shift vector then we explicitly have

$$d_j = \#\{\alpha\in\Delta \mid (\gamma_s,\alpha) = (j/N) \bmod 1\} + l\delta_{j,0} \,. \quad (3.9)$$

In terms of the multiplicities d_j, we can calculate, for instance, the characteristic polynomial

$$P_\sigma(t) \equiv \det_g(1 - t\sigma) = \prod_{j=0}^{N-1} (1 - t\epsilon^j)^{d_j}, \quad (3.10)$$

but they also appear in the very strange formula[19,20]

$$\frac{1}{2h^\vee}|\rho - h^\vee \gamma_s|^2 = \frac{\dim g}{24} - \frac{1}{4N^2}\sum_{j=0}^{N-1} j(N-j)d_j\,, \quad (3.11)$$

where

$$h^\vee = \sum_{i=0}^{l} a_i^\vee \quad (3.12)$$

is the dual Coxeter number of g, and $\rho\in h^*$ the *principal* vector

$$\rho = \frac{1}{2}\sum_{\alpha\in\Delta_+}\alpha \quad (3.13)$$

satisfying

$$2(\rho,\alpha_i)/(\alpha_i,\alpha_i) = 1, \quad \forall \text{ simple roots } \alpha_i\,. \quad (3.14)$$

The well-known Freudenthal–de Vries strange formula[21] appears as the $\sigma = 1$ case of (3.11). The advantage of describing an automorphism by means of an $(l+1)$-tuple s is that a lot of properties of σ_s can immediately be read off from s. We have for instance an explicit description of the invariant subalgebra $g_{(0)}$. It is the regular subalgebra of g given by the Dynkin diagram obtained by deleting all the vertices with $s_i \neq 0$ from the extended Dynkin diagram of g. To this, one has to add $u(1)$ factors such that rank $(g_{(0)}) = l$. An explicit description can also be given for the $g_{(0)}$ modules $g_{(1)}$ and $g_{(N-1)}$.[19]

So far we have been concerned with $(l+1)$-tuples $s = (s_0,\ldots,s_l)$ of relatively prime *non-negative* integers w.r.t. some chosen simple root basis $\Pi = \{\alpha_1,\ldots,\alpha_l\}$. Let us call such an $(l+1)$tuple *special* and the corresponding automorphism by a Kač automorphism. It is clear that very relatively prime $(l+1)$-tuple of integers (i.e., not necessarily non-negative) uniquely defines an automorphism by (3.4). How do we determine the Kač automorphism σ_s conjugated to σ? This is equivalent to the determination of a simple root basis Π' with respect to which the $(l+1)$-tuple s' is special. There exists an algorithm to achieve this.[22] Let $s = (s_0,\ldots,s_l)$ be an $(l+1)$-tuple w.r.t. a simple root system Π containing negative s_i. Pick an arbitrary $s_{i_0} < 0$ and consider the simple root system $\Pi' = r_{i_0}(\Pi)$ obtained by reflecting all simple roots in the hyperplane orthogonal to the simple root α_{i_0}. Correspondingly, the $(l+1)$-tuple s changes to

$$s_i' = s_i - \tilde a_{i i_0} s_{i_0}\,, \quad (3.15)$$

where $\tilde a_{ij}$ is the Cartan matrix of $g^{(k)}$. Now repeat the algorithm for the $(l+1)$-tuple s', etc.

One can prove that this algorithm terminates after a finite number of steps.[22] Let us give an example in $g = A_3$. Suppose we have the 4-tuple $s = (4,-1,2,-1)$. We obtain successively

$$(4,-1,2,-1) \to (3,1,1,-1) \to (2,1,0,1)\,, \quad (3.16)$$

so the automorphism of order 4, determined by $s = (4,-1,2,-1)$, is conjugated to the automorphism $s = (2,1,0,1)$. In the next section we will make extensive use of this algorithm when we consider the determination of the Kač automorphism conjugated to the lift of a Weyl group element in $\mathrm{Aut}_0(g)$.

IV. LIFT OF $\mathscr{W}$ IN AUT$_0$(g)

There exists a natural lift of $\mathscr{W}$ to a finite subgroup $\widetilde{\mathscr{W}}$ of $\mathrm{Aut}_0(g)$. The generators of $\widetilde{\mathscr{W}}$ are defined by[21]

$$\tilde r_\beta = \mathrm{Ad}(x), \quad \beta\in\Delta\,, \quad (4.1)$$

with

$$x = \exp[i\pi(2(\beta,\beta))^{-1/2}(E_\beta + E_{-\beta})]\,. \quad (4.2)$$

The lift of a general $w\in\mathscr{W}(g)$ is defined as

$$\tilde w = \tilde r_{\beta_1}\cdots\tilde r_{\beta_n}\,, \quad (4.3)$$

if $w = r_{\beta_1}\cdots r_{\beta_n}$ is a reduced decomposition of w into reflection. It is not hard to show that with this definition

$$\tilde w|_h = w\,, \qquad \tilde w(g_\alpha) \subset g_{w(\alpha)}\,. \quad (4.4)$$

If $w^N = 1$ then $\tilde w^{2N} = 1$. In general, it is, however, not true that $\tilde w^N = 1$. In fact the group D generated by the $\tilde r_i^2$ is a normal Abelian subgroup of $\widetilde{\mathscr{W}}$, such that $\mathscr{W} = \widetilde{\mathscr{W}}/D$.

Now we are ready to state the main problem of this paper. Given a $w\in\mathscr{W}$ how do we determine the Kač automorphism σ_s that is conjugated to $\tilde w$, the lift of w. In principle one can try to solve this problem by explicitly trying to find a CSA h' invariant under $\tilde w$, determining the root space decomposition under h' and the $(l+1)$-tuple $(s_0,\ldots,s_l)$ with respect to a simple root basis, and, if necessary, rotate back the associated shift vector γ_s to the positive Weyl chamber by the algorithm of Sec. III. In practice, however, it is difficult to find an invariant CSA explicitly. Fortunately, the problem can be reduced to the determination of the shift vector for the primitive elements $w\in\mathscr{W}$ in simple Lie algebras only.

In Sec. II we have stated the result that all Weyl group elements are conjugated to primitive elements in (not necessarily simple) regular subalgebras. If the shift vector $\gamma_s\in h^*$ is known in this regular subalgebra it is straightforward to determine an $(l+1)$-tuple $(s_0,\ldots,s_l)$ associated to $\gamma_s\in h^*$, by identifying the regular subalgebra in the extended Dynkin diagram of g. This $(l+1)$-tuple will contain negative s_i, but can be rotated back into the positive Weyl chamber by the algorithm outlined in Sec. III. Second, we observe that the shift vectors of semisimple subalgebras are just sums of shift vectors in the simple constituents. This follows straightforwardly from the definition (3.4) and the fact that the shift vectors in the simple constituents are orthogonal. So indeed, we have reduced the problem to finding the shift vectors for the primitive elements in simple Lie algebras.

For the Coxeter class $[w_C]$ we can use an old result of Kostant[23] that the lift $\tilde w_C$ of the Coxeter element (2.8) is conjugated to the automorphism σ_s with

$$s = \underbrace{(1,1,\ldots,1)}_{l+1} .$$

For simply Lie algebras g, the associated shift vector is therefore

$$\gamma_{s=(1,1,\ldots,1)} = (1/h)\rho . \tag{4.5}$$

For the other primitive conjugacy classes the general correspondence is not known. Using, however, some properties of γ_s, it is fairly easy to determine γ_s by exhaustion. We will come back to this in the next section.

Let us now give some examples in E_8. Consider first the reflection in a root α. This corresponds to the Coxeter element in A_1. Embed A_1, say, by identifying the root of A_1 with α_2 in E_8. This leads to

$$s = (4,-1,2,-1,0,0,0,0,0) , \tag{4.6}$$

where s_0 is determined by requiring that $\tilde{w}$ has the right order:

$$4 = \sum_{i=0}^{l} a_i^\vee s_i = s_0 + 2(-1) + 3(2) + 4(-1) . \tag{4.7}$$

Notice that we have *en-passant* multiplied the order by 2, which is required to have an integer labeling s. Rotating s back to the positive Weyl chamber one obtains

$$s = (2,1,0,0,0,0,0,0,0) . \tag{4.8}$$

From this example order-doubling is clearly seen to happen. In fact, it is easy to convince oneself that the automorphism corresponding to the conjugacy class $[w] = A_1$ has order 4 in every Lie algebra $g \neq A_1$. (The example in A_3 treated in Sec. III also corresponds to the class $[w] = A_1$.) Also it is easily seen that if $N = \mathrm{ord}(w)$ is odd, no order-doubling will occur. More explicit criteria for order-doubling can be found in Refs. 10 and 11. (These criteria are also implicit in the vertex operator construction of Refs. 5 and 6.)

The second example is the class $[w] = A_2 + A_1$. Embed $A_2 + A_1$ as in Fig. 1. We have

$$\begin{aligned} s_{A_1} &= (3,2,-1,0,0,0,0,0,0), \quad N=4 \\ s_{A_2} &= (3,0,-1,1,1,-1,0,0,0), \quad N=3 . \end{aligned} \tag{4.9}$$

So adding the shift vectors leads to

$$s_{A_2+A_1} = (9,6,-7,4,4,-4,0,0,0), \quad N=12 , \tag{4.10}$$

and after rotation to

$$s_{A_2+A_1} = (4,1,0,0,0,0,0,3,0) . \tag{4.11}$$

Using the knowledge of the primitive shift vectors we have explicitly computed the shift vectors for all simply laced simple Lie algebras of rank $\leqslant 8$. [See Tables II–XVII. A * indicates that the order of $\sigma \sim w$ is twice the order of w. Other interesting properties, such as the invariant subalgebra $g_{(0)}$, can immediately be read off from the shift vectors as explained in Sec. III. Throughout the text and tables we use the numbering of the Dynkin diagram vertices of Ref. 17, the extended root corresponding to the zeroth vertex.]

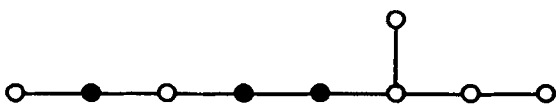


FIG. 1. The filled circles give the $A_2 + A_1$ regular subalgebra of E_8 used in the example of Sec. IV.

TABLE II. The shift vectors for the Weyl group conjugacy classes of A_1.

Γ	Ord σ	s_i
∅	1	(1,0)
A_1	2	(1,1)

TABLE III. The shift vectors for the Weyl group conjugacy classes of A_2.

Γ	Ord σ	s_i
∅	1	(1,0,0)
A_1	4*	(2,1,1)
A_2	3	(1,1,1)

TABLE IV. The shift vectors for the Weyl group conjugacy classes of A_3.

Γ	Ord σ	s_i
∅	1	(1,0,0,0)
A_1	4*	(2,1,0,1)
$2A_1$	2	(1,0,1,0)
A_2	3	(1,1,0,1)
A_3	4	(1,1,1,1)

TABLE V. The shift vectors for the Weyl group conjugacy classes of A_4.

Γ	Ord σ	s_i
∅	1	(1,0,0,0,0)
A_1	4*	(2,1,0,0,1)
$2A_1$	4*	(2,0,1,1,0)
A_2	3	(1,1,0,0,1)
A_2+A_1	12*	(4,1,3,3,1)
A_3	8*	(2,2,1,1,2)
A_4	5	(1,1,1,1,1)

TABLE VI. The shift vectors for the Weyl group conjugacy classes of A_5.

Γ	Ord σ	s_i
∅	1	(1,0,0,0,0,0)
A_1	4*	(2,1,0,0,0,1)
$2A_1$	4*	(2,0,1,0,1,0)
A_2	3	(1,1,0,0,0,1)
$3A_1$	2	(1,0,0,1,0,0)
A_2+A_1	12*	(4,1,3,0,3,1)
A_3	8*	(2,2,1,0,1,2)
$2A_2$	3	(1,0,1,0,1,0)
A_3+A_1	8*	(2,1,1,2,1,1,)
A_4	5	(1,1,1,0,1,1)
A_5	6	(1,1,1,1,1,1)

TABLE VII. The shift vectors for the Weyl group conjugacy classes of A_6.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0)
A_1	4*	(2,1,0,0,0,0,1)
$2A_1$	4*	(2,0,1,0,0,1,0)
A_2	3	(1,1,0,0,0,0,1)
$3A_1$	4*	(2,0,0,1,1,0,0)
A_2+A_1	12*	(4,1,3,0,0,3,1)
A_3	8*	(2,2,1,0,0,1,2)
A_2+2A_1	12*	(4,1,0,3,3,0,1)
$2A_2$	3	(1,0,1,0,0,1,0)
A_3+A_1	8*	(2,1,1,1,1,1,1)
A_4	5	(1,1,1,0,0,1,1)
A_3+A_2	24*	(6,1,5,3,3,5,1)
A_4+A_1	20*	(4,3,1,4,4,1,3)
A_5	12*	(2,2,2,1,1,2,2)
A_6	7	(1,1,1,1,1,1,1)

Let us now explain an important property of the automorphisms $\sigma\in\mathrm{Aut}_0(g)$ conjugated to lifts $\tilde{w}$ of $w\in\mathscr{W}(g)$. The property can be defined in terms of the multiplicity d_j of the eigenvalues ϵ^j of σ (resp. w), as defined in Secs. II and III. First extend d_j to a function on $\mathbb{N}$ by defining $d_j \equiv d_{(j \bmod N)}$, if N is the order of σ (resp. w). Now a function f: $\mathbb{N}\rightarrow\mathbb{N}$ of period N is called *quasirational* if $\gcd(i,N) = \gcd(j,N)$ implies $f(i) = f(j)$. We say that σ (resp. w) is quasirational if the corresponding function d_j is quasirational. It is evident that σ (resp. w) is quasirational if and only if its corresponding characteristic polynomial has rational coefficients. We observe (e.g., from Table I) that every Weyl group element is quasirational. This can be understood for instance by realizing that in a simple root basis the matrix of w only has integer entries. This quasirationality property of $w\in\mathscr{W}(g)$ carries over to its lift $\tilde{w}\in\mathrm{Aut}_0(g)$.[9]

Summarizing, an automorphism $\sigma\in\mathrm{Aut}_0(g)$ can only be conjugated to a lift $\tilde{w}$ of $w\in\mathscr{W}(g)$ if σ is quasirational. [The converse is not true. The automorphism σ_s, $s = (2,1,1,...,1)$, is a counterexample in almost every Lie algebra.] This severely restricts the possible lifts as there are only a finite number of conjugacy classes of quasirational elements in $\mathrm{Aut}_0(g)$.[24] If an automorphism τ is quasirational all its powers τ^j, $j\in\mathbb{Z}$, are also quasirational. If we have the property that τ^j is conjugated to τ for every j such that $\gcd(j,N) = 1$, then τ is called *rational*. This property clearly implies quasirationality. It can be verified that all Weyl group elements w are, in fact, rational. This property carries over to the lifts $\tilde{w}$ as well.

TABLE VIII. The shift vectors for the Weyl group conjugacy classes of A_7.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0,0)
A_1	4*	(2,1,0,0,0,0,0,1)
$2A_1$	4*	(2,0,1,0,0,0,1,0)
A_2	3	(1,1,0,0,0,0,0,1)
$3A_1$	4*	(2,0,0,1,0,1,0,0)
A_2+A_1	12*	(4,1,3,0,0,0,3,1)
A_3	8*	(2,2,1,0,0,0,2,1)
$4A_1$	2	(1,0,0,0,1,0,0,0)
A_2+2A_1	12*	(4,1,0,3,0,3,0,1)
$2A_2$	3	(1,0,1,0,0,0,1,0)
A_3+A_1	8*	(2,1,1,1,0,1,1,1)
A_4	5	(1,1,1,0,0,0,1,1)
$2A_2+A_1$	12*	(4,0,1,3,0,3,1,0)
A_3+2A_1	8*	(2,1,0,1,2,1,0,1)
A_3+A_2	24*	(6,1,5,3,0,3,5,1)
A_4+A_1	20*	(4,3,1,4,0,4,1,3)
A_5	12*	(2,2,2,1,0,1,2,2)
$2A_3$	4	(1,0,1,0,1,0,1,0)
A_4+A_2	15	(3,1,2,3,0,3,2,1)
A_5+A_1	6	(1,1,0,1,1,1,0,1)
A_6	7	(1,1,1,1,0,1,1,1)
A_7	8	(1,1,1,1,1,1,1,1)

TABLE IX. The shift vectors for the Weyl group conjugacy classes of A_8.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0,0,0)
A_1	4*	(2,1,0,0,0,0,0,0,1)
$2A_1$	4*	(2,0,1,0,0,0,0,1,0)
A_2	3	(1,1,0,0,0,0,0,0,1)
$3A_1$	4*	(2,0,0,1,0,0,1,0,0)
A_2+A_1	12*	(4,1,3,0,0,0,0,3,1)
A_3	8*	(2,2,1,0,0,0,0,2,1)
$4A_1$	4*	(2,0,0,0,1,1,0,0,0)
A_2+2A_1	12*	(4,1,0,3,0,0,3,0,1)
$2A_2$	3	(1,0,1,0,0,0,0,1,0)
A_3+A_1	8*	(2,1,1,1,0,0,1,1,1)
A_4	5	(1,1,1,0,0,0,0,1,1)
A_2+3A_1	12*	(4,1,0,0,3,3,0,0,1)
$2A_2+A_1$	12*	(4,0,1,3,0,0,3,1,0)
A_3+2A_1	8*	(2,1,0,1,1,1,1,0,1)
A_3+A_2	24*	(6,1,5,3,0,0,3,5,1)
A_4+A_1	20*	(4,3,1,4,0,0,4,1,3)
A_5	12*	(2,2,2,1,0,0,1,2,2)
$3A_2$	3	(1,0,0,1,0,0,1,0,0)
$2A_3$	8*	(2,0,2,0,1,1,0,2,0)
$A_3+A_2+A_1$	24*	(6,1,2,3,3,3,3,2,1)
A_4+2A_1	20*	(4,3,0,1,4,4,1,0,3)
A_4+A_2	15	(3,1,2,3,0,0,3,2,1)
A_5+A_1	12*	(2,2,0,2,1,1,2,0,2)
A_6	7	(1,1,1,1,0,0,1,1,1)
A_4+A_3	40*	(8,1,7,3,5,5,3,7,1)
A_5+A_2	12*	(2,1,1,2,1,1,2,1,1)
A_6+A_1	28*	(4,4,1,3,4,4,3,1,4)
A_7	16*	(2,2,2,2,1,1,2,2,2)
A_8	9	(1,1,1,1,1,1,1,1,1)

TABLE X. The shift vectors for the Weyl group conjugacy classes of D_4.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0)
A_1	4*	(2,0,1,0,0)
$(2A_1)'$	2	(1,1,0,0,0)
$(2A_1)''$	2	(1,0,0,1,0)
A_2	3	(1,0,1,0,0)
D_2	2	(1,1,0,0,0)
$(A_3)'$	4	(1,0,1,1,0)
$(A_3)''$	4	(1,0,1,0,1)
D_2+A_1	4*	(1,1,0,1,1)
D_3	4	(1,1,1,0,0)
$2D_2$	2	(0,0,1,0,0)
D_4	6	(1,1,1,1,1)
$D_4(a_1)$	4	(1,1,0,1,1)

TABLE XI. The shift vectors for the Weyl group conjugacy classes of D_5.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0)
A_1	4*	(2,0,1,0,0,0)
$2A_1$	4*	(2,0,0,0,1,1)
A_2	3	(1,0,1,0,0,0)
D_2	2	(1,1,0,0,0,0)
A_2+A_1	12*	(4,0,1,0,3,3)
A_3	8*	(2,0,2,0,1,1)
D_2+A_1	4*	(1,1,0,1,0,0)
D_3	4	(1,1,1,0,0,0)
A_4	5	(1,0,1,0,1,1)
D_2+A_2	6	(1,1,0,2,0,0)
D_3+A_1	4	(1,0,0,1,1,0)
$2D_2$	2	(0,0,1,0,0,0)
D_4	6	(1,1,1,1,0,0)
$D_4(a_1)$	4	(1,1,0,1,0,0)
D_3+D_2	4	(0,0,1,1,0,0)
D_5	8	(1,1,1,1,1,1)
$D_5(a_1)$	12	(2,2,1,1,2,2)

V. ORBIFOLDS VERSUS TWISTED WZW MODELS

In this section we will show how the results of the previous sections can be used to establish an explicit correspondence between a class of orbifolds and twisted WZW models. This correspondence gives rise to mathematically interesting relations, but also gives information on the moduli space of 2-D conformal field theories. Let us first recall the definition of a twisted WZW model[25] and determine its spectrum. In an untwisted WZW model[4] the basic field takes its values in some group manifold G. The conserved currents $J^a(z)$ and $\bar{J}^a(\bar{z})$ satisfy periodic boundary conditions and both satisfy the commutation relations of a Kač–Moody algebra $g^{(1)}$. The spectrum consists of integrable highest weight modules (HWM's) of $g^{(1)}$ at level k, where k is related to the topological charge of the WZW model.[26]

TABLE XII. The shift vectors for the Weyl group conjugacy classes of D_6.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0)
A_1	4*	(2,0,1,0,0,0,0)
$2A_1$	4*	(2,0,0,0,1,0,0)
A_2	3	(1,0,1,0,0,0,0)
D_2	2	(1,1,0,0,0,0,0)
$(3A_1)'$	2	(1,0,0,0,0,1,0)
$(3A_1)''$	2	(1,0,0,0,0,0,1)
A_2+A_1	12*	(4,0,1,0,3,0,0)
A_3	8*	(2,0,2,0,1,0,0)
D_2+A_1	4*	(1,1,0,1,0,0,0)
D_3	4	(1,1,1,0,0,0,0)
$2A_2$	3	(1,0,0,0,1,0,0)
A_3+A_1	8*	(2,0,1,0,1,2,0)
A_4	5	(1,0,1,0,1,0,0)
D_2+2A_1	4*	(1,1,0,0,0,1,1)
D_2+A_2	6	(1,1,0,2,0,0,0)
D_3+A_1	4	(1,1,0,0,1,0,0)
$2D_2$	2	(0,0,1,0,0,0,0)
D_4	6	(1,1,1,1,0,0,0)
$D_4(a_1)$	4	(1,1,0,1,0,0,0)
A_5	6	(1,0,1,0,1,1,0)
D_2+A_3	8*	(1,1,0,2,0,1,1)
$2D_2+A_1$	4*	(0,0,1,0,1,0,0)
D_3+A_2	12	(2,2,0,1,3,0,0)
D_4+A_1	12*	(2,2,1,0,1,2,2)
$D_4(a_1)+A_1$	4	(1,1,0,0,0,1,1)
D_3+D_2	4	(0,0,1,1,0,0,0)
D_5	8	(1,1,1,1,1,0,0)
$D_5(a_1)$	12	(2,2,1,1,2,0,0)
$3D_2$	2	(0,0,0,1,0,0,0)
$2D_3$	4	(0,0,1,0,1,0,0)
D_4+D_2	6	(0,0,1,1,1,0,0)
D_6	10	(1,1,1,1,1,1,1)
$D_6(a_1)$	8	(1,1,1,0,1,1,1)
$D_6(a_2)$	6	(1,1,0,1,0,1,1)

TABLE XIII. The shift vectors for the Weyl group conjugacy classes of D_7.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0,0)
A_1	4*	(2,0,1,0,0,0,0,0)
$2A_1$	4*	(2,0,0,0,1,0,0,0)
A_2	3	(1,0,1,0,0,0,0,0)
D_2	2	(1,1,0,0,0,0,0,0)
$(3A_1)$	4*	(2,0,0,0,0,0,1,1)
A_2+A_1	12*	(4,0,1,0,3,0,0,0)
A_3	8*	(2,0,2,0,1,0,0,0)
D_2+A_1	4*	(1,1,0,1,0,0,0,0)
D_3	4	(1,1,1,0,0,0,0,0)
A_2+2A_1	12*	(4,0,1,0,0,0,3,3)
$2A_2$	3	(1,0,0,0,1,0,0,0)
A_3+A_1	8*	(2,0,1,0,1,0,1,1)
A_4	5	(1,0,1,0,1,0,0,0)
D_2+2A_1	4*	(1,1,0,0,0,1,0,0)
D_2+A_2	6	(1,1,0,2,0,0,0,0)
D_3+A_1	4	(1,1,0,0,1,0,0,0)
$2D_2$	2	(0,0,1,0,0,0,0,0)
D_4	6	(1,1,1,1,0,0,0,0)
$D_4(a_1)$	4	(1,1,0,1,0,0,0,0)
A_3+A_2	24*	(6,0,1,0,5,0,3,3)
A_4+A_1	20*	(4,0,3,0,1,0,4,4)
A_5	12*	(2,0,2,0,2,0,1,1)
$D_2+A_2+A_1$	12*	(2,2,0,1,0,3,0,0)
D_2+A_3	8*	(1,1,0,2,0,1,0,0)
$2D_2+A_1$	4*	(0,0,1,0,1,0,0,0)
D_3+2A_1	4	(1,1,0,0,0,0,1,1)
D_3+A_2	12	(2,2,0,1,3,0,0,0)
D_4+A_1	12*	(2,2,1,0,1,2,0,0)
$D_4(a_1)+A_1$	4	(1,1,0,0,0,1,0,0)
D_3+D_2	4	(0,0,1,1,0,0,0,0)
D_5	8	(1,1,1,1,1,0,0,0)
$D_5(a_1)$	12	(2,2,1,1,2,0,0,0)
A_6	7	(1,0,1,0,1,0,1,1)
D_2+A_4	10	(1,1,0,2,0,2,0,0)
$2D_2+A_2$	6	(0,0,1,0,2,0,0,0)
$D_3+D_2+A_1$	4	(0,0,1,0,0,1,0,0)
D_3+A_3	8*	(1,1,0,1,1,0,1,1)
D_4+A_2	6	(1,1,0,0,1,1,0,0)
D_5+A_1	8	(1,1,1,0,0,1,1,1)
$D_5(a_1)+A_1$	12	(2,2,1,0,0,1,2,2)
$3D_2$	2	(0,0,0,1,0,0,0,0)
$2D_3$	4	(0,0,1,0,1,0,0,0)
D_4+D_2	6	(0,0,1,1,1,0,0,0)
D_6	10	(1,1,1,1,1,1,0,0)
$D_6(a_1)$	8	(1,1,1,0,1,1,0,0)
$D_6(a_2)$	6	(1,1,0,1,0,1,0,0)
D_3+2D_2	4	(0,0,0,1,1,0,0,0)
D_4+D_3	12	(0,0,2,1,1,2,0,0)
$D_4(a_1)+D_3$	4	(0,0,1,0,0,1,0,0)
D_5+D_2	8	(0,0,1,1,1,1,0,0)
D_7	12	(1,1,1,1,1,1,1,1)
$D_7(a_1)$	20	(2,2,2,1,1,2,2,2)
$D_7(a_2)$	24	(3,3,1,2,2,1,3,3)

TABLE XIV. The shift vectors for the Weyl group conjugacy classes of D_8.

Γ	Ord σ	s_i	Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0,0,0)	$D_3+A_2+A_1$	12	(2,2,0,1,0,0,3,0,0)
A_1	4*	(2,0,1,0,0,0,0,0,0)	$D_3+D_2+A_1$	4	(0,0,1,0,0,1,0,0,0)
$2A_1$	4*	(2,0,0,0,1,0,0,0,0)	D_3+A_3	8*	(1,1,0,1,1,0,1,0,0)
A_2	3	(1,0,1,0,0,0,0,0,0)	D_4+2A_1	12*	(2,2,1,0,0,0,1,2,2)
D_2	2	(1,1,0,0,0,0,0,0,0)	D_4+A_2	6	(1,1,0,0,1,1,0,0,0)
$3A_1$	4*	(2,0,0,0,0,0,1,0,0)	D_5+A_1	8	(1,1,1,0,0,1,1,0,0)
A_2+A_1	12*	(4,0,1,0,3,0,0,0,0)	$D_5(a_1)+A_1$	12	(2,2,1,0,0,1,2,0,0)
A_3	8*	(2,0,2,0,1,0,0,0,0)	$3D_2$	2	(0,0,0,1,0,0,0,0,0)
D_2+A_1	4*	(1,1,0,1,0,0,0,0,0)	$2D_3$	4	(0,0,1,0,1,0,0,0,0)
D_3	4	(1,1,1,0,0,0,0,0,0)	D_4+D_2	6	(0,0,1,1,1,0,0,0,0)
$(4A_1)'$	2	(1,0,0,0,0,0,0,0,1)	D_6	10	(1,1,1,1,1,1,0,0,0)
$(4A_1)''$	2	(1,0,0,0,0,0,0,1,0)	$D_6(a_1)$	8	(1,1,1,0,1,1,0,0,0)
A_2+2A_1	12*	(4,0,1,0,0,0,3,0,0)	$D_6(a_2)$	6	(1,1,0,1,0,1,0,0,0)
$2A_2$	3	(1,0,0,0,1,0,0,0,0)	$(A_7)'$	8	(1,0,1,0,1,0,1,1,0)
A_3+A_1	8*	(2,0,1,0,1,0,1,0,0)	$(A_7)''$	8	(1,0,1,0,1,0,1,0,1)
A_4	5	(1,0,1,0,1,0,0,0,0)	D_2+A_5	12*	(1,1,0,2,0,2,0,1,1)
D_2+2A_1	4*	(1,1,0,0,0,1,0,0,0)	$2D_2+A_3$	8*	(0,0,1,0,2,0,1,0,0)
D_2+A_2	6	(1,1,0,2,0,0,0,0,0)	$3D_2+A_1$	4*	(0,0,0,1,0,1,0,0,0)
D_3+A_1	4	(1,1,0,0,1,0,0,0,0)	D_3+A_4	20	(2,2,0,3,1,0,4,0,0)
$2D_2$	2	(0,0,1,0,0,0,0,0,0)	$D_3+D_2+A_2$	12	(0,0,2,0,1,3,0,0,0)
D_4	6	(1,1,1,1,0,0,0,0,0)	$2D_3+A_1$	4	(0,0,1,0,0,0,1,0,0)
$D_4(a_1)$	4	(1,1,0,1,0,0,0,0,0)	D_4+A_3	24	(3,3,0,1,4,1,0,3,3)
$2A_2+A_1$	12*	(4,0,0,0,1,0,3,0,0)	$D_4(a_1)+A_3$	8*	(1,1,0,1,0,1,0,1,1)
$(A_3+2A_1)'$	8*	(2,0,1,0,0,0,1,0,2)	$D_4+D_2+A_1$	12*	(0,0,2,1,0,1,2,0,0)
$(A_3+2A_1)''$	8*	(2,0,1,0,0,0,1,2,0)	D_5+A_2	24	(3,3,1,0,2,3,3,0,0)
A_3+A_2	24*	(6,0,1,0,5,0,3,0,0)	$D_5(a_1)+A_2$	12	(2,2,0,0,1,1,2,0,0)
A_4+A_1	20*	(4,0,3,0,1,0,4,0,0)	D_6+A_1	20*	(2,2,2,1,0,1,2,2,2)
A_5	12*	(2,0,2,0,2,0,1,0,0)	$D_6(a_1)+A_1$	8	(1,1,1,0,0,0,1,1,1)
D_2+3A_1	4*	(1,1,0,0,0,0,0,1,1)	$D_6(a_2)+A_1$	12*	(2,2,0,1,0,1,0,2,2)
$D_2+A_2+A_1$	12*	(2,2,0,1,0,3,0,0,0)	D_3+2D_2	4	(0,0,0,1,1,0,0,0,0)
D_2+A_3	8*	(1,1,0,2,0,1,0,0,0)	D_4+D_3	12	(0,0,2,1,1,2,0,0,0)
$2D_2+A_1$	4*	(0,0,1,0,1,0,0,0,0)	$D_4(a_1)+D_3$	4	(0,0,1,0,0,1,0,0,0)
D_3+2A_1	4	(1,1,0,0,0,0,1,0,0)	D_5+D_2	8	(0,0,1,1,1,1,0,0,0)
D_3+A_2	12	(2,2,0,1,3,0,0,0,0)	D_7	12	(1,1,1,1,1,1,1,0,0)
D_4+A_1	12*	(2,2,1,0,1,2,0,0,0)	$D_7(a_1)$	20	(2,2,2,1,1,2,2,0,0)
$D_4(a_1)+A_1$	4	(1,1,0,0,0,1,0,0,0)	$D_7(a_2)$	24	(3,3,1,2,2,1,3,0,0)
D_3+D_2	4	(0,0,1,1,0,0,0,0,0)	$4D_2$	2	(0,0,0,0,1,0,0,0,0)
D_5	8	(1,1,1,1,1,0,0,0,0)	$2D_3+D_2$	4	(0,0,0,1,0,1,0,0,0)
$D_5(a_1)$	12	(2,2,1,1,2,0,0,0,0)	D_4+2D_2	6	(0,0,0,1,1,1,0,0,0)
$(2A_3)'$	4	(1,0,0,0,1,0,0,1,0)	$2D_4$	6	(0,0,1,0,1,0,1,0,0)
$(2A_3)''$	4	(1,0,0,0,1,0,0,0,1)	$2D_4(a_1)$	4	(0,0,1,0,0,0,1,0,0)
A_4+A_2	15	(3,0,1,0,2,0,3,0,0)	D_5+D_3	8	(0,0,1,1,0,1,1,0,0)
$(A_5+A_1)'$	6	(1,0,1,0,0,0,1,1,0)	$D_5(a_1)+D_3$	12	(0,0,2,1,0,1,2,0,0)
$(A_5+A_1)''$	6	(1,0,1,0,0,0,1,0,1)	D_6+D_2	10	(0,0,1,1,1,1,1,0,0)
A_6	7	(1,0,1,0,1,0,1,0,0)	D_8	14	(1,1,1,1,1,1,1,1,1)
D_2+2A_2	6	(1,1,0,0,0,2,0,0,0)	$D_8(a_1)$	12	(1,1,1,1,0,1,1,1,1)
D_2+A_4	10	(1,1,0,2,0,2,0,0,0)	$D_8(a_2)$	30	(3,3,2,1,3,1,2,3,3)
$2D_2+2A_1$	4*	(0,0,1,0,0,0,1,0,0)	$D_8(a_3)$	8	(1,1,0,1,0,1,0,1,1)
$2D_2+A_2$	6	(0,0,1,0,2,0,0,0,0)			

To every subgroup $\mathscr{H}$ of the automorphism group $\mathrm{Aut}(g)$ one can associate a twisted WZW model, by identifying points in the group manifold G which differ by the action of an automorphism $\sigma \in \mathscr{H}$. For simplicity we will restrict the discussion to cyclic subgroups of $\mathrm{Aut}_0(G)$. Let N be the order of a generating element σ. The model has various twisted sectors. The currents in a sector, labeled by an integer $s = 0,1,\ldots,N-1$, satisfy boundary conditions

$$J(e^{2\pi i}z) = \sigma^s J(z) . \tag{5.1}$$

An analogous equation holds for the right-moving components.

In every sector the currents $J^a(z)$ still satisfy the commutation relations of a Kač–Moody algebra,[13] which, for inner automorphisms, is isomorphic to the untwisted Kač–Moody algebra. The difference, however, lies in the way the conformal algebra is realized on the spectrum. In particular, the commutation relation between the Virasoro generator $L_0(\sigma^s)$ in the σ^s sector and the step operators $E_\alpha(z)$ changes into

$$[L_0(\sigma^s),E_\alpha(z)] = -\left(z\frac{d}{dz} + s(\gamma_s,\alpha)\right)E_\alpha(z) , \tag{5.2}$$

if the automorphism σ is rotated into the standard form (3.7).

To obtain the physical spectrum, i.e., the torus partition function $\mathscr{Z}$, of the twisted WZW model, we have to combine the various sectors and project, in each sector, on the states variant under the automorphism σ. The projection operator is given by

TABLE XV. The shift vectors for the Weyl group conjugacy classes of E_6.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0)
A_1	4*	(2,0,0,0,0,0,1)
$2A_1$	4*	(2,1,0,0,0,1,0)
A_2	3	(1,0,0,0,0,0,1)
$3A_1$	4*	(1,0,0,1,0,0,0)
A_2+A_1	12*	(4,3,0,0,0,3,1)
A_3	8*	(2,1,0,0,0,1,2)
$4A_1$	2	(0,0,0,0,0,0,1)
A_2+2A_1	12*	(2,0,1,2,1,0,0)
$2A_2$	3	(1,1,0,0,0,1,0)
A_3+A_1	8*	(2,1,1,0,1,1,0)
A_4	5	(1,1,0,0,0,1,1)
D_4	6	(1,0,0,1,0,0,1)
$D_4(a_1)$	4	(1,0,0,1,0,0,0)
$2A_2+A_1$	12*	(1,1,0,3,0,1,0)
A_3+2A_1	8*	(0,0,1,0,1,0,2)
A_4+A_1	20*	(3,1,3,1,3,1,0)
A_5	12*	(2,2,1,0,1,2,1)
D_5	8	(1,1,0,1,0,1,1)
$D_5(a_1)$	24*	(4,3,1,2,1,3,2)
$3A_2$	3	(0,0,0,1,0,0,0)
A_5+A_1	6	(0,0,1,0,1,0,1)
E_6	12	(1,1,1,1,1,1,1)
$E_6(a_1)$	9	(1,1,1,0,1,1,1)
$E_6(a_2)$	6	(1,1,0,1,0,1,0)

$$P=\frac{1}{N}\sum_{r=0}^{N-1}\sigma^r . \tag{5.3}$$

So we have

$$\mathscr{Z}=\frac{1}{N}\sum_{0<r,s<N-1}\mathscr{Z}^{(r,s)}, \tag{5.4}$$

where

$$\mathscr{Z}^{(r,s)}=(q\bar{q})^{-c/24}\,\mathrm{Tr}(\sigma^r q^{L_0(\sigma^s)}\bar{q}^{\bar{L}_0(\sigma^s)}),\quad q=e^{2\pi i\tau}. \tag{5.5}$$

The partition function of a twisted WZW model can be expressed in terms of specializations of the affine Kač–Moody characters.

The character of an integrable HWM $L(\Lambda)$ at level k is defined as[20]

$$\chi_\Lambda(z,\tau,t)=q^{(\Lambda,\Lambda+2\rho)/2(k+h^\vee)-c/24}e^{-2\pi ikt} \times\mathrm{Tr}_{L(\Lambda)}(e^{-2\pi i(\lambda,z)}q^{-(\lambda,\Lambda_0)}), \tag{5.6}$$

where $z\in h^*$, the dual Cartan subalgebra of g, and Λ_0 is the element dual to the central charge (i.e., the highest weight of the basic representation).

A specialization (α,β) of χ_Λ is an expression of the form

$$\chi_{\Lambda|(\alpha,\beta)}(z,\tau,t)\equiv\chi_\Lambda\big(z-\alpha+\tau\beta,\tau,t-(\beta,z) -(\tau/2)(\beta,\beta)+\tfrac{1}{2}(\alpha,\beta)\big), \tag{5.7}$$

where α, $\beta\in h^*$. If the automorphism $\sigma\in\mathrm{Aut}_0(g)$ corresponds to a shift vector $\gamma_s\in h^*$, then by (5.2) a state $|\lambda\rangle$ in the HWM $L(\Lambda)$ has an energy $-(\lambda,\Lambda_0+\gamma_s)$ relative to the vacuum energy, and an eigenvalue $e^{2\pi i(\lambda,\gamma_s)}$ under σ. Using (5.5) it is now easy to convince oneself that the contribution of the HWM $L(\Lambda)$ to the partition function $\mathscr{Z}^{(r,s)}$ in the (r,s) section is given by

TABLE XVI. The shift vectors for the Weyl group conjugacy classes of E_7.

Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0,0)
A_1	4*	(2,1,0,0,0,0,0,0)
A_2	3	(1,1,0,0,0,0,0,0)
$2A_1$	4*	(2,0,0,0,0,1,0,0)
A_3	8*	(2,2,0,0,0,1,0,0)
A_2+A_1	12*	(4,1,0,0,0,3,0,0)
$(3A_1)'$	4*	(1,0,1,0,0,0,0,0)
$(3A_1)''$	2	(1,0,0,0,0,0,1,0)
A_4	5	(1,1,0,0,0,1,0,0)
$(A_3+A_1)'$	8*	(2,0,0,1,0,1,0,0)
$(A_3+A_1)''$	8*	(2,1,0,0,0,1,2,0)
$2A_2$	3	(1,0,0,0,0,1,0,0)
A_2+2A_1	12*	(2,0,2,1,0,0,0,0)
$(4A_1)'$	2	(0,1,0,0,0,0,0,0)
$(4A_1)''$	4*	(1,0,0,0,0,0,1,1)
D_4	6	(1,1,1,0,0,0,0,0)
$D_4(a_1)$	4	(1,0,1,0,0,0,0,0)
$(A_5)'$	12*	(2,1,0,1,0,2,0,0)
$(A_5)''$	6	(1,1,0,0,0,1,1,0)
A_4+A_1	20*	(3,0,1,3,0,1,0,0)
A_3+A_2	24*	(4,0,2,1,0,5,0,0)
$(A_3+2A_1)'$	8*	(0,2,0,1,0,0,0,0)
$(A_3+2A_1)''$	8*	(1,0,1,0,1,0,1,0)
$2A_2+A_1$	12*	(1,0,3,0,0,1,0,0)
A_2+3A_1	6	(1,0,0,0,0,0,1,2)
$5A_1$	4*	(0,1,0,0,0,1,0,0)
D_5	8	(1,1,1,0,0,1,0,0)
D_4+A_1	12*	(2,1,0,0,0,1,2,2)
$D_5(a_1)$	24*	(4,2,2,1,0,3,0,0)
$D_4(a_1)+A_1$	4	(1,0,0,0,0,0,1,1)
A_6	7	(1,0,0,1,0,1,0,0)
$(A_5+A_1)'$	12*	(1,0,1,1,1,0,1,0)
$(A_5+A_1)''$	6	(0,1,0,1,0,0,0,0)
A_4+A_2	15	(1,0,2,1,0,2,0,0)
$2A_3$	4	(0,1,0,0,0,1,0,0)
$A_3+A_2+A_1$	24*	(1,0,3,0,3,0,1,2)
A_3+3A_1	8*	(0,1,0,0,0,1,0,2)
$3A_2$	3	(0,0,1,0,0,0,0,0)
$6A_1$	4*	(0,0,0,1,0,0,0,0)
E_6	12	(1,1,1,1,0,1,0,0)
D_6	20*	(2,2,1,0,1,2,2,1)
D_5+A_1	8	(1,0,0,1,0,0,1,1)
D_4+2A_1	12*	(0,2,0,1,0,2,0,0)
$D_6(a_1)$	8	(1,1,0,0,0,1,1,1)
$D_6(a_2)$	12*	(2,0,1,0,1,0,2,1)
$E_6(a_1)$	9	(1,1,0,1,0,1,0,0)
$E_6(a_2)$	6	(1,0,1,0,0,1,0,0)
$D_5(a_1)+A_1$	24*	(3,0,1,2,1,0,3,2)
A_7	8	(0,1,0,1,0,1,0,0)
A_5+A_2	6	(0,0,1,0,1,0,0,0)
$2A_3+A_1$	4	(0,0,0,1,0,0,0,0)
$7A_1$	2	(0,0,0,0,0,0,0,1)
E_7	18	(1,1,1,1,1,1,1,1)
D_6+A_1	10	(0,1,0,1,0,1,0,1)
$D_6(a_2)+A_1$	6	(0,1,0,0,0,1,0,1)
D_4+3A_1	6	(0,0,0,1,0,0,0,1)
$E_7(a_1)$	14	(1,1,1,0,1,1,1,1)
$E_7(a_2)$	12	(1,1,0,1,0,1,1,1)
$E_7(a_3)$	30	(3,2,1,2,1,2,3,1)
$E_7(a_4)$	6	(1,0,0,1,0,0,1,0)

$$e^{-\pi irs(\gamma_s,\gamma_s)}\,\mathrm{Tr}_{L(\Lambda)}(\sigma^r q^{L_0(\sigma^s)-c/24})=\chi_{\Lambda|(r\gamma_s,s\gamma_s)}(0,\tau,0) \equiv\chi_\Lambda^{(r,s)}(\tau). \tag{5.8}$$

If the partition function of the untwisted WZW model is written in terms of the untwisted character $\chi_\Lambda\equiv\chi_\Lambda^{(0,0)}$ as

TABLE XVII. The shift vectors for the Weyl group conjugacy classes of E_8.

Γ	Ord σ	s_i	Γ	Ord σ	s_i
$\varnothing$	1	(1,0,0,0,0,0,0,0,0)	$A_4+A_2+A_1$	60*	(1,0,3,0,5,0,4,0,3)
A_1	4*	(2,1,0,0,0,0,0,0,0)	A_4+A_3	40*	(1,0,3,0,4,1,0,2,0)
$2A_1$	4*	(2,0,0,0,0,0,0,1,0)	A_5+2A_1	12*	(0,1,0,0,0,1,1,0,0)
A_2	3	(1,1,0,0,0,0,0,0,0)	A_5+A_2	12*	(0,0,2,0,0,1,0,0,0)
$3A_1$	4*	(1,0,1,0,0,0,0,0,0)	A_6+A_1	28*	(1,0,3,0,0,1,3,0,0)
A_2+A_1	12*	(4,1,0,0,0,0,0,3,0)	$(A_7)'$	16*	(1,0,1,1,0,1,0,1,0)
A_3	8*	(2,2,0,0,0,0,0,1,0)	$(A_7)''$	8	(0,1,0,1,0,0,0,1,0)
$(4A_1)'$	4*	(1,0,0,0,0,0,0,0,1)	D_4+3A_1	12*	(0,0,0,2,0,0,1,0,0)
$(4A_1)''$	2	(0,1,0,0,0,0,0,0,0)	D_4+A_3	24*	(0,3,0,0,0,1,2,2,0)
A_2+2A_1	12*	(2,0,2,1,0,0,0,0,0)	$D_4(a_1)+A_3$	8*	(0,1,0,0,0,1,0,0,0)
$2A_2$	3	(1,0,0,0,0,0,0,1,0)	D_5+2A_1	8	(0,1,0,0,0,0,1,1,0)
A_3+A_1	8*	(2,0,0,1,0,0,0,1,0)	D_5+A_2	24	(1,0,2,1,0,0,0,2,3)
A_4	5	(1,1,0,0,0,0,0,1,0)	$D_5(a_1)+A_2$	24*	(1,0,3,0,0,1,0,1,2)
D_4	6	(1,1,1,0,0,0,0,0,0)	D_6+A_1	20*	(0,2,0,2,0,1,0,1,0)
$D_4(a_1)$	4	(1,0,1,0,0,0,0,0,0)	$D_6(a_2)+A_1$	12*	(0,2,0,0,0,1,0,1,0)
$5A_1$	4*	(0,1,0,0,0,0,0,1,0)	E_6+A_1	12	(1,0,0,1,0,0,1,0,1)
A_2+3A_1	12*	(2,0,0,0,0,0,1,0,2)	$E_6(a_1)+A_1$	36*	(3,0,1,3,0,1,3,0,0)
$2A_2+A_1$	12*	(1,0,3,0,0,0,0,1,0)	$E_6(a_2)+A_1$	12*	(1,0,1,0,1,0,0,0,1)
$(A_3+2A_1)'$	8*	(1,0,1,0,0,0,1,0,0)	D_7	24*	(2,1,0,1,1,0,1,2,1)
$(A_3+2A_1)''$	8*	(0,2,0,1,0,0,0,0,0)	$D_7(a_1)$	40*	(4,1,0,3,1,0,1,2,3)
A_3+A_2	24*	(4,0,2,1,0,0,0,5,0)	$D_7(a_2)$	24	(3,0,0,1,2,0,0,2,1)
A_4+A_1	20*	(3,0,1,3,0,0,0,1,0)	E_7	36*	(2,2,2,2,1,0,1,2,1)
A_5	12*	(2,1,0,1,0,0,0,2,0)	$E_7(a_1)$	28*	(2,2,2,0,1,0,1,2,1)
D_4+A_1	12*	(2,1,0,0,0,0,0,1,2)	$E_7(a_2)$	12	(1,1,0,1,0,0,0,1,1)
$D_4(a_1)+A_1$	4	(1,0,0,0,0,0,0,0,1)	$E_7(a_3)$	60*	(6,4,2,3,0,1,2,4,2)
D_5	8	(1,1,1,0,0,0,0,1,0)	$E_7(a_4)$	12*	(2,0,0,1,0,1,0,0,0)
$D_5(a_1)$	24*	(4,2,2,1,0,0,0,3,0)	$8A_1$	2	(0,0,0,0,0,0,0,1,0)
$6A_1$	4*	(0,0,0,1,0,0,0,0,0)	$4A_2$	3	(0,0,0,0,0,0,0,0,1)
A_2+4A_1	6	(0,1,0,0,0,0,0,2,0)	$2A_3+2A_1$	4	(0,0,0,0,0,0,1,0,0)
$2A_2+2A_1$	12*	(1,0,0,0,1,0,0,0,2)	$2A_4$	5	(0,0,0,0,1,0,0,0,0)
$3A_2$	3	(0,0,1,0,0,0,0,0,0)	$A_5+A_2+A_1$	6	(0,0,0,0,0,1,0,0,0)
A_3+3A_1	8*	(0,1,0,0,0,1,0,0,0)	A_7+A_1	8	(0,0,0,1,0,0,1,0,0)
$A_3+A_2+A_1$	24*	(1,0,3,0,2,0,1,0,0)	A_8	9	(0,0,1,0,0,1,0,0,0)
$(2A_3)'$	4	(0,1,0,0,0,0,0,1,0)	D_4+4A_1	6	(0,0,0,0,0,0,1,1,0)
$(2A_3)''$	8*	(1,0,0,0,1,0,0,1,0)	$2D_4$	6	(0,0,0,1,0,0,0,1,0)
A_4+2A_1	20*	(2,0,1,0,0,0,3,0,1)	$2D_4(a_1)$	4	(0,0,0,1,0,0,0,0,0)
A_4+A_2	15	(1,0,2,1,0,0,0,2,0)	$D_5(a_1)+A_3$	12	(0,0,0,2,0,0,1,0,0)
$(A_5+A_1)'$	6	(0,1,0,1,0,0,0,0,0)	D_6+2A_1	10	(0,0,0,1,0,0,1,1,0)
$(A_5+A_1)''$	12*	(1,0,1,1,0,0,1,0,0)	D_8	14	(0,1,0,1,0,1,0,1,0)
A_6	7	(1,0,0,1,0,0,0,1,0)	$D_8(a_1)$	12	(0,1,0,1,0,1,0,0,0)
D_4+2A_1	12*	(0,2,0,1,0,0,0,2,0)	$D_8(a_2)$	30	(0,3,0,2,0,1,2,1,0)
D_4+A_2	6	(1,0,0,0,0,0,0,1,1)	$D_8(a_3)$	8	(0,1,0,0,0,1,0,0,0)
$D_4(a_1)+A_2$	12	(2,0,0,0,0,0,1,0,2)	E_6+A_2	12	(0,0,1,0,0,1,0,0,1)
D_5+A_1	8	(1,0,0,1,0,0,0,0,1)	$E_6(a_2)+A_2$	6	(0,0,1,0,0,0,0,0,1)
$D_5(a_1)+A_1$	24*	(3,0,1,2,0,0,1,0,2)	E_7+A_1	18	(0,1,0,1,0,1,1,1,0)
D_6	20*	(2,2,1,0,0,0,1,2,1)	$E_7(a_2)+A_1$	12	(0,1,0,1,0,0,1,1,0)
$D_6(a_1)$	8	(1,1,0,0,0,0,0,1,1)	$E_7(a_4)+A_1$	6	(0,1,0,0,0,0,1,0,0)
$D_6(a_2)$	12*	(2,0,1,0,0,0,1,0,1)	E_8	30	(1,1,1,1,1,1,1,1,1)
E_6	12	(1,1,1,1,0,0,0,1,0)	$E_8(a_1)$	24	(1,1,1,1,1,0,1,1,1)
$E_6(a_1)$	9	(1,1,0,1,0,0,0,1,0)	$E_8(a_2)$	20	(1,1,1,0,1,0,1,1,1)
$E_6(a_2)$	6	(1,0,1,0,0,0,0,1,0)	$E_8(a_3)$	12	(1,0,1,0,0,1,0,1,0)
$7A_1$	4*	(0,0,0,0,0,0,1,0,0)	$E_8(a_4)$	18	(1,1,1,0,1,0,1,0,1)
$3A_2+A_1$	12*	(0,0,1,0,0,0,0,0,3)	$E_8(a_5)$	15	(1,1,0,1,0,1,0,1,0)
A_3+4A_1	8*	(0,0,0,1,0,0,0,2,0)	$E_8(a_6)$	10	(1,0,1,0,0,1,0,0,0)
$A_3+A_2+2A_1$	24*	(0,1,0,0,0,3,0,2,0)	$E_8(a_7)$	12	(1,1,0,0,1,0,1,0,0)
$2A_3+A_1$	4	(0,0,0,1,0,0,0,0,0)	$E_8(a_8)$	6	(1,0,0,0,1,0,0,0,0)

$$\mathscr{Z}_{\text{untwisted}} = \sum_{\Lambda,\Lambda'} \overline{\chi_\Lambda} \mathscr{N}_{\Lambda,\Lambda'} \chi_{\Lambda'}\,, \tag{5.9}$$

where the sum is over the integral HWM's at level k, then the twisted WZW model has partition function

$$\mathscr{Z}_{\text{twisted}} = \frac{1}{N} \sum_{0 \leqslant r,s \leqslant N-1} \sum_{\Lambda,\Lambda'} \overline{\chi_\Lambda^{(r,s)}} \mathscr{N}_{\Lambda,\Lambda'} \chi_{\Lambda'}^{(r,s)} \tag{5.10}$$

[In particular, for simply connected group manifolds we have $\mathscr{N}_{\Lambda,\Lambda'} = \delta_{\Lambda,\Lambda'}$ (Ref. 26)]. If $g_{(0)}$ is semisimple [no U(1) factors] then (5.10) can easily be rewritten in terms of level-k characters of the Kač–Moody algebra $(g_{(0)})^{(1)}$. Observe thereto that in (5.10) only left ⊗ right weight combinations (λ,λ') survive for which

$$(\lambda - \lambda', \gamma_s) \in \mathbf{Z}\,. \tag{5.11}$$

To determine these it suffices to look at the decomposition of the HWM's $L(\Lambda) \otimes L(\Lambda')$ of $g^{(1)} \otimes g^{(1)}$ into HWM's $L(\overline{\Lambda}) \otimes L(\overline{\Lambda}')$ of $(g_{(0)})^{(1)} \otimes (g_{(0)})^{(1)}$. (This decomposition is finite, see, e.g., Ref. 27 and references therein.) Then determine the set $(\overline{\Lambda},\overline{\Lambda}')$ of highest weights satisfying (5.11). If (5.11) is satisfied for the highest weights, then all the weights in the HWM satisfy (5.11) because they differ from the highest weight by the subtraction of simple roots of $g_{(0)}$ which are all orthogonal to γ_s.

The modular group SL(2,**Z**) acts on χ_Λ by

$$\chi_{\Lambda|A}(z,\tau,t) = \chi_\Lambda\left(\frac{z}{c\tau+d},\frac{a\tau+b}{c\tau+d},t+\frac{c}{2}\frac{|z|^2}{c\tau+d}\right), \quad A=\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in \mathrm{SL}(2,\mathbf{Z}) . \tag{5.12}$$

There is a beautiful interplay between specialization and modular transformations[20]

$$\chi^{(r,s)}_{\Lambda|A}(\tau) = \chi_\Lambda^{(dr-bs,\,-cr+as)}\left(\frac{a\tau+b}{c\tau+d}\right), \tag{5.13}$$

which shows that the twisted partition function (5.10) is, in fact, modular invariant. The ground state mass shift in (5.8) is adjusted such that the set $\Lambda^{(r,s)}_\Lambda$, $0\leqslant r,s\leqslant N-1$, $\Lambda\in P^k_+$, forms an invariant subspace under modular transformations (5.13). For simply laced Lie algebras of rank l at level 1, the twisted character (5.8) is explicitly given by[20]

$$\chi^{(r,s)}_\Lambda = \frac{1}{\eta(\tau)^l}\, e^{i\pi rs(\gamma_s,\gamma_s)} \sum_{\lambda\in\Lambda_R+\Lambda-s\gamma_s} e^{2\pi i r(\lambda,\gamma_s)} q^{(1/2)|\lambda|^2} . \tag{5.14}$$

Compared to the untwisted character, the theta function is now taken over a lattice shifted by $-s\gamma_s$. This explains the name "shift vector." It should be remarked that (5.14) is often taken as the definition of a shifted torus compactification. We want to stress that the use of a shift vector is not restricted to level-1, simply laced, Kač–Moody modules, but applies to arbitrary levels and also non-simply-laced. The interpretation of γ_s as a shift of the compactification lattice holds only for simply laced at level 1, though.

The class of orbifolds we are considering are obtained by identifying points on a maximal torus $\mathbf{T}^l=\mathbf{R}^l/\Lambda_R$ of a Lie algebra g, related by a subgroup of the automorphism group of the root lattice Λ_R. The automorphism group of Λ_R is given by $\mathrm{Aut}(\Lambda_R)=\mathscr{W}(g)\times\mathscr{D}(g)$, where $\mathscr{D}(g)$ is a finite Abelian group isomorphic to the symmetry group of the Dynkin diagram of g. As before we restrict ourselves to cyclic subgroups of $\mathrm{Aut}(\Lambda_R)$ generated by an element $w\in\mathscr{W}(g)$. A closed string propagating on a torus $\mathbf{T}^l$ or an orbifold version thereof is described by a set of l scalar fields $X^I(z,\bar z)$ coupled to a background constant rank l antisymmetric tensor field B^{IJ}. If the field B^{IJ} is chosen as (Ref. 28)

$$B=(1+w_C)/(1-w_C) , \tag{5.15}$$

then the torus string is known to be equivalent to the level-1 WZW model based on the algebra g. [The crucial property is $(1-w)\Lambda_W=\Lambda_R$. Because this holds for all primitive elements w one could alternatively take any other primitive element.] For the orbifold models we also take (5.15). The construction of the spectrum of a closed string propagating on such an orbifold proceeds along the same lines as explained for the twisted WZW model. A Kač–Moody symmetry is realized on the spectrum. The vertex operators in the various sectors can be explicitly constructed in terms of the scalar fields of the model, where the scalar field has boundary conditions twisted by powers of w. The difficult part of this construction is the determination of the zero mode sector. For details on the vertex operator construction we refer to Refs. 5–8.

For example, the contribution of the singlet HWM to the partition function $\mathscr{Z}^{(0,1)}$ in the $(r,s)=(0,1)$ sector is given by

$$c_w q^{\xi-l/24}\frac{\sum_{\lambda\in P_{(0)}(\Lambda_R)} q^{(1/2)|\lambda|^2}}{\prod_{j=0}^{N-1}\prod_{n\geqslant 1}(1-q^{n-j/N})^{d_j}} . \tag{5.16}$$

The denominator

$$\prod_{j=0}^{N-1}\prod_{n\geqslant 1}(1-q^{n-j/N})^{d_j}=\prod_{j\geqslant 1}(1-q^{j/N})^{d_{\bar j}}, \quad d_{\bar j}\equiv d_{j \bmod N}, \tag{5.17}$$

gives the contribution of the twisted oscillator modes $\alpha_{n-j/N}$ to the partition function. The sum in the numerator is the zero mode contribution. Here, $P_{(0)}(\Lambda_R)$ is the projection of the root lattice Λ_R on $h^*_{(0)}$. The mass shift of the ground state

$$\xi=\frac{1}{4}\sum_{j=0}^{N-1}\frac{j}{N}\left(1-\frac{j}{N}\right)d_j , \tag{5.18}$$

is most easily calculated by ζ-function regularization.[29] Finally the so-called *defect* value c_w gives the degeneration of the ground state. It can be computed by[5]

$$c_w=\left[\det_{h^{*}_{(0)}{}^{\perp}}(1-w)\right]^{1/2}\frac{\mathrm{vol}\,P_{(0)}(\Lambda_R)}{\mathrm{vol}\,\Lambda_R} . \tag{5.19}$$

Some remarks as to what extent different conjugacy classes in $\mathscr{W}(g)$ or $\mathrm{Aut}_0(g)$ give rise to different string models are in place. It is clear from the construction of the spectrum that, by combining the various twisted sectors, the string model only depends on the equivalence class of the cyclic subgroup. For two cyclic subgroups, generated by τ_1 and τ_2, respectively, to be conjugated it is necessary and sufficient that the generator τ_1 is conjugated to some power τ_2^j of the generator τ_2. [The conjugacy class of τ^j, for $\sigma\in\mathrm{Aut}_0(g)$, can also be calculated by means of the algorithm of Sec. IV.[22]] Two cyclic subgroups generated by elements τ_1, $\tau_2\in\mathscr{W}(g)$ are conjugated if and only if τ_1 and τ_2 themselves are conjugated, because of the rationality property (see Sec. IV). In $\mathrm{Aut}_0(g)$, however, it can certainly happen that two cyclic subgroups generated by $\tau_1,\tau_2\in\mathrm{Aut}_0(g)$ are conjugated while τ_1 is not in the same conjugacy class as τ_2. More serious is the fact that for two nonconjugated Weyl group elements w_1 and w_2 it can happen that their lifts $\tilde w_1$ and $\tilde w_2$ are conjugated in $\mathrm{Aut}_0(g)$, implying also that the corresponding string models are equivalent. This phenomenon can only occur in the exceptional groups $g=E_n$. We have for example $D_4(a_1)\sim 3A_1$.

The equivalence between the orbifold string model defined by a $w\in\mathscr{W}(g)$ and the level-1 WZW model twisted by the lift $\tilde w\in\mathrm{Aut}_0(g)$ of w can be concluded from the fact that the untwisted models are equivalent and by observing the analogy in the construction of the twisted spectrum out of the untwisted spectrum. If we compare the two ways of calculating the contribution of the basis HWM to the string partition function in the (0,1) sector, we find the following identity:

$$c_w q^{\xi - l/24} \frac{\sum_{\lambda\in P_{(0)}(\Lambda_R)} q^{(1/2)|\lambda|^2}}{\prod_{j>1}(1-q^{j/N})^{d_j}} = q^{-l/24} \frac{\sum_{\lambda\in\Lambda_R} q^{(1/2)|\lambda+\gamma_s|^2}}{\prod_{j>1}(1-q^j)^l}. \tag{5.20}$$

By modular transformations one can compute the corresponding identities in other twisted sectors. It is possible to give a rigorous proof of (5.20), without making use of the explicit form of the shift vector for a given Weyl group element.[9] The relation (5.20) contains a lot of previously known identities such as Macdonald's η function and Gauss and Jacobi identities. [In Ref. 13 the examples $[w] = 8A_1$ and $[w] = 4A_1$ in $g = E_8$ are treated in detail. In the first case we have $w^2 = \sigma_s^2 = 1$, $\dim h^*_{(0)} = 0$, $\dim h^*_{(1)} = 8$, $\xi = \frac{1}{2}$, and $c_w = 16$. The equality (5.20) is then based on Jacobi's triple product formula and Jacobi's "*æquatio identico satis abstrusa.*"]

Instead of looking at the complete identity (5.20), it is instructive to compare parts by expanding it in powers of q. Comparing the mass shift of the vacuum in the (0,1) sector we find the identity

$$\xi = \frac{1}{4N^2}\sum_{j=0}^{N-1} j(N-j)d_j = \frac{1}{4N^2}\sum_{j=1}^{l} m_j(N-m_j) = \frac{1}{2}|\gamma_s|^2, \tag{5.21}$$

where d_j and m_j correspond to the eigenvalues of $w\in\mathscr{W}$ as in Sec. II.

For the defect c_w, i.e., the degeneracy of the vacuum, (5.20) gives

$$c_w = \#\{\alpha\in\Delta | (\alpha,\gamma_s) = 1\} + 1 . \tag{5.22}$$

The possible values of the defect c_w have been given in Ref. 5. By inspection we find the following result (see Ref. 10 for E_8): If we denote the rhs of (5.22) by c_s and we associate the possible defect values b_i to the vertices of the extended Dynkin diagram of g, as indicated in Fig. 2, then

$$c_s = \min\{b_i | s_i \neq 0\} . \tag{5.23}$$

We have no proof of this other than a case by case verification, but it is clear that this observation explains the observation of Sec. II on the number of possible values of the defect c_w.

After a modular transformation $\tau\to -1/\tau$ of (5.20) the following identity is obtained[9]:

$$\frac{\sum_{\lambda\in(\Lambda_R)_{(0)}} q^{(1/2)|\lambda|^2}}{\prod_{j>1}\det_h(1-q^j w)} = \frac{\sum_{\lambda\in\Lambda_R} e^{2\pi i(\gamma_s,\lambda)} q^{(1/2)|\lambda|^2}}{\prod_{n>1}(1-q^n)^l}, \tag{5.24}$$

where $(\Lambda_R)_{(0)} = h^*_{(0)} \cap \Lambda_R$. Comparing the coefficient of the term linear in q we find for the trace

$$\operatorname{Tr}\sigma = l + \sum_{\alpha\in\Delta} e^{2\pi i(\gamma_s,\alpha)} = \operatorname{Tr} w + \#\{\alpha\in\Delta | w\alpha = \alpha\} . \tag{5.25}$$

It is these four properties; quasirationality, mass shift, defect, and the trace that allow for an easy determination of the shift vectors by exhaustion. This is the way in which we

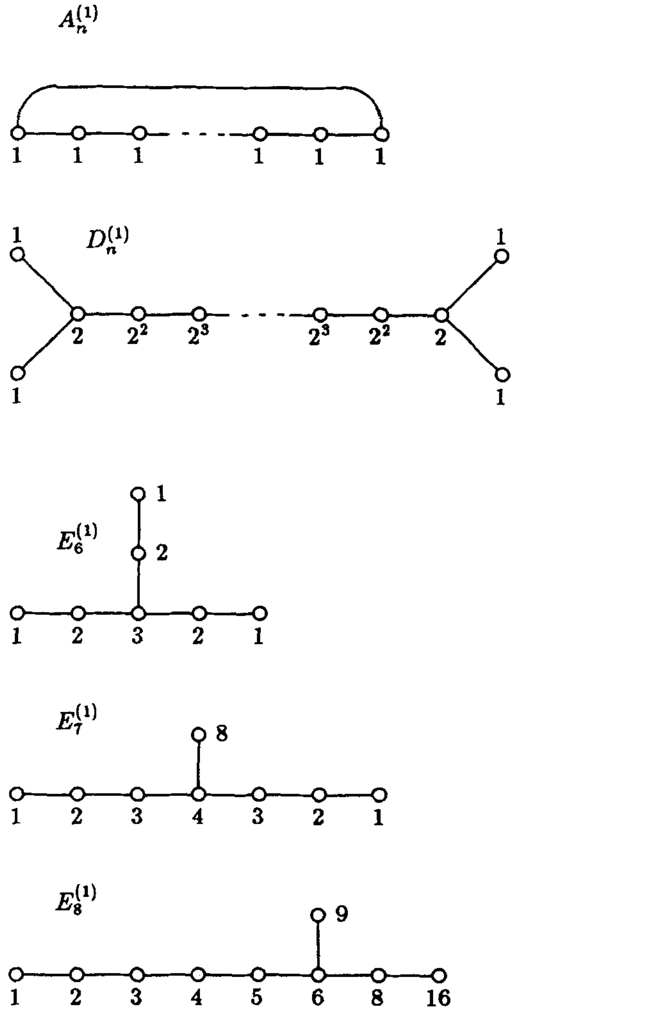


FIG. 2. The extended Dynkin diagrams of the simply laced Lie algebras with the labels b_i, from which the defects c_w can be computed [see (5.23)].

computed the shift vectors of the nonprincipal primitive elements.

We conclude this section by making some remarks on the moduli space of two-dimensional conformal field theories. For $c<1$ this moduli space is well understood. There exists a discrete series of allowed c values, and for every allowed c value there are only a finite number of CFT's. For $c=1$ there already exists a continuum of CFT's. Apart from a few discrete cases, the moduli space consists of two lines.[14] These lines are realized by a free scalar field compactified on a one-dimensional torus of arbitrary radius R, and a $\mathbf{Z}_2$ orbifold version thereof. The level-1 SU(2) WZW model and its twisted versions are all rational models on the torus line, and the Weyl group orbifold is the intersection point of the two branches. (A similar picture exists for the $\hat{c} = 1$, $N = 1$ superconformal moduli space.[30])

For $c>1$ the general situation is not well understood. The WZW model at a certain level k together with its twisted versions provide an infinite number of rational CFT's at $c = (k \dim g)/(k + h^\vee)$. Apart from taking tensor products these are in general the only CFT's which are well understood. Only for integer c the picture is more complete.[15] We have a continuum of torus compactified scalar fields,[12] and orbifold versions thereof. The equivalence of twisted WZW models and orbifold models as discussed in this paper gives intersection points (i.e., multicritical points) of these possible branches.

ACKNOWLEDGMENTS

I am indebted to Michiel Bos for drawing my attention to and for correspondence on Ref. 22. Furthermore, I would like to thank Sander Bais and Bert Schellekens for discussions.

I would like to thank the Stichting voor Fundamenteel Onderzoek der Materie (F.O.M.) for financial support.

APPENDIX: A_1 SUBALGEBRAS

In this Appendix we make some speculations on the existence of certain extended conformal algebras.[31] These speculations are based on the interesting observation that there seems to be an intimate relation between the classification of conjugacy classes in $\mathscr{W}(g)$ and equivalence classes of A_1 subalgebras of g.

An A_1 subalgebra is conveniently described in terms of a *defining vector* $\delta\in h^*$, such that the positive root $\hat{\alpha}$ of A_1 is given by $2\delta/(\delta,\delta)$.[17] The A_1-isospin value of a weight $\lambda\in h^*$ is given by $\frac{1}{2}(\lambda,\delta)$, as can be seen by projecting λ onto δ [in particular we have that $(\lambda,\delta)\in\mathbf{Z}$ for every weight λ of g]

$$\lambda\rightarrow\frac{(\lambda,\delta)}{(\delta,\delta)}\delta=\frac{1}{2}(\lambda,\delta)\hat{\alpha}\,. \tag{A1}$$

The Dynkin index[17] of the embedding $A_1\hookrightarrow g$ equals $j=\frac{1}{2}(\delta,\delta)$, Equivalence classes of A_1 subalgebras are classified by considering certain primitive A_1 subalgebras that are maximal in the regular subalgebras of g. Therefore we can calculate defining vectors of g if the primitive defining vectors in the regular subalgebras are known, using the same algorithm as for the determination of the shift vectors. The correspondence becomes even more striking when we consider the primitive A_1 subalgebra with the largest Dynkin index. This is called the *principal* A_1 subalgebra. Its defining vector is given by

$$(\delta,\alpha_i)=2\,,\quad i=1,\ldots,l, \tag{A2}$$

where the α_i are the simple roots of g. The properties of the principal A_1 subalgebra are hence remarkably similar to those of the Coxeter class in $\mathscr{W}(g)$. For simply laced Lie algebras g, Eq. (A2) implies that for the principal A_1 subalgebra $\delta=2\rho$. Hence, for all Weyl group elements in $\mathscr{W}(g)$ for which the graph Γ is a disconnected sum of Dynkin diagrams of *isomorphic* simple Lie algebras, the shift vector coincides up to normalization with the defining vector of the A_1 subalgebra, which is principal in the regular subalgebra described by Γ. Note, however, that this does not hold for general simple subalgebras because, in general, one needs different normalizations for the shift vectors in the simple constituents of Γ.

As an aside let us mention that Dynkin classified all A_1 subalgebras up to linear equivalence (i.e., on the level of branching rules). For classical Lie algebras this is the same as equivalence under conjugation. For the exceptional Lie algebras this is not necessarily so. By comparing Dynkin's lists with the list of conjugacy classes in $\mathscr{W}(g)$ the abovementioned correspondence seems to indicate that Dynkin's lists are incomplete if equivalence up to conjugation is considered. (We have made no attempt to prove this conjecture.) For every A_1 subalgebra of g (g not necessarily simply laced!), there exists an analog of the mass formula (5.21). Define thereto the set $\{m_i\}$ ($2m_i\in\mathbf{Z}_{>0}$) as the spins of the irreducible A_1 representations contained in the decomposition of the adjoint representation of g. Let δ be the defining vector of this A_1 subalgebra. We have

$$-\sum_i\,[-m_i\,][m_i+1]=2(\rho,\delta), \tag{A3}$$

where $[x]=\max\{n\in\mathbf{Z}|n\leqslant x\}$ denotes the integral part of x. The proof can be given by inserting (3.13) and using the definition of δ. (Compare Ref. 11 for a special case.) In particular, for the principal A_1 subalgebra the spins $\{m_i\}$ are precisely the exponents $\{e_j\}$, of g.[23] If, in addition, g is simply laced, one can use (A2), the Freudenthal–de Vries strange formula, and the relation $\dim(g)=l(h+1)$ to write (A3) as

$$\sum e_j(e_j+1)=\frac{1}{3}lh(h+1)\,. \tag{A4}$$

This result is exactly the same as what one would obtain from the mass formula (5.21) for the Coxeter class $[w_C]$.

In Ref. 31 (see also Ref. 32) Eq. (A4) was shown to be related to the BRST quantization of certain conformal field theories with extended symmetries based on the Casimirs of g. [In Refs. 31 and 32 a third-order polynomial relation in the exponents was found (see also Ref. 11). This relation is related to the computation of the Dynkin index of the principal A_1 embedding in g, by means of the branching rule for the adjoint representation of g. It can straightforwardly be generalized to all A_1 subalgebras of g. Because, unlike (A4), it does not seem to be related to any underlying conformal structure, we omit the formulas.] The set $\{e_j\}$ gives the conformal dimensions of the fields present in the symmetry algebra. It is tempting to suggest that a similar interpretation is possible either for all conjugacy classes in $\mathscr{W}(g)$ or all integral A_1 subalgebras of g. In fact, such an interpretation might be the clue to establish closure of the operator product algebra of these extended conformal field theories.

[1] I. Frenkel and V. G. Kač, Inv. Math. **62**, 23 (1980).
[2] G. Segal, Commun. Math. Phys. **80**, 301 (1981).
[3] P. Goddard and D. I. Olive, Int. J. Mod. Phys. A **1** 303 (1986).
[4] J. Wess and B. Zumino, Phys. Lett. **37**, 95 (1971); E. Witten, Commun. Math. Phys. **92**, 455 (1984).
[5] V. G. Kač and D. H. Peterson, *Symposium on Anomalies, Geometry and Topology*, Argonne, 1985, edited by W. A. Bardeen and A. R. White (World Scientific, Singapore, 1985).
[6] J. Lepowsky, Proc. Natl. Acad. Sci. USA **82**, 8295 (1985).
[7] D. Altschuler, Ph. Beran, J. Lacki, and I. Roditi, *Proceedings of the 10th Johns Hopkins Workshop* (World Scientific, Singapore, 1987).
[8] P. Sorba and B. Torresani, Int. J. Mod. Phys. A **3**, 1451 (1988).
[9] V. G. Kač, Adv. Math. **35**, 264 (1980).
[10] T. J. Hollowood and R. G. Myhill, Int. J. Mod. Phys. A **3**, 899 (1988).
[11] A. N. Schellekens and N. P. Warner, Nucl. Phys. B **308**, 397 (1988).
[12] K. S. Narain, Phys. Lett. B **169**,41 (1986); K. S. Narain, M. H. Sarmadi, and E. Witten, Nucl. Phys. B **279**, 93 (1986); S. Elitzur, E. Gross, E. Rabinovici, and N. Seiberg, *ibid.* **283**, 413 (1987).

[13]L. Dixon, J. A. Harvey, C. Vafa, and E. Witten, Nucl. Phys. B **261**, 93 (1986); **274**, 285 (1986).
[14]R. Dijkgraaf, E. Verlinde, and H. Verlinde, Commun. Math. Phys. **115**, 649 (1988); P. Ginsparg, Nucl. Phys. B **295** (FS21), 153 (1988).
[15]R. Dijkgraaf, E. Verlinde, and H. Verlinde, in *Proceedings of the 1987 Cargèse School on Nonperturbative Quantum Field Theory*, edited by G. 't Hooft *et al.* (Plenum, New York, 1988).
[16]R. W. Carter, Comp. Math. **25**, 1 (1972).
[17]Dynkin, Am. Math. Soc. Transl. **6**, 111 (1957).
[18]V. G. Kač, Funct. Anal. Appl. **3**, 252 (1969); S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces*, (Academic, New York, 1978).
[19]V. G. Kač, *Infinite Dimensional Lie Algebras* (Cambridge U.P., London, 1985).
[20]V. G. Kač and D. H. Peterson, Adv. Math. **53**, 125 (1984).
[21]H. Freudenthal and H. de Vries, *Linear Lie Groups* (Academic, New York, 1969).
[22]M. Bos, Math. thesis (in Dutch), Nijmegen, 1983.
[23]B. Kostant, Am. J. Math. **81**, 973 (1959); Adv. Math. **20**, 179 (1976).
[24]V. G. Kač, Adv. Math. **30**, 85 (1978).
[25]P. Bouwknegt, Amsterdam preprint, ITFA 87-08 Jan. 1987 (unpublished).
[26]D. Gepner and E. Witten, Nucl. Phys. B **278**, 493 (1986).
[27]P. Bouwknegt, Nucl. Phys. B **290** (FS20) 507 (1987); V. G. Kač and M. Wakimoto, Adv. Math. **70**, 156 (1988).
[28]K. Miki, Mod. Phys. Lett. A **3**, 191 (1988).
[29]C. Vafa, E. Witten, Phys. Lett. B **159**, 265 (1985).
[30]L. Dixon, P. Ginsparg, and J. Harvey, Nucl. Phys. B **306**, 470 (1988).
[31]F. A. Bais, P. Bouwknegt, K. Schoutens, and M. Surridge, Nucl. Phys. B **304**, 348 (1988).
[32]J. Thierry-Mieg, Phys. Lett. B **197**, 368 (1987); in *Proceedings of the 1987 Cargèse school on Nonperturbative Quantum Field Theory*, edited by G. 't Hooft *et al.* (Plenum, New York, 1988).

Octonions and subalgebras of the exceptional algebras

F. Buccella
Istituto di Fisica Teorica, Napoli, Italy
A. Della Selva and A. Sciarrino
Dipartimento di Fisica, Napoli, Italy and Istituto Nazionale de Fisica Nucleare, Sez di Napoli, Italy

(Received 25 June 1987; accepted for publication 19 October 1988)

The vector space of the generators of E_8 is realized in terms of 3×3 traceless matrices, two independent sets of octonions imaginary units, and the two G_2 acting on them. In this way one gets an appropriate framework to describe in a simple way how exceptional algebras and their fundamental representations transform under their subalgebras.

I. INTRODUCTION

The exceptional Lie algebras, in particular the largest of them E_8, recently came to the attention of the physicists involved in constructing unified theories including gravity starting from string theories in higher dimensions.[1] The relationship between the exceptional subalgebras and octonions has already been brought to the attention of physicists several years ago by Gunaydin and Gursey.[2] Here we want to exploit this relationship to build, explicitly, for the adjoint and the lowest-dimensional representations of the exceptional algebras, the vector spaces corresponding to their maximal subalgebras.

In Sec. II the properties of the octonionic algebra and the structure of the Lie algebra of their automorphisma G_2, the smallest exceptional group, are recalled.

In Sec. III the vector space of the E_8 generators is represented in terms of 3×3 matrices, two independent sets of imaginary octonionic units and the related G_2's.[3]

Also we show the vector spaces corresponding to the exceptional subalgebras of E_8 and the way exceptional algebras embed each other.

In Sec. IV we show the vector spaces corresponding to some peculiar classical subalgebras of E_8, including the maximal ones.

In Sec. V the behavior of the adjoint and fundamental representations of E_7, E_6, and F_4 with respect to their maximal classical subalgebras is discussed through an explicit construction of the vector spaces of these subalgebras.

II. OCTONIONS AND THE EXCEPTIONAL ALGEBRAS

The octonionic algebra is completely characterized by the product of the seven independent imaginary units[2]:

$$e_i e_j = -\delta_{ij} + c_{ijk} e_k \quad (i,j,k = 1,\ldots,7)\,. \tag{1a}$$

Here c_{ijk} is fully antisymmetric and has only the following nonvanishing elements:

$$c_{123} = c_{246} = c_{435} = c_{367} = c_{651} = c_{572} = c_{714} = 1\,. \tag{1b}$$

The set of automorphisms of the octonionic algebra form a group that is a subgroup of the SO(7) group of the rotations in the seven-dimensional space spanned by the e_i's. SO(7) is the group of invariance of the symmetric product of two octonions $e_i e_j + e_j e_i = -2\delta_{ij}$. Its Lie algebra can be specified by the 21 generators T_{ij} $(i,j = 1,\ldots,7)$,

$$T_{ij} = -T_{ji} = -T_{ij}^{+}\,, \tag{2}$$

while the Lie algebra corresponding to the group of automorphism leaving invariant equation (1) is the 14-dimensional algebra G_2 which can be written by the combinations of SO(7) generator[2]:

$$D_{ij}^{(r)} = (1/\sqrt{6})(2T_{ij} - T_{lm} - T_{kn})\,, \tag{3}$$

where, with fixed r, the three different couples (ij), (lm), (kn) are such that $c_{ijr} = c_{lmr} = c_{knr} = 1$. We shall omit from now on the upper index r.

The 21 D_{ij}'s satisfy the following seven equations:

$$\sum_{ij} c_{ijr} D_{ij} = 0\,. \tag{4}$$

By defining a scalar product in the space of the SO(7) generators

$$(T_{ij}, T_{hk}) = \delta_{ih}\delta_{jk} - \delta_{ik}\delta_{jh}\,, \tag{5}$$

one realizes that the D_{ij} are all orthogonal to the seven combinations:

$$\frac{1}{2\sqrt{3}} \sum_{ij} c_{ijk} T_{ij}\,. \tag{6}$$

Moreover we have

$$(D_{ij}, D_{ij}) = 1, \quad (D_{ij}, D_{hk}) = -\tfrac{1}{2}\,, \tag{7}$$

if $i \neq h$ and $c_{ijr} = c_{hkr}$, otherwise

$$(D_{ij}, D_{hk}) = 0\,.$$

The root diagram of G_2 consists of two hexagons with the same center, which correspond to the two-dimensional Cartan subalgebra of G_2, rotated by an angle of 30° with each other and with sides in the ratio $\sqrt{3}$.

The subalgebra of G_2 that does not act on one of the imaginary units, for instance, e_7, is SU(3) and the generators of G_2 decomposed with respect to it as

$$14 \to 8 + 3 + \bar{3}\,.$$

As G_2 has only real representations in the decomposition in terms of SU(3) representations only real combinations can appear. The SU(3) generators are the combinations of D_{ij} of Eq. (2) that do not contain any T_{k7}. The SU(3) generators that satisfy

$$[F_i, F_j] = f_{ijk} F_k \quad (i,j,k = 1,\ldots,8)\,, \tag{8}$$

where f_{ijk}, the usual Gell-Mann structure constants, are

 0022-2488/89/030585-09$02.50

$$F_1 = \sqrt{\tfrac{3}{2}}(D_{51} + \tfrac{1}{3}D_{24} + \tfrac{2}{3}D_{73}) ,$$
$$= \sqrt{\tfrac{1}{6}}(D_{51} - D_{24}) ,$$
$$F_2 = \sqrt{\tfrac{1}{6}}(D_{54} - D_{12}) ,$$
$$F_3 = \sqrt{\tfrac{1}{6}}(D_{25} - D_{14}) ,$$
$$F_4 = \sqrt{\tfrac{1}{6}}(D_{43} - D_{16}) , \tag{9}$$
$$F_5 = \sqrt{\tfrac{1}{6}}(D_{31} - D_{46}) ,$$
$$F_6 = \sqrt{\tfrac{1}{6}}(D_{62} - D_{35}) ,$$
$$F_7 = \sqrt{\tfrac{1}{6}}(D_{56} - D_{23}) ,$$
$$F_8 = \sqrt{\tfrac{1}{2}}D_{36} ,$$
$$(F_i,F_j) = \tfrac{1}{2}\delta_{ij} .$$

The other maximal subalgebra of G_2 is $SU(2)_L \otimes SU(2)_S$ spanned by the generators of G_2 that do not act on a quaternionic subalgebra of the octonionic algebra (for instance, the one spanned by e_3, e_6, and e_7) and by D_{pq} $(p,q = 3,6,7)$, respectively. By defining

$$T_{La} = F_a \quad (a = 1,2,3) ,$$
$$T_{S1} = \sqrt{\tfrac{3}{2}}D_{73} \quad T_{S2} = \sqrt{\tfrac{3}{2}}D_{67}, \quad T_{S3} = \sqrt{\tfrac{3}{2}}D_{36} , \tag{10}$$

the two SU(2) are defined by $(a,b,c = 1,2,3)$

$$[T_{La},T_{Lb}] = \epsilon_{abc}T_{Lc} , \quad [T_{Sa},T_{Sb}] = \epsilon_{abc}T_{Sc} ,$$
$$[T_{La},T_{Sb}] = 0 . \tag{11}$$

Remark that

$$(T_{La},T_{Lb}) = \tfrac{1}{2}\delta_{ab}, \quad (T_{Sa},T_{Sb}) = \tfrac{3}{2}\delta_{ab} . \tag{12}$$

With respect to $SU(2)_L \otimes SU(2)_S$ the generators of G_2 transform according to

$$14 \to (3,1) + (1,3) + (2,4) . \tag{13}$$

The seven imaginary octonionic units transform under G_2 and also under SO(7) as the seven-dimensional fundamental representation which, under the two maximal subalgebras, transforms as

$$7 \xrightarrow{SU(3)} 3 + \bar{3} + 1 ,$$
$$7 \xrightarrow{SU(2)_L \otimes SU(2)_S} (2,2) + (1,3) . \tag{14}$$

It is useful to decompose the seven imaginary units in the so-called split basis,[2] with definite properties under SU(3) and $SU(2)_L \otimes SU(2)_S$ described in Table I.

In the following we will denote collectively with e_+ (e_-) the split octonions transforming as a $3(\bar{3})$ under $SU(3)_7$.

The transformation properties under $SU(3)_7$ and $SU(2)_S \otimes SU(2)_L$ of the generators D_{ij} of Eq. (2) can be obtained from Table I according to the Wigner–Eckart (WE) theorem; for instance we have

$$T_{\pm L} = D_{(e_1 \pm ie_4)/\sqrt{2},(e_2 \mp ie_5)/\sqrt{2}} ,$$
$$T_{\pm S} = D_{(e_3 \pm ie_6)/\sqrt{2},e_7} . \tag{15}$$

TABLE I. $SU(3)$, T_{3L}, and T_{3S} assignments for the imaginary octonionic units in the split basis.

	$SU(3)_7$	T_{3L}	T_{3S}
$\epsilon_1 = e_1 + ie_4$		$\frac{1}{2}$	$\frac{1}{2}$
$\epsilon_2 = e_2 + ie_5$	3	$-\frac{1}{2}$	$\frac{1}{2}$
$\epsilon_3 = e_3 + ie_6$		0	-1
$\bar{\epsilon}_1 = e_1 - ie_4$		$-\frac{1}{2}$	$-\frac{1}{2}$
$\bar{\epsilon}_2 = e_2 - ie_5$	$\bar{3}$	$\frac{1}{2}$	$-\frac{1}{2}$
$\bar{\epsilon}_3 = e_3 - ie_5$		0	1
$\sqrt{2}e_7$	1	0	0

III. REALIZATION OF THE EXCEPTIONAL ALGEBRAS

The algebra of E_8 can be written in terms of 3×3 antisymmetric (**A**) and symmetric traceless (**S**) matrices, two independent sets of imaginary octonionic units (e_i and e'_i, $i = 1,\ldots,7$), and their corresponding automorphism algebras G_2 and G'_2.

The vector space of E_8 can be spanned by[3]

$$A \oplus G_2 \oplus S \otimes e_i \oplus A \otimes e_i \otimes e'_i \oplus S \otimes e'_i \oplus G'_2 \tag{16}$$

of dimensions

$$3 + 14 + 5\times 7 + 3\times 7\times 7 + 5\times 7 + 14 = 248 .$$

Equation (16) defines a Lie algebra wth respect to the following composition law: (i) $D_{ij}(D'_{ij})$ act on the corresponding octonionic units according to Eq. (2); (ii) the matrix **A** acts only on the matrices **A** and **S** according to the usual commutator;

(iii) $$[S_a \otimes e_i, S_b \otimes e_j] = \{S_a,S_b\} \otimes e_ie_j - \delta_{ab}\sqrt{\tfrac{2}{3}}D_{ij} - \delta_{ij}[S_a,S_b] , \tag{17a}$$

(iv) $$[S_a \otimes e_i, A_l \otimes e_j \otimes e'_h] = \{S_a,A_l\} \otimes e_ie_j \otimes e'_h - \delta_{ij}[S_a,A_l] \otimes e'_h , \tag{17b}$$

(v) $$[S_a \otimes e_i, S_b \otimes e'_j] = [S_a,S_b] \otimes e_i \otimes e'_j , \tag{17c}$$

(vi) $$[A_l \otimes e_i \otimes e'_h, A_m \otimes e_j \otimes e'_k] = -\{A_l,A_m\} \otimes (\delta_{ij}e'_he'_k + \delta_{hk}e_ie_j) - \delta_{lm}\sqrt{\tfrac{2}{3}}(\delta_{ij}D'_{hk} + \delta_{hk}D_{ij}) + [A_l,A_m](\delta_{ij}\delta_{hk} + \otimes e_ie_j \otimes e'_he'_k) , \tag{17b}$$

(vii) the rules one can find with the exchange $e_i \leftrightarrow e'_i$ ($\{S_a,S_b\}$ and $\{A_l,A_m\}$ are the traceless part of the anticommutator).

One can show that Eq. (16) with composition law (17) defines an algebra that satisfies the Jacobi identity. It is a simple algebra with 248 generators and one can see from Eq. (17) that its rank is 8, in fact at most eight generators commute each other.

The algebras E_7, E_6, and F_4 can be obtained from Eqs. (16) and (17) as follows.

(a) Restricting one of the set of octonionic units to the imaginary units of a quaternionic subalgebra (for instance, e'_3, e'_6, and e'_5). In this case the corresponding automorphism algebra G'_2 restricts to the automorphism algebra of the quaternions $SU(2)_Q$. The explicit form of the vector space of E_7 is $(p = 3,6,7)$

$$A \oplus G_2 \oplus S \otimes e_i \oplus A \otimes e_i \otimes e'_p \oplus S \otimes e'_p \oplus SU(2)'_Q \,, \qquad 3 + 14 + 5\times 7 + 3\times 7\times 3 + 5\times 3 + 3 = 133 \,. \tag{18}$$

(b) Restricting one of the sets of the octonionic units to one imaginary unit [of a complex subalgebra (for instance, e'_7)]. In this case G'_2 does not appear any more. The explicit form of vector space for E_6 is

$$A \oplus G_2 \oplus S \otimes e_i \oplus A \otimes e_i \otimes e'_7 \oplus S \otimes e'_7 \,, \qquad 3 + 14 + 5\times 7 + 3\times 7 + 5 = 78 \,. \tag{19}$$

(c) Suppressing a whole set of octonionic units and the corresponding automorphism algebra (for instance G'_2). The explicit form of the vector space of F_4 is

$$A \oplus G_2 \oplus S \otimes e_i \,, \quad 3 + 14 + 5\times 7 = 52 \,. \tag{20}$$

The fact that in cases (b) and (c) G'_2 completely disappears is a reflex of the fact that the real and complex numbers have no continuous symmetry with respect to their products.

Equations (16) and (18)–(20) give an explicit expression for the generators of the exceptional algebras that allow us to discuss in a relatively simple and straightforward way the embedding of the maximal subalgebras and the decomposition of the adjoint representations in terms of representation of subalgebras.

Let us remark that the following results, already known, are usually derived by using the properties of the root and weight spaces,[4] while we use only the peculiar parametrization of the generators.

In the following sometimes we will make explicit use of a basis for the matrices $A \equiv (i\lambda_2, i\lambda_5, i\lambda_7)$ and $S \equiv (\lambda_1, \lambda_3, \lambda_4, \lambda_5, \lambda_8)$, where λ_i are the usual Gell-Mann matrices.[5]

For the octonionic imaginary units we shall write e_i or e_j $(i,j = 1,...,7)$, e_h or e_k $(h,k = 1,...,6)$, e_p or e_q $(p,q = 3,6,7)$, and e_r or e_s $(r, s = 1,2,4,5)$.

The parametrization of the vector space of E_8 given by Eq. (16) allows us to see easily the way exceptional algebras are embedded in each other. In fact, the E_8 generators that commute with the E_7 of Eq. (18) build clearly a $SU(2)'_L$; the remaining generators defining $E_8/E_7\times SU(2)'_L$ transform as doublets of the form

$$(A \otimes e_i \oplus S) \otimes e'_r \oplus G'_2/SU(3)'_7 \tag{21}$$

under $SU(2)'_L$ and as a 56 under E_7. Thus we have

$$248 \xrightarrow{E_7 \otimes SU(2)} (133,1) + (1,3) + (56,2) \,. \tag{22}$$

Similarly the generators of E_8 commuting with E_6 of Eq. (19) build a $SU(3)'_7$ and the generators of $E_8/E_6\times SU(3)'_7$,

$$(A \otimes e_i \oplus S) \otimes e'_h \oplus G'_2/SU(3)'_7 \,, \tag{23}$$

transform as $(27,3) + (\overline{27,3})$ under $E_6\times SU(3)$; thus we get

$$248 \xrightarrow{E_6 \otimes SU(3)} (78,1) + (1,8) + (27,3) + (\overline{27},\overline{3}) \,. \tag{24}$$

The generators commuting with the F_4 of Eq. (20) build the subalgebra G'_2 and the generators of $E_8/F_4\times G'_2$,

$$(A \otimes e_i \oplus S) \otimes e'_j \,, \tag{25}$$

transform as a (26,7) under $F_4 \otimes G'_2$; thus we get

$$248 \xrightarrow{F_4 \otimes G'_2} (52,1) + (1,14) + (26,7) \,. \tag{26}$$

In a similar way one can discuss the embedding of the other exceptional algebras and get the scheme reported in Table II.

IV. CLASSICAL SUBALGEBRAS OF E_8

It is useful to start by considering some particular subalgebras of E_8 beyond the ones considered previously, in order to explicitly construct the maximal classical subalgebras of the exceptional algebras in terms of the vector spaces introduced in Eq. (16).

They are as follows.

(i) $SU(3)_G \equiv A \oplus S \otimes e_7$, which is also contained in F_4 and commutes with $SU(3)_7$ and G'_2. The residual generators of E_8 transform under $SU(3)_G \otimes SU(3)_7 \otimes G'_2$ as

$$\begin{aligned} &G_2/SU(3)_7 \oplus S \otimes e_h\colon\ (6,3,1) + (\overline{6},\overline{3},1) \,, \\ &(S \oplus A \otimes e_7) \otimes e'_i\colon\ (8,1,7) \,, \\ &A \otimes e_h \otimes e'_i\colon\ (3,\overline{3},7) + (\overline{3},3,7) \,. \end{aligned} \tag{27}$$

(ii) $SU(6) \equiv SU(2)_S \oplus SU(3)'_G \equiv (A \oplus S \otimes e'_7) \oplus [S \oplus A \otimes e'_7] \otimes e_q \,,$

TABLE II. Embedding of the exceptional algebras in each other and vector spaces of the residual generators. The $SU(2)_w$ is spanned by the A's, introduced in Eq. (16).

E_8	$E_7 \otimes SU(2)'_L$ $(A\otimes e_i \oplus S)\otimes e'_r \oplus \dfrac{G'_2}{SU(2)'_L \otimes SU(2)'_S}$ $(56,2)\,(r = 1,2,4,5)$	$E_6 \otimes SU(3)'_7$ $(A\otimes e_i \oplus S)\otimes e'_h \oplus \dfrac{G'_2}{SU(3)'_7}$ $(27,3) + (\overline{27,3})\,(h \neq 7)$	$F_4 \otimes G'_2$ $(A\otimes e_i \oplus S)\otimes e'_j$ $(26,7)$
E_7	$E_6 \otimes T'_{3S}$ $(A\otimes e_i \oplus S)\otimes(e'_3 \oplus e'_6)\oplus T'_{\pm S}$ $(27,-1) + (\overline{27},+1)$	$F_4 \otimes SU(2)'_S$ $(A\otimes e_i \oplus S)\otimes e'_p$ $(26,3)\,(p = 3,6,7)$	$G_2 \otimes Sp(6)$ $e_i \otimes (A\otimes e'_p \oplus S)$ $(7,14)\,(p = 3,6,7)$
E_6	F_4 $(A\otimes e_i \oplus S)\otimes e'_7$ 26	$G_2 \otimes SU(3)'_7$ $e_i \otimes (A\otimes e'_7 \oplus S)$ $(7,8)$	
F_4	$G_2 \otimes SU(2)_w$ $e_i \otimes S$ $(7,5)$		

which decomposes under $SU(2)_S \otimes SU(3)'_G$ *à la* Gursey and Radicati[6]: (3,1) + (1,8) + (3,8). It is contained in E_6 and commutes with $SU(2)_L \otimes SU(3)'_7$. The residual generators of E_8 transform under $SU(6) \otimes SU(2)_L \otimes SU(3)'_7$ as

$$\frac{G_2}{SU(2)_L \otimes SU(2)_S} \oplus [S \oplus A \otimes e'_7] \otimes e_r:(20,2,1)\,,$$

$$S \oplus (A \otimes e_q) \otimes e'_h \oplus \frac{G'_2}{SU(3)'_7}:(15,1,3) + (\overline{15},1,\bar{3})\,, \tag{28}$$

$A \otimes e_r \otimes e'_k:(\bar{6},2,3) + (6,2,\bar{3})$.

(iii) $SU(8) \equiv SU(6) \oplus SU(2)_L \oplus T'_{3S} \oplus A \otimes e_r \otimes (e'_3 \oplus e'_6)$.

It is contained in E_7; therefore it commutes with $SU(2)'_L$. The residual generators transform under $SU(8) \otimes SU(2)'_L$ as [we write also for each set of generators their behavior under the $SU(6) \otimes SU(2)_L \otimes U(1)_{T'_{3S}}$ subalgebra of $SU(8)$]

	(70,1)	
$\frac{G_2}{SU(2)_S \otimes SU(2)_L} \oplus [S \oplus A \otimes e'_7] \otimes e_r$	(20,2,0)	
$(S \oplus A \otimes e_q) \otimes (e'_3 \oplus e'_6) \oplus T'_{\pm S}$	$(15,1,-1) \oplus (\overline{15},1,+1)$	
	(28,2)	$(\overline{28},2)$
$(S \oplus A \otimes e_p) \otimes e'_r \oplus D'_{7S}$	$(15,1,+\frac{1}{2})$	$(\overline{15},1,-\frac{1}{2})$
$A \otimes e_r \otimes e'_s$	$(6,2,-\frac{1}{2})$	$(\bar{6},2,+\frac{1}{2})$
$D'_{3 \pm i6,2 \pm i5} \oplus D'_{3 \pm i6,1 \pm i4}$	$(1,1,-\frac{3}{2})$	$(1,1,+\frac{3}{2})$.

(29)

It is interesting to remark that the 70 and the $28 \otimes \overline{28}$ are the fourth- and second-order antisymmetric tensors, respectively. In conclusion, E_8 decomposes under $SU(8) \otimes SU(2)'_L$ as

$$248 \to (63 + 70,1) + (28 + \overline{28},2) + (1,3)\,. \tag{30}$$

(iv) $SU(9) \equiv SU(6) \oplus SU(3)'_7 \oplus T_{3L} \oplus A \otimes \{[(e_1 + ie_4) \oplus (e_2 - ie_5)] \otimes e'_- \oplus [(e_1 - ie_4) \oplus (e_2 + ie_5)] \otimes e'_+\}$.

The residual generators transform as the third-order antisymmetric tensors $84 + \overline{84}$ [the behavior under the $SU(6) \otimes SU(3)'_7 \otimes U(1)T_{3L}$ subalgebra of $SU(9)$ is also written].

	84	$\overline{84}$
$\frac{G_2}{SU(2)_L \otimes SU(2)_S} \oplus [S \oplus A \otimes e'_7] \otimes e_r$	$(20,1,+\frac{1}{2})$	$(20,1,-\frac{1}{2})$
$(S \oplus A \otimes e_q) \otimes e'_h \oplus \frac{G'_2}{SU(3)'_7}$	(15,3,0)	$(\overline{15},\bar{3},0)$
$A \otimes \{[(e_1 - ie_4) \oplus (e_2 + ie_5)] \otimes e'_- \oplus [(e_1 + ie_4) \otimes (e_2 - ie_5)] \otimes e'_+\}$	$(6,\bar{3},-\frac{1}{2})$	$(\bar{6},3,+\frac{1}{2})$
$T_{\mp L}$	$(1,1,-1)$	$(1,1,+1)$.

(31)

(v) $SO(9) \equiv G_2 \oplus i\lambda_2 \oplus (\lambda_1 \oplus \lambda_3 \oplus \lambda_8) \otimes e_i$,

which is contained in F_4 and commutes with $SU(2)'_A \otimes SU(2)'_B \otimes SU(2)'_L$, where

$$\begin{aligned} SU(2)_A &\equiv \left[\sqrt{\tfrac{2}{3}}D_{67} - (1/\sqrt{3})\lambda_8 \otimes e_3\right] \\ &\oplus \left[\sqrt{\tfrac{2}{3}}D_{73} - (1/\sqrt{3})\lambda_8 \otimes e_6\right] \\ &\oplus \left[\sqrt{\tfrac{2}{3}}D_{36} - (1/\sqrt{3})\lambda_8 \otimes e_7\right] \\ &\equiv \left[\tfrac{2}{3}T_{Sq} - (1/\sqrt{3})\lambda_8 \otimes e_q\right]\,, \\ SU(2)_B &\equiv \tfrac{1}{3}T_{Sq} + (1/\sqrt{3})\lambda_8 \otimes e_q\,. \end{aligned} \tag{32}$$

The residual generators transform under $SO(9) \otimes SU(2)'_A \otimes SU(2)'_B \otimes SU(2)'_L$ as

$$\begin{aligned} &(\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_i) \otimes e'_q \ \ (9,3,1,1), \\ &i\lambda_5 \oplus i\lambda_7 \oplus (\lambda_4 \oplus \lambda_6) \otimes (e_i \oplus e'_q) \oplus (i\lambda_5 \oplus i\lambda_7) \\ &\quad \otimes e_i \otimes e'_q \ \ (16,2,2,1)\,, \\ &[\lambda_4 \oplus \lambda_6 \oplus (i\lambda_5 \oplus i\lambda_7) \otimes e_i] \otimes e'_r \ \ (16,2,1,2)\,, \\ &(\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_i) \otimes e'_r \ \ (9,1,2,2)\,, \\ &\frac{G'_2}{SU(2)'_S \otimes SU(2)'_L} \oplus \lambda_8 \otimes e'_r (1,3,2,2)\,. \end{aligned} \tag{33}$$

Let us remark that (a) the generators of $SO(9) \otimes SU(2)'_B \otimes SU(2)_L$, and the generators tranforming as (9,1,2,2) [fourth row of Eq. (33)] build up an $SO(13)$ that commutes with $SU(2)'_A$; (b) $SU(2)'_B$ is isomorphic to

$SU(2)'_L$; (c) $SU(2)'_A$ is isomorphic to the $SU(2)_w$ generated by $i\lambda_2 \oplus (\lambda_1 \oplus \lambda_3) \otimes e'_7$; (d) $G_2 \oplus \lambda_8 \otimes e_i$ and $G_2 \oplus (\lambda_3 \oplus \lambda_8) \otimes e_i$ build up, respectively, an SO(7) and SO(8) algebra [$SO(7) \subset SO(8) \subset SO(9)$].

(vi) $SO(12) \equiv SU(2)_S \oplus SU(2)'_S$

$$\oplus A \oplus S \otimes (e_q \oplus e'_p) \oplus A \otimes e_q \otimes e'_p ,$$

which is contained in E_7 and commutes with $SU(2)_L \otimes SU(2)'_L$. The other generators transform under $SO(12) \otimes SU(2)_L \otimes SU(2)'_L$ as

$$\begin{aligned} &\frac{G_2}{SU(2)_S \otimes SU(2)_L} \oplus (S \oplus A \otimes e'_q) \otimes e_r\colon\ (32,2,1) ,\\ &\frac{G'_2}{SU(2)'_S \otimes SU(2)'_L} \oplus (S \oplus A \otimes e_p) \otimes e'_s\colon\ (32',1,2) ,\\ &A \otimes e_r \otimes e'_s\colon\ (12,2,2) . \end{aligned} \tag{34}$$

This SO(12) is isomorphic to the one generated by

$$\begin{aligned} &SO(9) \oplus SU(2)'_A \oplus (\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_i) \otimes e'_q \\ &\quad \equiv G_2 \oplus i\lambda_2 \oplus (\lambda_1 \oplus \lambda_3) \otimes (e_i \oplus e'_q) \\ &\qquad \oplus i\lambda_2 \otimes e_i \otimes e'_p \oplus \lambda_8 \otimes e_i \oplus SU(2)'_A . \end{aligned}$$

(vii) The generators

$$SO(12) \oplus SU(2)_L \oplus SU(2)'_L \oplus A \otimes e_r \otimes e'_s$$

build up a SO(16) with the remaining generators transforming as the spinorial 128. By rotating by 120° in the weight space the e''s, i.e., replacing (e'_3, e'_6, e'_7) by (e'_1, e'_4, e'_7), and the corresponding G'_2 one gets an SO(16) which contains the SU(8) defined in (iii) as one can see explicitly:

$$\begin{aligned} &SU(2)_S \oplus SU(2)_L \oplus A \oplus S \otimes (e_3 \oplus e_6 \oplus e_7 \oplus e'_1 \oplus e'_4 \oplus e'_7) \\ &\quad \oplus A \otimes (e_1 \oplus e_2 \oplus e_4 \oplus e_5) \otimes (e'_2 \oplus e'_3 \oplus e'_5 \oplus e'_6) \\ &\quad \oplus T'_{3S} \oplus T'_{3L} \oplus D'_{3 \mp i6, 2 \mp i5} \oplus D'_{7,1 \mp i4} \oplus A \\ &\quad \otimes (e_3 \oplus e_6 \oplus e_7) \otimes (e'_1 \oplus e'_4 \oplus e'_7) . \end{aligned} \tag{35}$$

(viii) $SO(16) \equiv G_2 \oplus G'_2 \oplus i\lambda_2 \oplus (\lambda_1 \oplus \lambda_3 \oplus \lambda_8)$

$$\otimes (e_i \oplus e'_i) \oplus i\lambda_2 \otimes e_i \otimes e'_i$$

with the other generators

$$i\lambda_5 \oplus i\lambda_7 \oplus (\lambda_4 \oplus \lambda_6) \otimes (e_i \oplus e'_i) \oplus (i\lambda_5 \oplus i\lambda_7) \otimes e_i \otimes e'_i$$

transforming as the spinorial 128. This SO(16) is isomorphic to the two SO(16) defined in (vii). It contains $SO(13) \otimes SU(2)'_A$ where SO(13) is given by

$$SO(9) \oplus SU(2)'_B \oplus SU(2)'_L \oplus (\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_i) \otimes e'_r$$

[see (v)].

(ix) $SU(5) \otimes SU(5)$—with the two SU(5) generated, respectively, by

$$\begin{aligned} &\underset{(8,1,0)}{SU(3)_7} \oplus \underset{+\,(1,3,0)}{(\lambda_1 \oplus \lambda_3) \otimes (e_7 + e'_7) \oplus i\lambda_2 \otimes (-1 + e_7 \otimes e'_7)} \\ &\oplus \underset{(1,1,0)}{(1/\sqrt{10})\left[\sqrt{3}\lambda_3 \otimes (e_7 - e'_7) + 2\lambda_8 \otimes e_7\right]} \\ &\oplus \underset{(3,2,-\frac{5}{6}) + (\bar{3},2,+\frac{5}{6})}{(\lambda_1 \pm \lambda_2 \otimes e'_7) \otimes e_\pm \oplus \left\{\sqrt{\tfrac{2}{3}}D_{7,\pm} \pm i[\lambda_3 + (\lambda_8/\sqrt{3})] \otimes e_\pm\right\}} \end{aligned} \tag{36a}$$

[we write the transformation properties under $SU(3)_7 \otimes SU(2) \otimes U(1)$] and by

$$\begin{aligned} &SU(3)'_7 \oplus \{\lambda_6 \oplus [\tfrac{1}{2}\lambda_3 + (\sqrt{3}/2)\lambda_8]\} \otimes (e'_7 - e_7) \oplus i\lambda_7 \otimes (1 + e_7 \otimes e'_7) \oplus (1/2\sqrt{10})\left[(3\lambda_8 - \sqrt{3}\lambda_3) \otimes e_7 + (5\lambda_8 + \sqrt{3}\lambda_3) \otimes e'_7\right] \\ &\quad \oplus (\lambda_6 \pm \lambda_7 \otimes e_7) \otimes e'_\pm \oplus \sqrt{\tfrac{2}{3}}(D'_{7,\pm} \mp i\sqrt{2}\lambda_8 \otimes e'_\pm) . \end{aligned} \tag{36b}$$

The residual generators transform under $SU(5) \otimes SU(5)$ as $(10,5) + (\overline{10},\bar{5}) + (\bar{5},10) + (5,\overline{10})$.

(ix) $Sp(6) \equiv SU(2)_S \oplus A \oplus S \otimes e_p$.

It is contained in F_4 and commutes with $SU(2)_L \otimes G'_2$. The residual generators transform under $Sp(6) \otimes SU(2)_L \otimes G'_2$ as

$$\begin{aligned} &\frac{G_2}{SU(2)_L \otimes SU(2)_S} \oplus S \otimes e_r\ \ (14',2,1) ,\\ &(S \oplus A \otimes e_p) \otimes e'_i\ \ (14,1,7) ,\\ &A \otimes e_r \otimes e'_i\ \ (6,2,7) , \end{aligned} \tag{37}$$

(x) $Sp(8) \equiv Sp(6) \oplus SU(2)_L \oplus A \otimes e_r \otimes e'_7$.

It is contained in E_6 and commutes with $SU(3)'_7$. The residual generators transform under $Sp(8) \otimes SU(3)'_7$ as

$$\frac{G_2}{SU(2)_S \otimes SU(2)_L} \oplus S \otimes e_r \oplus A \otimes e_p \otimes e'_7\ \ (42,1) ,$$

$$(S \oplus A \otimes e_i) \otimes e'_h \oplus \frac{G'_2}{SU(3)'_7}\ \ (27,3+\bar{3}) . \tag{38}$$

The vector spaces corresponding to the subalgebras of E_8 introduced in this section are written in Table III in such a way to show how they are contained in each other.

V. CLASSICAL MAXIMAL SUBALGEBRAS OF E_7, E_6, AND F_4

We have already shown how the 248 of E_8 transforms under its classical maximal subalgebras SU(9), SO(16), and $SU(5) \otimes SU(5)$. Let us now consider the behavior of the adjoint and the fundamental representations of the other exceptional groups under their classical maximal subalgebras.

For E_7 they are SU(8), $SU(6)' \otimes SU(3)_7$, and $SO(12) \otimes SU(2)_L$.

One should compare the decomposition of the 248 into $(E_7, SU(8), SU(6)' \otimes SU(3)_7, SO(12) \otimes SU(2)_L) \otimes SU(2)'_L$ given by Eqs. (22), (30), (28), and (34), respectively:

TABLE III. Subalgebras of E_8 defined in Sec. IV. The explicit form of the underlined algebras is given.

G_2	$SU(3)_7$		
F_4	$\underline{SU(3)_G} \otimes SU(3)_7$ $A \oplus S \otimes e_7$	$\underline{SO(9)}$ $G_2 \oplus (\lambda_1 \oplus \lambda_3 \oplus \lambda_8) \otimes e_i \oplus i\lambda_2$	
E_6	$\underline{SU(3)_L \otimes SU(3)_R} \otimes SU(3)_7$ $A \oplus S \otimes (e_7 \oplus e'_7) \oplus A \otimes e_7 \otimes e'_7$	$\underline{SO(10)} \otimes U(1)$ $SO(9) \oplus (\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_i) \otimes e'_7$ $G_2 \oplus (\lambda_1 \oplus \lambda_3 \oplus \lambda_8) \otimes (e_i \oplus e'_7) \oplus i\lambda_2 \oplus i\lambda_2 \otimes e_i \otimes e'_7$	
E_7	$\underline{SU(6)} \otimes SU(3)_7$ $SU(3)_G \oplus SU(2)'_S \oplus (S \oplus A \otimes e_7) \otimes e_r$	$\underline{SO(12)} \otimes SU(2)'_B$ $SO(10) \oplus (\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_i) \otimes (e'_3 \oplus e'_6) \oplus SU(2)'_A$ $G_2 \oplus SU(2)'_S \oplus (\lambda_1 \oplus \lambda_3 \oplus \lambda_8) \otimes (e_i \oplus e'_p) \oplus i\lambda_2 \oplus i\lambda_2 \otimes e_i \otimes e'_p$	
E_8	$\underline{SU(9)}$ $SU(6)' \oplus T_{3L} \oplus A \otimes \{e_- \otimes [(e'_1 + ie'_4) \oplus$ $(e'_2 - ie'_5)] \oplus e_+ \otimes [(e'_1 - ie'_4) \oplus (e'_2 + ie'_5)]\}$ $\oplus SU(3)_7$	$\underline{SO(16)}$ $G_2 \oplus G'_2 \oplus (\lambda_1 \oplus \lambda_3 \oplus \lambda_8) \otimes (e_i \oplus e'_j) \oplus i\lambda_2 \oplus i\lambda_2 \otimes e_i \otimes e'_j$	
G_2		$SU(2)_S \otimes SU(2)_L$	
F_4	$\underline{Sp(6)} \otimes SU(2)_L$ $\overline{SU(2)_S \oplus}$ $\oplus A \oplus S \otimes e_p$		
E_6	$\underline{Sp(8)}$ $SU(2)_S \oplus \overline{SU(2)_L} \oplus A \oplus$ $S \otimes e_p \oplus A \otimes e_r \otimes e'_3$	$\underline{SU(6)} \otimes SU(2)_L$ $\overline{SU(2)_S} \oplus A \oplus$ $S \otimes (e_p \oplus e'_7) \oplus A \otimes e_p \otimes e'_7$	
E_7	$\underline{SU(8)}$ $SU(2)_S \oplus \overline{SU(2)_L} \oplus T'_{3S} \oplus$ $A \oplus S \otimes (e_p \oplus e'_7) \oplus A \otimes e_r \otimes (e'_3 \oplus e'_6)$ $\oplus A \otimes e_p \otimes e'_7$		$\underline{SO(12)} \otimes SU(2)_L$ $\overline{SU(2)_S \oplus SU(2)'_S}$ $\oplus A \oplus S \otimes (e_p \oplus e'_q) \oplus A \otimes e_p \otimes e'_q$
E_8	$\underline{SO(16)}$ $SU(2)_S \oplus SU\overline{(2)_L \oplus T'_{3S}} \oplus D'_{7.1 \mp i4} \oplus$ $T'_{3L} \otimes D'_{3 \mp i6.2 \mp i5} \oplus$ $A \oplus S \otimes (e_p \oplus e'_1 \oplus e'_4 \oplus e'_7)$ $\oplus A \otimes e_r \otimes (e'_2 \oplus e'_3 \oplus e'_5 \oplus e'_6)$ $\oplus A \otimes e_p \otimes (e'_1 \oplus e'_4 \oplus e'_7)$		$\underline{SO(16)}$ $SU(2)_S \oplus SU\overline{(2)_L \oplus SU}(2)'_S \oplus SU(2)'_L$ $\oplus A \oplus S \otimes (e_p \oplus e'_q) \oplus A \otimes e_r \otimes e'_s$ $\oplus A \otimes e_p \otimes e'_q$

$$
248 \to \begin{cases}
E_7 \otimes SU(2)'_L: & (133,1) + (56,2) + (1,3)\,, \\
SU(8) \otimes SU(2)'_L: & (63 + 70,1) + (28 + \overline{28},2) + (1,3)\,, \\
SU(6)' \otimes SU(3)_7 \otimes SU(2)'_L: & (35,1,1) + (1,8,1) + (1,1,3) + (20,1,2) \\
 & + (15,3,1) + (\overline{15},\overline{3},1) + (\overline{6},3,2) \\
 & + (6,\overline{3},2)\,, \\
SO(12) \otimes SU(2)_L \otimes SU(2)'_L: & (66,1,1) + (1,3,1) + (1,1,3) \\
 & + (32,2,1) + (32',1,2) + (12,2,2)\,,
\end{cases} \tag{39}
$$

by identifying the terms that are singlets or doublets under $SU(2)'_L$ one gets

$$
133 \to \begin{cases}
SU(8): & 63 + 70\,, \\
SU(6)' \otimes SU(3)_7: & (35,1) + (1,8) + (15,3) + (\overline{15},\overline{3})\,, \\
SO(12) \otimes SU(2)_L: & (66,1) + (1,3) + (32,2)\,,
\end{cases} \tag{40}
$$

$$
56 \to \begin{cases}
SU(8): & 28 + \overline{28} \\
SU(6)' \otimes SU(3)_7: & (20,1) + (\overline{6},3) + (6,\overline{3})\,, \\
SO(12) \otimes SU(2)_L: & (32,1) + (12,2)\,.
\end{cases} \tag{41}
$$

The transformation properties of the adjoint and fundamental representations of E_6 under its maximal and classical subalgebras $SU(6) \otimes SU(2)_L$ and $SO(10) \otimes U(1)$ may be derived by comparing the decomposition of the 248 into

$$\{E_6, SU(6) \otimes SU(2)_L, SO(10) \otimes U(1)\} \otimes SU(3)'_7$$

given the first two in Eqs. (24) and (28),

$$248 \to \begin{cases} E_6 \otimes SU(3)'_7: \ (78,1) + (1,8) + (27,3) + (\overline{27},\bar{3}), \\ SU(6) \otimes SU(2)_L \otimes SU(3)'_7: \ (35,1,1) + (1,3,1) + (1,1,8) + (20,2,1) \\ \quad + (15,1,3) + (\overline{15},1,\bar{3}) + (\bar{6},2,3) + (6,2,\bar{3}), \\ SO(16) \to SO(10) \otimes SO(6) \to SO(10) \otimes U(1)_{Y'} \otimes SU(3)'_7: \\ \quad (45,0,1) + (1,0,8) + (1,+\frac{4}{3},3) + (1,-\frac{4}{3},\bar{3}) + (1,0,1) + (10,-\frac{2}{3},3) \\ \quad + (10,\frac{2}{3},\bar{3}) + (16,-1,1) + (16,+\frac{1}{3},3) + (\overline{16},+1,1) + (\overline{16},-\frac{1}{3},\bar{3}). \end{cases} \tag{42}$$

By identifying the terms that are singlet, triplets, or antitriplets under $SU(3)'_7$ one gets

$$78 \to \begin{cases} SU(6) \otimes SU(2)_L: \ (35,1) + (1,3) + (20,2), \\ SO(10) \otimes U(1)_{Y'}: \ (45,0) + (1,0) + (16,-1) + (\overline{16},+1), \end{cases}$$

$$27 \to \begin{cases} SU(6) \otimes SU(2): \ (15,1) + (\bar{6},2), \\ SO(10) \otimes U(1)_{Y'}: \ (1,+\frac{4}{3}) + (10,-\frac{2}{3}) + (16,+\frac{1}{3}). \end{cases}$$

To find the transformation properties under the $SU(3)^3$ maximal subalgebra of E_6 spanned by

$$SU(3)_7 \oplus A \oplus S \otimes (e_7 \oplus e'_7) \oplus A \otimes e_7 \otimes e'_7,$$

it is useful to compare the decomposition of the 133 representation of E_7 given in Table II and by Eq. (40):

TABLE IV. Decomposition of the exceptional algebras under their maximal classical subalgebras. Here $SU(4)_7$ is the subalgebra of $SO(7)$, defined in Sec. IV (v), (d) not acting on e_7. The $SU(2)_P$ is generated by $\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_7$.

E_8	$SO(16)$	$120 + 128$
E_7	$SO(12) \otimes SU(2)_L$	$(66,1) + (1,3) + (32,2)$
E_6	$SO(10) \otimes U(1)$	$(45,1) + (1,1) + (16,1) + (\overline{16},-1)$
	$Sp(8)$	$36 + 42$
F_4	$SO(9)$	$36 + 16$
	$Sp(6) \otimes SU(2)_L$	$(21,1) + (14',2) + (1,3)$
G_2	$SU(3)_7$	$8 + 3 + \bar{3}$
	$SU(2)_S \otimes SU(2)_L$	$(3,1) + (1,3) + (4,2)$
F_4	$SU(4)_7 \otimes SU(2)_P$	$(15,1) + (1,3) + (4,2)$ $+ (\bar{4},2) + (6,3)$
	$SU(3)_7 \otimes SU(3)_G$	$(8,1) + (1,8) + (3,6) + (\bar{3},\bar{6})$
E_6	$SU(6) \otimes SU(2)_L$	$(35,1) + (1,3) + (20,2)$
	$SU(3)_7 \otimes SU(3)_L \otimes SU(3)_R$	$(8,1,1) + (1,8,1) + (1,1,8)$ $+ (3,3,3) + (\bar{3},\bar{3},\bar{3})$
E_7	$SU(8) \supset SU(6)_{GR} \otimes SU(2)_L \otimes U(1)_{T'_{3S}}$	$63 + 70$
	$SU(6)'_{GR} \otimes SU(3)_7$	$(8,1) + (1,35) + (3,15) + (\bar{3},\overline{15})$
	$SU(4)_7 \otimes SU(4) \otimes SU(2)_Q$	$(15,1,1) + (1,15,1) + (1,1,3)$ $+ (4,\bar{4},2) + (\bar{4},4,2) + (6,6,1)$
E_8	$SU(9) \supset SU(6)_{GR} \otimes SU(3)'_7 \otimes U(1)_{T_{3L}}$	$80 + 84 + \overline{84}$
	$SU(5) \otimes SU(5)$	$(24,1) + (1,24) + (10,5)$ $(\overline{10},\bar{5}) + (\bar{5},10) + (5,\overline{10})$

$$
133 \rightarrow \begin{cases} E_6 \otimes U(1)_{T'_{3S}}\colon\ (78,0) + (27,-1) + (\overline{27},+1) + (1,0)\,, \\ SU(3)_7 \otimes SU(6)' \rightarrow SU(3)^3 \otimes U(1)_{T'_{3S}}\colon \\ (1,8,1,0) + (1,1,8,0) + (1,1,1,0) + (1,3,\bar{3},-1) + (1,\bar{3},3,+1) \\ \quad + (8,1,1,0) + (3,\bar{3},1,-1) + (3,1,\bar{3},+1) + (3,3,3,0) \\ \quad + (\bar{3},3,1,+1) + (\bar{3},1,3,-1) + (\bar{3},\bar{3},\bar{3},0)\,. \end{cases}
$$

By identifying the terms corresponding to the same value of T'_{3S} one gets the well-known result[7] [with respect to Ref. 7 we choose the different convention $3 \leftrightarrow \bar{3}$ for the fundamental representation of $SU(3)_7$]

$$
\begin{aligned} &78 \rightarrow (8,1,1) + (1,8,1) + (1,1,8) + (3,3,3) + (\bar{3},\bar{3},\bar{3})\,, \\ &27 \rightarrow (1,3,\bar{3}) + (3,\bar{3},1) + (\bar{3},1,3)\,. \end{aligned} \tag{43}
$$

The transformation properties under Sp(8) may be obtained from Eq. (38):

$$
78 \rightarrow 36 + 42, \quad 27 \rightarrow 27\,. \tag{44}
$$

The 42(27) representation of Sp(8) is contained in the fourth (fourth- and second-) order antisymmetric tensor. Let us consider now the maximal subalgebras of F_4.

To find the transformation properties of the 52 and the 26 representations under SO(9) one has to compare the decompositions of the 248 of E_8:

$$
248 \rightarrow \begin{cases} F_4 \otimes G'_2\colon\ (52,1) + (26,7) + (1,14)\,, \\ SO(16) \rightarrow SO(9) \otimes SO(7) \rightarrow SO(9) \otimes G'_2\colon \\ \quad (36,1) + (1,14) + (1,7) + (9,7) + (16,7) + (16,1)\,. \end{cases} \tag{45}
$$

By identifying the G'_2 singlets and seven-dimensional representations one gets

$$
52 \rightarrow 36 + 16, \quad 26 \rightarrow 16 + 9 + 1\,. \tag{46}
$$

Finally we want to find the transformation properties under the maximal subalgebras F_4, $Sp(6) \otimes SU(2)_L$, and $SU(3)_7 \otimes SU(3)_G$.

To this extent we compare the following different decompositions of the 133 of E_7:

TABLE V. Decomposition of the fundamental representations of the exceptional algebras under their maximal subalgebras.

	$SU(3)_7$	$3 + \bar{3} + 1$
G_2	$SU(2)_L \otimes SU(2)_S$	$(2,2) + (1,3)$
	$G_2 \otimes SU(2)_W$	$(1,5) + (7,3)$
	$SU(3)_7 \otimes SU(3)_{GM}$	$(1,8) + (3,\bar{3}) + (\bar{3},3)$
F_4	$SP(6) \otimes SU(2)_L$	$(14,1) + (6,2)$
	$SU(4)_7 \otimes SU(2)_W$	$(6,1) + (1,3) + (4,2) + (\bar{4},2) + (1,1)$
	$SO(9)$	$16 + 9 + 1$
	F_4	$26 + 1$
	$G_2 \otimes SU(3)'_7$	$(7,3) + (1,6)$
E_6	$SU(3)_7 \otimes SU(3)_L \otimes SU(3)_R$	$(3,\bar{3},1) + (1,3,\bar{3}) + (\bar{3},1,3)$
	$SU(6)_{GR} \otimes SU(2)_L$	$(15,1) + (\bar{6},2)$
	$SO(10) \otimes U(1)$	$(16,\frac{1}{2}) + (10,-1) + (1,2)$
	$Sp(8)$	27
	$E_6 \otimes U(1)$	$(27,-\frac{1}{2}) + (\overline{27},\frac{1}{2}) + (1,\frac{3}{2}) + (1,-\frac{3}{2})$
	$F_4 \otimes SU(2)'_S$	$(26,2) + (1,4)$
	$G_2 \otimes Sp(6)$	$(7,6) + (1,14)$
E_7	$SU(6)'_{GR} \otimes SU(3)_7$	$(20,1) + (6,3) + (\bar{6},\bar{3})$
	$SU(8)$	$28 + \overline{28}$
	$SU(4)_7 \otimes SU(4) \otimes SU(2)_P$	$(6,1,2) + (1,6,2) + (4,4,1) + (\bar{4},\bar{4},1)$
	$SO(12) \otimes SU(2)_L$	$(12,2) + (32',1)$
	$E_7 \otimes SU(2)'_L$	$(133,1) + (1,3) + (56,2)$
	$E_6 \otimes SU(3)'_7$	$(78,1) + (1,8) + (27,3) + (\overline{27},\bar{3})$
	$F_4 \otimes G_2$	$(26,7) + (52,1) + (1,14)$
E_8	$SU(9)$	$80 + 84 + \overline{84}$
	$SU(5) \otimes SU(5)$	$(24,1) + (1,24) + (10,5) + (\overline{10},\bar{5}) + (\bar{5},10) + (\overline{10},5)$
	$SO(16)$	$120 + 128$

TABLE VI. Vector spaces of the fundamental representations of the exceptional algebras.

$(56,2)$ of $E_7 \otimes SU(2)'_L$ $\dfrac{G'_2}{SU(2)'_L \otimes SU(2)'_S}$	SU(8)		$SO(12) \otimes SU(2)_L$		$SU(6)' \otimes SU(3)_7$
					$(20,1)$
$(S \oplus A \otimes e_7) \otimes e'_r$	$28 + \overline{28}$		$(32',1)$		
$A \otimes e_{3,6} \otimes e'_r$					$(\bar{6},3) + (6,\bar{3})$
$A \oplus e_s \otimes e'_r$			$(12,2)$		
$(27,3)$ of $E_6 \otimes SU(2)'_7$ D'_{e_+,e_7}			$SU(6) \otimes SU(2)_L$		$SU(3)_{e_7} \otimes SU(3)_L \otimes SU(3)_R$
					$(1,3,\bar{3})$
$(S \oplus A \otimes e_7) \otimes e'_+$			$(15,1)$		
$A \otimes e_{3,6} \otimes e'_+$					
					$(3,\bar{3},1) + (\bar{3},1,3)$
$A \otimes e_r \otimes e'_+$			$(\bar{6},2)$		
$(27,3)$ of $E_6 \otimes SU(3)'_7$				$SO(10) \otimes U(1)'_Y$	
$(1/\sqrt{3})D'_{+7} + (i\sqrt{2}/\sqrt{3})\lambda_8 e'_+$				$(1,+2)$	
$[\lambda_4 \oplus \lambda_6 \oplus (i\lambda_5 \oplus i\lambda_7) \otimes e_i] \otimes e'_i$				$(16,+\frac{1}{2})$	
$(\lambda_1 \oplus \lambda_3 + \oplus i\lambda_2 \otimes e_i) \otimes e'_+$					
$\oplus \left[\sqrt{\tfrac{2}{3}} D'_{+7} - (i/\sqrt{3})\lambda_8 e'_+ \right]$				$(10,-1)$	
$(26,7)$ of $F_4 \otimes G_2$		$Sp(6) \otimes SU(2)_L$			$SU(3)_G \otimes SU(3)_7$
$(S_s \oplus A \otimes e_7) \otimes e'_i$					$(8,1)$
		$(14,1)$			
$A \otimes e_{3,6} \otimes e'_i$					$(3,\bar{3})$
$A \otimes e_r \otimes e'_i$		$(6,2)$			$(\bar{3},3)$
$(26,7)$ of $F_4 \otimes G_2$				SO(9)	
$\lambda_8 \otimes e_i$				1	
$[\lambda_4 \oplus \lambda_6 \oplus (i\lambda_5 \oplus i\lambda_7) \otimes e_i] \otimes e'_i$				16	
$(\lambda_1 \oplus \lambda_3 \oplus i\lambda_2 \otimes e_i) \otimes e'_i$				9	

$$
133 \to \begin{cases} F_4 \otimes SU(2)'_S : (52,1) + (26,3) + (1,3)\,, \\ SO(12) \otimes SU(2)_L \to Sp(6) \otimes SU(2)'_S \otimes SU(2)_L : \\ \quad (21,1,1) + (14,3,1) + (1,3,1) + (14',1,2) + (6,3,2) + (1,1,3) \\ SU(3)_7 \otimes SU(6)' \to SU(3)_7 \otimes SU(3)_G \otimes SU(2)'_S : (8,1,1) + (1,8,1) + (1,1,3) + (1,8,3) + (3,6,1) \\ \quad + (3,\bar{3},3) + (\bar{3},\bar{6},1) + (\bar{3},3,3)\,. \end{cases} \tag{47}
$$

By identifying the $SU(2)'_S$ singlets and triplets one easily finds

$$
\begin{aligned} &52 \to \begin{cases} Sp(6) \otimes SU(2)_L :\ (21,1) + (14',2) + (1,3), \\ SU(3)_7 \otimes SU(3)_G :\ (8,1) + (1,8) + (3,6) + (\bar{3},\bar{6})\,, \end{cases} \\ &26 \to \begin{cases} Sp(6) \otimes SU(2)_L :\ (14,1) + (6,2), \\ SU(3)_7 \otimes SU(3)_G :\ (1,8) + (3,\bar{3}) + (\bar{3},3)\,. \end{cases} \end{aligned} \tag{48}
$$

The 14′ (14) is contained on the third- (second-) order antisymmetric tensor of Sp(6).

The inclusion of the maximal classical subalgebras of the exceptional algebras and the transformation properties of the adjoint and fundamental representations are described in Tables III–V. The explicit construction of the vector spaces corresponding to the fundamental representations of the exceptional algebras and their transformation properties under their maximal subalgebras are described in Table VI.

[1]M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 117 (1984).

[2]M. Günaydin and F. Gürsey, J. Math. Phys. **14**, 1651 (1973); F. Gürsey, "Johns Hopkins University Workshop on Current Problems in High Energy Particle Theory," Baltimore, MD, 1974; presented at the Kyoto Conference in Mathematical Problems in Theoretical Physics, Kyoto, Japan, 1975.

[3]R. D. Schafer, *An Introduction to Nonassociative Algebras* (Academic, New York, 1966), and the references therein; N. Jacobson, *Exceptional Lie Algebras* (Dekker, New York, 1971).

[4]R. C. King and A. K. A. Al-Qubanchi, J. Phys. A: Math. Gen. **14**, 13 (1981).

[5]M. Gell-Mann, Phys. Rev. **125**, 1067 (1962).

[6]F. Gürsey and L. A. Radicati, Phys. Rev. Lett. **13**, 173 (1964); B. Sakita, Phys. Rev. B **136**, 1756 (1964).

[7]F. Gürsey, P. Ramond, and P. Sikivie, Phys. Lett. B **60**, 177 (1986).

Boson operator realizations of su(2) and su(1,1) and unitarization

H. D. Doebner
Arnold Sommerfeld Institute for Mathematical Physics, Technical University Clausthal, D 3392 Clausthal, West Germany
Bruno Gruber
Physics Department, Southern Illinois University, Carbondale, Illinois 62901 and Arnold Sommerfeld Institute for Mathematical Physics, Technical University Clausthal, D 3392 Clausthal, West Germany
Miguel Lorente[a)]
Departamento de Métodos Matemáticos de la Física, Universidad Complutense de Madrid, 28040 Madrid, Spain

(Received 16 June 1987; accepted for publication 10 August 1988)

Boson operator realizations of su(2) and su(1,1) are obtained. Scalar products are introduced on "Fock spaces" (Verma modules) spanned by generators of the Heisenberg algebra H and by generators of su(2). These scalar products unitarize certain of the representations of H, or of su(1,1). It is shown that the Gel'fand–Dyson realization of su(1,1) implies a scalar product that unitarizes H, while the Primakoff–Holstein realizations imply a scalar product that unitarizes su(1,1). The relationship between the Gel'fand–Dyson boson operators $a^\dagger$ and the Primakoff–Holstein boson operators $b^\dagger$ is obtained making use of the two distinct scalar products. Generalized "vacuum states" are defined that are formed by polynomials in the creation–annihilation operator pairs $a^\dagger a$. A representation ρ of H and su(1,1) on the states $a^{\dagger m}$ and a^n is discussed. For this representation $a\mathbb{1} = a|0\rangle \neq 0$, but rather $(a^\dagger a)|0\rangle = 0$. The states of this representation space consist of boson states and boson-hole states. All the familiar results of H and su(1,1) representation theory are preserved within the representation ρ.

I. INTRODUCTION

A natural space carrying linear representaions of a Lie algebra L is its universal enveloping algebra $U(L)$, and the quotient spaces $U(L)/J$, where J is a (left sided) ideal generated through algebraic relations in $U(L)$. The canonical representation $\hat{\rho}$ of L—the so-called master representation—can be calculated directly by acting with basis elements of L on the Poincaré–Birkhoff–Witt basis in $U(L)$. On the quotient spaces $U(L)/J$ one obtains induced representations $\rho(L)$ from $\hat{\rho}$. Depending on the choice of J, different types of representations, e.g., reducible, irreducible, nondecomposable, can be constructed, analyzed, and related (see, e.g., Refs. 1–3. If one is modeling physics via $\hat{\rho}(L)$ then one can show that the particular choice of J carries physical information, for example, in the appearance of "vacuum states."

The algebras $U(L)$ and $U(L)/J$ are linear spaces. A natural definition of an inner product faces difficulties, but is necessary for physical applications. One would like to discuss Hermiticity properties and a closure of $U(L)/J$ is a candidate for a physical Hilbert space. For semisimple L and for certain $U(L)/J$, the Harish-Chandra sesquilinear form[4,5] induces a positive definite inner product and the representation space of $\rho(L)$ has the structure of a Verma module, i.e., it is a space generated by a cyclic element, namely the identity $\mathbb{1}$ of $U(L)$ (the vacuum state). For nonsemisimple L an analog construction would be useful, especially for the n-dimensional Heisenberg algebra H_n [basis $a_1,\dots,a_n$, $a^\dagger_1,\dots,a^\dagger_n$, E (see Sec. II, $H_1 = H$)]. In Sec. II we construct such a sesquilinear form on $U(H_1)$, which induces an inner product on $U(H_1)/K_1$, with K_1 a (left) ideal generated by $a\mathbb{1}$, $(E-\mu)\mathbb{1}$, $\mu\in\mathbb{R}$, fixed. Here $\mathbb{1}$ is the identity in $U(H_1)$. This is shown in Sec. III. In this section other representations will also be discussed which are induced from other ideals. Especially from K_2, generated by $(E-\mu)\mathbb{1}$, $\mu\in R$, which has a very peculiar kind of vacuum state and from K_3, generated by $(E-\mu)\mathbb{1}$, $a^\dagger a\mathbb{1}$, $\mu\in R$, both of which are not Fock-like representations.

The Heisenberg algebra has another physically important and mathematically convenient property. Because of the simple structure of the representation $\hat{\rho}(L)$ [see, e.g., (3.2) or (4.1)] on $U(L)$, or $\rho(L)$ on $U(L)/J$, one can realize the same representation on $U(H_n)$, n suitably chosen, with the generators $\hat{\rho}(L)$ given as finite-order polynomials in "bose operators" a_i, $a^\dagger_i$. The representations $\hat{\rho}$ and ρ yield *canonically* a boson realization for L on $U(H_n)$. [See Refs. 6 and 7. Similarly, quotient spaces like $U(H_n)/K$ carry boson realizations.] Mathematically, there is in general no preferred J, the background for all the realizations is the same. But the physical systems described via these realizations can be of very different structure.[8]

We explain the construction and application of canonical boson realizations in connection with our inner product on $U(H)/K$ in Secs. IV–VI. The master representation $\hat{\rho}(L)$ and its boson realization for su(2) and su(1,1) are given in Sec. IV for different ideals; also, the Gel'fand–Dyson representation is obtained. It was shown in Ref. 5 that there is an inner product on U(su(2))/J induced by Harish-Chandra's sesquilinear form. Hence the inner product can be transferred to $U(H)/K$ via the boson realization and we have two inner products on $U(H)/K$ that differ by the norm scale. We

[a)] Present address: Departamento de Física, Universidad de Oviedo, 33007 Oviedo, Spain.

 0022-2488/89/030594-07$02.50

analyze the Hermiticity properties of the representations, depending on the inner products. For $U(H)/K_3$ we get, e.g., an indecomposable representation of H and two irreducible Hermitian representations of su(1,1) that are living on polynomials of a and $a^\dagger$, respectively, and which have the Primakoff–Holstein form. For a certain parameter value ($l=\frac{1}{2}$) these representations are exceptionally simple and they possess formal extensions to larger representation spaces (Sec. IV).

Our presentation should indicate how to exploit the algebraical, and also the geometrical, structure of $U(L)$ and its quotient spaces for a discussion of certain aspects of Lie algebra representations, especially boson realizations of Lie algebras as symmetries and dynamical elements of physical systems.

II. DEFINITIONS

The Heisenberg algebra H, a nilpolent algebra, is a linear space over $\mathbb{C}$, spanned by three elements a, b, c. Following convential notation we replace the symbols a, b, c by a, $a^\dagger$, E, but we have to keep in mind that $a^\dagger$ is not necessarily the adjoint of a, and E is not necessarily the identity operator. We have

$$H:\{a,a^\dagger,E\}, \quad \mathbb{C},$$

with the Lie products

$$[a,a^\dagger]=E, \quad [a,E]=[a^\dagger,E]=0. \tag{2.1}$$

A basis for the enveloping algebra $\Omega\equiv U(H)$ is given by

$$U(H):\ \{B(m,n,r)=a^{\dagger m}a^nE^r;\ m,n,r\in\mathbb{N}\}, \tag{2.2}$$

with $B(0,0,0)=\mathbb{1}$ denoting the identity of $U(H)$. As usual the product for the basis elements $B(m,n,r)$ is the ordered tensor product. The general element of $U(H)$ can be represented as

$$B=\sum_{m,n,r} C_{mnr}B(m,n,r), \quad C_{mnr}\in\mathbb{C}, \tag{2.3}$$

with only a finite number of the $C_{mnr}\neq 0$. The center $Z(H)$ of $U(H)$ has a basis

$$Z(H):\ B(0,0,r), \quad r\in\mathbb{N}. \tag{2.4}$$

Within $U(H)$ the following relations hold:

$$\begin{aligned} aa^{\dagger n}&=a^{\dagger n}a+na^{\dagger(n-1)}E,\\ a^\dagger a^n&=a^na^\dagger-na^{(n-1)}E. \end{aligned} \tag{2.5}$$

We define on $U(H)$ a sesquilinear form in close analogy to the sesquilinear form on universal enveloping algebras for simple Lie algebras.[5] Let σ denote the conjugation of H,

$$\begin{aligned} \sigma(ca)&\doteq-\bar{c}a^\dagger, \quad \sigma(ca^\dagger)=-\bar{c}a,\\ \sigma(cE)&=-\bar{c}E, \quad c\in\mathbb{C}, \end{aligned} \tag{2.6}$$

where the overbar denotes complex conjugation. This conjugation is extended to an antilinear antiautomorphism η of $U(H)$,

$$\begin{aligned} \eta(\mathbb{1})&=\mathbb{1}, \quad \eta(BB')=\eta(B')\eta(B),\\ \eta(B)&=-\sigma(B), \quad B,B'\in U(H). \end{aligned} \tag{2.7}$$

We define a linear map (projection) π from $U(H)$ to $Z(H)$,

$$\begin{aligned} \pi(B(m,n,r))&=B(0,0,r), \quad \text{if } m=n=0,\\ &=0, \quad \text{if } m\neq 0 \text{ and/or } n\neq 0, \end{aligned} \tag{2.8}$$

and a map Λ from $Z(H)$ to $\mathbb{C}$,

$$\Lambda(B(0,0,r))=\mu^r, \quad r\in\mathbb{N}, \tag{2.9}$$

with μ some arbitrarily chosen, but fixed, complex number. We denote the composition of maps π, Λ by

$$\xi_\Lambda=\Lambda\circ\pi. \tag{2.10}$$

We obtain a sesquilinear form $S(B,B')$ on $U(H)$ as

$$S(B,B')=\xi_\Lambda(\eta(B)B'),$$

which is generally degenerate and not positive definite. It follows, using (2.5),

$$\begin{aligned} &S(B(m,n,r),B(s,t,u))\\ &=\Lambda\circ\pi\bigg(\sum_{\alpha=0}^{m}\frac{m!}{(m-\alpha)!\alpha!}\frac{s!}{(s-\alpha)!}\\ &\qquad\times B(n+s-\alpha,t+m-\alpha,r+u+\alpha)\bigg)\\ &=\Lambda(m!B(0,0,r+u+m))\delta_{s,m}\delta_{n,0}\delta_{t,0}\\ &=m!\mu^{r+u+m}\delta_{s,m}\delta_{n,0}\delta_{t,0}, \quad \mu\in\mathbb{C}, \quad \mu\neq 0, \end{aligned} \tag{2.11}$$

since the projection π requires for a nonzero result $\alpha=m$, $t=0$, and $n+s-m=0$. But $s\geqslant m$ and thus $n=0$, $s=m$.

III. INDUCED SCALAR PRODUCTS

The sesquilinear form (2.1) induces in some physically relevant cases a scalar product.

(1) Consider the ideal K_1 generated by[6]

$$K_1:\ a\mathbb{1},\ (E-\mu)\mathbb{1},\ \mu\in\mathbb{C},\ \mu\neq 0;$$

we form the quotient space $U(H)/K_1$ with basis

$$U(H)/K_1:\ B(m)=a^{\dagger m}, \quad m\in\mathbb{N}, \quad B(0)=\mathbb{1}.$$

The space $U(H)/K_1$ has the following properties: E acts as a multiple of the identity operator $\mathbb{1}$. Moreover, there exists a vacuum state, or "lowest state" $\mathbb{1}$, since $a\mathbb{1}=0$ (i.e., a acting upon the *state* $\mathbb{1}$ maps $\mathbb{1}$ to zero). Such a space, generated by a cyclic element, namely the element $\mathbb{1}$, is called a Verma module. Here it corresponds to the familiar Fock space. On $U(H)/K_1$ the sesquilinear form yields

$$S(a^{\dagger m},a^{\dagger n})=m!\mu^m\delta_{m,n}. \tag{3.1}$$

For $\mu\in\mathbb{R}^\dagger$ the sesquilinear form is nondegenerate and positive definite. Thus it defines a scalar product. For $\mu=1$ the familiar physical case is obtained with the orthonormalized basis states $a^{\dagger m}/\sqrt{m!}$. This is seen as follows.

On the space $U(H)$ the algebra H has a canonical representation, called master representation $\hat{\rho}$, given as[6]

$$\begin{aligned} \hat{\rho}(a^\dagger)B(m,n,r)&=B(m+1,n,r),\\ \hat{\rho}(a)B(m,n,r)&=B(m,n+1,r)\\ &\quad+mB(m-1,n,r+1),\\ \hat{\rho}(E)B(m,n,r)&=B(m,n,r+1). \end{aligned} \tag{3.2}$$

On the quotient space $U(H)/K_1$ the representation $\hat{\rho}$ induces a corresponding representation ρ (here and in the following we do not use labels to characterize the representations ρ in order to simplify the notation),

$$\begin{aligned} &\rho(a^\dagger)B(m) = B(m+1), \\ &\rho(a)B(m) = m\mu B(m-1), \\ &\rho(E)B(m) = \mu B(m). \end{aligned} \tag{3.3}$$

The $B(m)$ form, in $\mathrm{U}(H)/K_1$, an orthogonal basis with respect to the scalar product (3.1). They can be normalized to 1 and are denoted in this case as ket vectors (used as physical states)

$$|m\rangle = \frac{B(m)}{\sqrt{m!\mu}} = \frac{a^{\dagger m}}{\sqrt{m!\mu}}. \tag{3.4}$$

Then

$$\begin{aligned} &\rho(a^\dagger)|m\rangle = \sqrt{\mu(m+1)}\,|m+1\rangle, \\ &\rho(a)|m\rangle = \sqrt{\mu m}\,|m-1\rangle, \\ &\rho(E)|m\rangle = \mu|m\rangle. \end{aligned} \tag{3.5}$$

For $\mu = 1$ this is the familiar harmonic oscillator representation. The space $\mathrm{U}(H)/K_1$, with the sesquilinear form as a scalar product, can be closed to a Hilbert space. Its elements are weak limits,

$$Y = \sum_{m=0}^{\infty} C_m|m\rangle, \quad C_m \in \mathbb{C}. \tag{3.6}$$

(2) A further representation of H is induced by the master representation $\hat{\rho}$ on the quotient space $\mathrm{U}(H)/K_2$, where K_2 denotes the ideal generated by $(E-\mu)\mathbb{1}$, with basis

$$\begin{aligned} &\mathrm{U}(H)/K_2\colon \{B(m,n);\ m,n\in\mathbb{N},\ B(m,n) \equiv B(m,n,0)\}, \\ &\rho(a^\dagger)B(m,n) = B(m+1,n), \\ &\rho(a)B(m,n) = B(m,n+1) + \mu m B(m-1,n), \\ &\rho(E)B(m,n) = \mu B(m,n). \end{aligned} \tag{3.7}$$

The representation (3.7) has a remarkable property. In $\mathrm{U}(H)/K_2$ any $B(m,n)$ can be factored. With

$$\begin{aligned} &\Psi_n(m) = \prod_{s=1}^{k} (a^\dagger a - (n-s)\mathbb{1}), \\ &\qquad k = n \text{ for } m \geqslant n \geqslant 1, \quad k = m \text{ for } n \geqslant m \geqslant 1, \\ &\qquad \psi_0(m) = \psi_n(0) = \mathbb{1}, \end{aligned}$$

the $B(m,n)$ splits (uniquely) in

$$B(m,n) = \begin{cases} (a^\dagger)^{m-n}\Psi_n(m) = B(m-n,0)\psi_n, & m \geqslant n \geqslant 1, \\ a^{n-m}\Psi_n(m) = B(0,n-m)\psi_n(m), & n \geqslant m \geqslant 1, \end{cases}$$

and one obtains after cancellation of $\psi_n(m)$

$$\begin{aligned} &\rho(a^\dagger)B(n-m,0) = B(n-m+1,0), \quad m > n, \\ &\rho(a^\dagger)B(0,n-m) \\ &\quad = B(0,n-m-1)(a^\dagger a - (n-m-1)\mathbb{1}), \\ &\qquad\qquad\qquad\qquad n > m+1, \\ &\rho(a)B(m-n,0) = B(m-n-1,0)(a^\dagger a + (m-n)\mathbb{1}), \\ &\qquad\qquad\qquad\qquad m > n+1, \\ &\rho(a)B(0,n-m) = B(0,n-m+1), \quad n > m, \\ &\rho(E)B(m-n,0) = B(m-n,0), \quad m > n, \\ &\rho(E)B(0,n-m) = B(0,n-m), \quad n > m. \end{aligned} \tag{3.8}$$

The $\psi_n(m)$ can be interpreted as vacuum states containing boson creation–annihilation pairs $(a^\dagger a)$ in the form of a polynomial of order k. Since the order of the polynomials $\psi_n(m)$ under the action of ρ remains either the same, or increases by $(a^\dagger a)$, we obtain a representation by collecting those terms for which the order of the polynomials remains. Thus

$$\begin{aligned} &\rho(a^\dagger)B(m,0) = B(m+1,0), \quad m \geqslant 0, \\ &\rho(a^\dagger)B(0,n) = -(n-1)B(0,n-1), \quad n \geqslant 1, \\ &\rho(a)B(m,0) = mB(m-1,0), \quad m \geqslant 1, \\ &\rho(a)B(0,n) = B(0,n+1), \quad n \geqslant 0, \\ &\rho(E)B(m,0) = B(m,0), \\ &\rho(E)B(0,n) = B(0,n), \quad m,n \geqslant 0, \end{aligned} \tag{3.9}$$

with the representation space V spanned by

$$V\colon \{B(m,0), B(0,n);\ m,n\in\mathbb{N}\}. \tag{3.10}$$

This result is identical to the representation (3.22) of Ref. 6 for the values $\lambda = 0$, $\mu = 1$. Obviously it is induced by (3.7) on the quotient space $\mathrm{U}(H)/K_3$ with respect to the ideal K_3, generated by $(E-\mu)\mathbb{1}$, $a^\dagger a\mathbb{1}$. Note that $\rho(a)\mathbb{1} = a \neq 0$. But since $\rho(a^\dagger)a = 0$, the elements $B(0,n)$ span an invariant subspace V'. The quotient space of V with respect to this invariant subspace has the basis elements $B(m,0)$, and on it holds $\rho(a)\mathbb{1} = 0$.

(3) Equation (3.9) can be brought into a more suitable form by introducing the basis

$$\begin{aligned} V\colon\ &\{|m\rangle \equiv (a^\dagger)^m/\sqrt{m!}, \quad m\in\mathbb{N}; \\ &|-n\rangle \equiv a^n/\sqrt{(n-1)!}, \quad n\in\mathbb{N}^+\}, \end{aligned} \tag{3.11}$$

where we have introduced the kets $|-n\rangle$. One obtains

$$\begin{aligned} &\rho(a^\dagger)|m\rangle = \sqrt{m+1}\,|m+1\rangle, \quad m \geqslant 0, \\ &\rho(a)|m\rangle = \sqrt{m}\,|m-1\rangle, \quad m \geqslant 1, \quad \rho(a)|0\rangle = |-1\rangle, \\ &\rho(a^\dagger)|-n\rangle = -\sqrt{n-1}\,|-n+1\rangle, \quad n \geqslant 1, \\ &\rho(a)|-n\rangle = \sqrt{n}\,|-n-1\rangle, \quad n \geqslant 1, \\ &\rho(E)|m\rangle = |m\rangle, \\ &\rho(E)|-n\rangle = |-n\rangle. \end{aligned} \tag{3.12}$$

The subspace spanned by the $|-n\rangle$, $n\in\mathbb{N}^+$, is invariant. The quotient space of V with respect to this invariant subspace V' carries the familiar harmonic oscillator representation (3.5) (for $\mu = 1$). The entire space V carries an indecomposible representation. (See Fig. 1.)

IV. BOSON REALIZATIONS OF su(2)

In this section we will derive boson operator realizations of the master representation $\hat{\rho}$ of (complex) su(2), and rep-

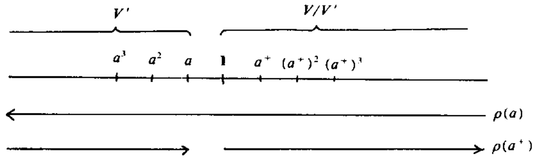


FIG. 1. The space V and the action of the operators $(a^\dagger)$ and (a) in the representations (3.9) and (3.12).

resentations ρ induced by $\hat{\rho}$. The su(2) representations discussed are among those which were derived in Ref. 1.

(1) The master representation of (complex) su(2) on the space of its universal enveloping algebra U(su(2)), spanned by

$$\{X(m,n,r) = l_+^m\, l_-^n\, l_3^r;\ m,n,r\in\mathbb{N}\},\quad X(0,0,0) = \mathbb{1},$$

is obtained as[1]

$$\begin{aligned}
\rho(l_3)X(n,m,r) &= X(n,m,r+1) + (n-m)X(n,m,r),\\
\rho(l_+)X(n,m,r) &= X(n+1,m,r),\\
\rho(l_-)X(n,m,r) &= X(n,m+1,r) - nX(n-1,m,r+1)\\
&\quad + n(m - \tfrac{1}{2}(n-1))X(n-1,m,r),\\
\rho(C)X(n,m,r) &= 2X(n+1,m+1,r) + X(n,m,r+2)\\
&\quad - (2m+1)X(n,m,r+1)\\
&\quad + m(m+1)X(n,m,r).
\end{aligned} \tag{4.1}$$

Here $C = 2l_+l_- + l_3(l_3-1)$ is the Casimir operator. The basis states of $\hat{\rho}$ are given in terms of three independent parameters m,n,r. This implies that the representation can be reexpressed in terms of three independent Heisenberg operators $a^\dagger_k$, a_k that satisfy the Lie products

$$[a_k,a^\dagger_s] = \delta_{ks},\quad s,k = 1,2,3.$$

The boson operator realizations will then act upon the boson basis $B^{(3)}$,

$$\begin{aligned}
B^{(3)}(m,n,r) &= B_1(m,0,0)B_2(n,0,0)B_3(r,0,0)\\
&= (a_1^\dagger)^m(a_2^\dagger)^n(a_3^\dagger)^r.
\end{aligned}$$

We obtain, following a procedure given in Ref. 6, Sec. V.3 (see also Ref. 7),

$$\begin{aligned}
\rho(l_3) &= a_1^\dagger a_1 - a_2^\dagger a_2 - a_3,\\
\rho(l_+) &= a_2,\\
\rho(l_-) &= a_2^\dagger(a_1^\dagger a_1 - \tfrac{1}{2}a_2^\dagger a_2 - a_3) + a_1,\\
\rho(C) &= 2a_1a_2 - (2a_1^\dagger a_1 - a_3 + 1)a_3 + a_1^\dagger a_1(a_1^\dagger a_1 + 1).
\end{aligned} \tag{4.2}$$

(2) The master representation $\hat{\rho}$ induces and subduces other representations on quotient spaces and invariant subspaces of U(su(2)). The elements $l_-^m\, l_+^r$, $m\in\mathbb{N}^+$, generate an ideal T_1 in U(su(2)). On the quotient space U(su(2))/T_1 with basis $\{X(n,r) = l_+^n\, l_3^r,\ n,r\in\mathbb{N}\}$ we obtain

$$\begin{aligned}
\rho(l_3)X(n,r) &= X(n,r+1) + nX(n,r),\\
\rho(l_+)X(n,r) &= X(n+1,r),\\
\rho(l_-)X(n,r) &= -nX(n-1,r+1)\\
&\quad - \tfrac{1}{2}(n-1)nX(n-1,r),\\
\rho(C)X(n,r) &= X(n,r+2) - X(n,r+1).
\end{aligned} \tag{4.3}$$

In the boson realization this representation becomes

$$\begin{aligned}
\rho(l_3) &= -a_2^\dagger a_2 - a_3,\\
\rho(l_+) &= a_2,\\
\rho(l_-) &= -\tfrac{1}{2}a_2^{\dagger 2}a_2 - a_2^\dagger a_3,\\
\rho(C) &= a_3(a_3-1).
\end{aligned} \tag{4.4}$$

(3) The master representation $\hat{\rho}$ induces on U(su(2))/T_2 with the ideal T_2 generated by $(l_3 - \Lambda)\mathbb{1}$, $\Lambda\in\mathbb{C}$, a further representation on the space spanned by

$$\begin{aligned}
&\mathrm{U(su(2))}/T_2\colon\ \{X(m,n) \equiv l_+^m\, l_-^n;\ m,n\in\mathbb{N}\},\\
\rho(l_3)X(n,m) &= (\Lambda + n - m)X(n,m),\\
\rho(l_+)X(n,m) &= X(n+1,m),\\
\rho(l_-)X(n,m) &= X(n,m+1)\\
&\quad - n(\Lambda - m + \tfrac{1}{2}(n-1))X(n-1,m),\\
\rho(C)X(n,m) &= 2X(n+1,m+1)\\
&\quad + (\Lambda - m)(\Lambda - m - 1)CX(n,m),
\end{aligned} \tag{4.5}$$

$\Lambda\in\mathbb{C}$, with boson realization;

$$\begin{aligned}
\rho(l_3) &= a_1^\dagger a_1 - a_2^\dagger a_2 - \Lambda,\\
\rho(l_+) &= a_2,\\
\rho(l_-) &= a_2^\dagger(a_1^\dagger a_1 - \tfrac{1}{2}a_2^\dagger a_2 - \Lambda) + a_1,\\
\rho(C) &= 2a_1a_2 - 2\Lambda a_1^\dagger a_1\\
&\quad + a_1^\dagger a_1(a^\dagger a_1 + 1) + \Lambda(\Lambda - 1)\mathbb{1}.
\end{aligned} \tag{4.6}$$

(4) A very unusual representation is induced by $\hat{\rho}$ on U(su(2))/T_3, with the ideal T_3 generated by $(l_- - \lambda)\mathbb{1}$, $\lambda\in\mathbb{C}$ and the basis spanned by $\{X(n,r) = l_+^n\, l_3^r,\ n,r\in\mathbb{N}\}$,

$$\begin{aligned}
\rho(l_3)X(n,r) &= X(n,r+1) + nX(n,r),\\
\rho(l_+)X(n,r) &= X(n+1,r),\\
\rho(l_-)X(n,r) &= \lambda\sum_{k=0}^{r}\frac{r!}{(r-k)!k!}X(n,k)\\
&\quad - nX(n-1,r+1)\\
&\quad - \tfrac{1}{2}(n-1)nX(n-1,r),\\
\rho(C)X(n,r) &= 4\lambda\sum_{k=0}^{r}\frac{r!}{(r-k)!k!}X(n+1,k)\\
&\quad + X(n,r+2) - X(n,r+1),
\end{aligned} \tag{4.7}$$

with boson realization

$$\begin{aligned}
\rho(l_3) &= -a_2^\dagger a_2 - a_3,\\
\rho(l_+) &= a_2,\\
\rho(l_-) &= \lambda e^{a_3^\dagger} - a_2^\dagger(a_3 + \tfrac{1}{2}a_2^\dagger a_2),\\
\rho(C) &= 4\lambda e^{a_3^\dagger}a_2 + a_3(a_3-1).
\end{aligned} \tag{4.8}$$

The boson realizations can be expressed in terms of differential operators by the substitution

$$a_1^\dagger\to x,\quad a_1\to\partial_x,\quad a_2^\dagger\to y,\quad a_2\to\partial_y,\quad a_3^\dagger\to z,\quad a_3a_3\to\partial_z.$$

For example, for (4.8) one obtains

$$\begin{aligned}
\rho(l_3) &= -y\,\partial_y - \partial_z,\\
\rho(l_+) &= \partial_y,\\
\rho(l_-) &= \lambda e^z - y\,\partial_z - \tfrac{1}{2}y^2\,\partial_y,\\
\rho(C) &= 4\lambda e^z y + \partial_z^2 - \partial_z,
\end{aligned} \tag{4.9}$$

the last equation being equivalent to a partial differential equation invariant under group su(2).

(5) Familiar realizations of su(2) and (su(1,1)) are readily identified as special cases of the realizations given

above. The Gel'fand–Dyson su(2) representation[9,10] is induced by the representation (4.6) modulo the ideal generated by $a_1\mathbb{1}$. One obtains (dropping the suffix 2) for su(2)

$$\begin{aligned} &\rho(l_3) = -\Lambda - a^\dagger a, \quad \rho(l_+) = a,\\ &\rho(l_-) = -\Lambda a^\dagger - \tfrac{1}{2}a^{\dagger 2}a, \\ &\rho(C) = \Lambda(\Lambda-1)\mathbb{1}, \end{aligned} \tag{4.10a}$$

and for su(1,1)

$$\begin{aligned} &\rho'(l_3) = \Lambda + a^\dagger a,\\ &\rho'(l_+) = a^\dagger, \\ &\rho'(l_-) = \Lambda a + \tfrac{1}{2}a^\dagger a^2, \end{aligned} \tag{4.10b}$$

and

$$\begin{aligned} &J_z = \rho(l_3) = -\Lambda - a^\dagger a,\\ &J_x = (1/\sqrt{2})(-\Lambda a^\dagger - (\tfrac{1}{2}a^{\dagger 2} - 1)a), \\ &J_y = (i/\sqrt{2})(-\Lambda a^\dagger - (\tfrac{1}{2}a^{\dagger 2} + 1)a), \\ &[J_x,J_y] = iJ_z \quad \text{(cyclic)}. \end{aligned} \tag{4.11}$$

The familiar form of the Gel'fand–Dyson representation ($\Lambda = -j$) is obtained by defining primed operators

$$J'_x = \sqrt{2}J_x - \tfrac{1}{2}a, \quad J'_y = \sqrt{2}J_y + \tfrac{1}{2}a, \quad J'_z = J_z,$$

followed by an application of the antilinear antiautomorphism η, Eqs. (2.6) and (2.7).

(6) Equation (4.6) induces also a representation of the type of the generalized Gel'fand–Dyson representation.[9] Setting $\Lambda = 0$,

$$\begin{aligned} &\rho(l_3) = a_1^\dagger a_1 - a_2^\dagger a_2,\\ &\rho(l_+) = a_2, \\ &\rho(l_-) = a_2^\dagger a_1^\dagger a_1 - \tfrac{1}{2}a_2^{\dagger 2}a_2 + a_1, \\ &\rho(C) = 2a_1a_2 + a_1^\dagger a_1(a_1^\dagger a_1 + 1). \end{aligned} \tag{4.12}$$

Setting $a_1 = b_1$, $a_2 = a$,

$$\begin{aligned} J_z &= \rho(l_3) = b^\dagger b - a^\dagger a,\\ J_x &= (1/\sqrt{2})(\rho(l_+) + \rho(l_-)) \\ &= (1/\sqrt{2})(a^\dagger b^\dagger b - \tfrac{1}{2}a^{\dagger 2}a + a + b), \\ J_y &= (-i/\sqrt{2})(\rho(l_+) - \rho(l_-)) \\ &= (i/\sqrt{2})(a^\dagger b^\dagger b - \tfrac{1}{2}a^{\dagger 2}a - a + b). \end{aligned} \tag{4.13}$$

Then,

$$\begin{aligned} &J'_x = \sqrt{2}J_x - \tfrac{1}{2}(a^\dagger b^\dagger b + a) - b,\\ &J'_y = \sqrt{2}J_y - (i/2)(a^\dagger b^\dagger b - a) - ib, \\ &J'_z = J_z - \tfrac{1}{2}b^\dagger b, \end{aligned}$$

are identical in form to Ref. 9, up to the map η.

V. SCALAR PRODUCTS AND EQUIVALENCE

In Sec. II we introduced a sesquilinear form on U(H) that induced a scalar product on the Fock space, Sec. III, Eq. (3.1). In this section we will make use of a sesquilinear form on U(su(2)) that becomes a scalar product on certain invariant subspaces and quotient spaces of U(su(2)). In Sec. IV we have given realizations of su(2) and su(1,1) in terms of boson operators. Thus two *distinct scalar products* are available to study Hermiticity properties of both the algebras H and su(2) (su(1,1)).

In the harmonic oscillator (orthonormalized) basis $|m\rangle$, Eq. (3.4), the Heisenberg algebra H was obtained in the form (3.5) while the su(1,1) realization (4.10b) takes on the form $[\rho'(l_+) \to -\rho'(l_+)$ for su(1,1)$]$,

$$\begin{aligned} &\rho(l_3)|m\rangle = (l+m)|m\rangle,\\ &\rho(l_+)|m\rangle = \sqrt{m+1}|m+1\rangle, \\ &\rho(l_-)|m\rangle = \sqrt{m}(l + \tfrac{1}{2}(m-1))|m-1\rangle, \\ &\rho(C)|m\rangle = (l(l-1) + m(m-2) + 4ml)|m\rangle, \quad m\in\mathbb{N} \end{aligned} \tag{5.1}$$

with $\Lambda = l$, and deleting primes.

Equation (3.5) shows that in the basis $|m\rangle$ the operator $a^\dagger$ is really the adjoint of a and it holds

$$S(|m'\rangle, a^\dagger|m\rangle) = S(a|m'\rangle, |m\rangle). \tag{5.2}$$

This is however not the case for su(1,1) as Eq. (5.1) shows. We thus ask the question whether there exists some scalar product S' such that

$$S'(\rho(l_+)a^{\dagger k}, a^{\dagger n}) = S'(a^{\dagger k}, \rho(l_-)a^{\dagger n}), \tag{5.3}$$

with respect to the (unnormalized) basis $\{a^{\dagger m}\}$. The condition (5.3) implies

$$\begin{aligned} &S'(a^{\dagger k}, a^{\dagger n}) = \delta_{kn} n! \prod_{s=1}^{n} \left(l + \frac{1}{2}(s-1)\right) S'(\mathbb{1},\mathbb{1}),\\ &S'(\mathbb{1},\mathbb{1}) = 1, \end{aligned} \tag{5.4}$$

where we have set $\Lambda = l \neq -\tfrac{1}{2}(s-1)$, $s\in\mathbb{N}^+$.

But this is precisely the scalar product S_1 (Ref. 5) induced by the sesquilinear form on the space spanned by the (noncompact) generators $\hat{l}_-$,

$$\{\hat{l}^{\,n}_-,\ n\in\mathbb{N}\}, \tag{5.5}$$

of su(1,1) [see Ref. 1, Eq. (3.14), $l \to -l$]. If we introduce the new states $|n)$, orthonormalized with respect to the scalar product S_1,

$$|n) = \left\{ n! \prod_{s=1}^{n} \left(l + \frac{1}{2}(s-1)\right)\right\}^{-1/2} a^{\dagger n}, \quad n\in\mathbb{N}^+, \tag{5.6}$$

then the Heisenberg algebra H takes on the form

$$\begin{aligned} &\rho(a^\dagger)|n) = \sqrt{n+1}\sqrt{l + \tfrac{1}{2}n}|n+1),\\ &\rho(a)|n) = [\sqrt{n}/\sqrt{l + \tfrac{1}{2}(n-1)}]|n-1), \\ &\rho(E)|n) = |n), \end{aligned} \tag{5.7}$$

$n\in\mathbb{N}$. The second of Eq. (5.7) becomes singular for values $l = -\tfrac{1}{2}(n-1)$.

The algebra su(1,1), Eq. (4.10b) for $\Lambda = l > 0$, takes on the unitary form,

$$\begin{aligned} &\rho(l_3)|n) = (l+n)|n),\\ &\rho(l_+)|n) = \sqrt{n+1}\sqrt{l + \tfrac{1}{2}n}|n+1), \\ &\rho(l_-)|n) = \sqrt{n}\sqrt{l + \tfrac{1}{2}(n-1)}\,|n-1), \\ &\rho(C)|n) = (l(l-1 + n(n-2) + 4nl)|n). \end{aligned} \tag{5.8}$$

Thus while for the Heisenberg algebra it holds that

$\rho(a)^\dagger = \rho(a^\dagger)$ in the basis $|m\rangle$ and with respect to the scalar product S, it holds for the noncompact real algebra su(1,1) that $\rho(l_+)^\dagger = \rho(l_-)$ in the basis $|n)$ and with respect to the scalar product S_1.

The relationship between the two bases is

$$
\begin{aligned}
|n) &= \left\{ \prod_{k=1}^{n} \left(l + \frac{1}{2}(k-1) \right) \right\}^{-1/2} |n\rangle, \\
&= \left\{ \prod_{k=1}^{n} \left(l + \frac{1}{2}(k-1) \right) \right\}^{-1/2} (n!)^{-1/2} a^{\dagger n} \mathbf{1}, \\
|0) &= |0\rangle = \mathbf{1}.
\end{aligned} \tag{5.9}
$$

On the basis $|n)$ another representation of the Heisenberg algebra can be defined. The generators are $b^\dagger$, b, E with $(b^{\dagger n}/\sqrt{n!})|0) = |n), |0) = 1$. The relation between the $b^\dagger$ and $a^\dagger$ is given by

$$
\begin{aligned}
& b^\dagger = a^\dagger(\rho'(l_3) - \tfrac{1}{2}N)^{-1/2}, \quad b = (\rho'(l_3) - \tfrac{1}{2}N)^{1/2} a, \\
& N = a^\dagger a = b^\dagger b, \\
& b^{\dagger n}\mathbf{1} = \left\{ 1 \prod_{k=1}^{n} \left(l + \frac{1}{2}(k-1) \right) \right\}^{-1/2} a^{\dagger n}\mathbf{1}, \quad l>0, \quad n \geqslant 1.
\end{aligned} \tag{5.10}
$$

With $b^\dagger b\,|n) = n|n)$, $b^\dagger|n) = \sqrt{n+1}\,|n+1)$, $b\,|n) = \sqrt{n}\,|n-1)$, we obtain the Primakoff–Holstein realization of su(1,1),

$$
\begin{aligned}
\rho(l_3)|n) &= (l + b^\dagger b)|n), \\
\rho(l_+)|n) &= b^\dagger \sqrt{l + \tfrac{1}{2} b^\dagger b}\,|n), \\
\rho(l_-)|n) &= \sqrt{l + \tfrac{1}{2} b^\dagger b}\, b\,|n), \\
\rho(C)|n) &= ((l + b^\dagger b)(l + b^\dagger b - 1) \\
&\quad + 2(l + \tfrac{1}{2} b^\dagger b) b^\dagger b)|n).
\end{aligned} \tag{5.11}
$$

The relationship between the two realizations (4.10b) and (5.11) was discussed by Moshinsky using a different approach.[10,11]

The two bases $\{|m\rangle\}$ and $\{|m)\}$ are bases for the same (algebraic) space. They are related to each other by scaling.

The two realizations of su(1,1) as well as the two realizations of H, are equivalent, however they are different in the scaling of the scalar product.

VI. AN EXAMPLE

We discuss a special representation for H and su(1,1) for $l = \frac{1}{2}$. Equation (5.7) becomes (replacing n by m),

$$
\begin{aligned}
\rho(a^\dagger)|m) &= (1/\sqrt{2})(m+1)|m+1), \\
\rho(a)|m) &= \sqrt{2}|m-1), \\
\rho(E)|m) &= |m),
\end{aligned} \tag{6.1}
$$

and $\rho(a)|0) \neq 0$. In fact, we can formally extend the values of m to $m \in \mathbb{Z}$ to form a larger representation. With the new basis elements

$$
\begin{aligned}
|m) &= 2^{m/2}(a^{\dagger m}/m!), \quad m \in \mathbb{N}, \\
|-n) &= 2^{(n-1)/2}[a^n/(n-1)!], \quad n \in \mathbb{N}^+,
\end{aligned} \tag{6.2}
$$

we obtain the relations

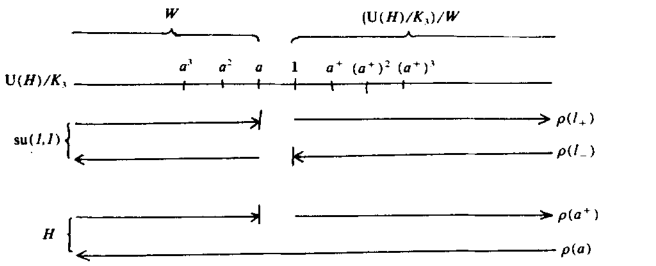


FIG. 2. The two su(1,1) irreducible representations (6.4) and (6.5) and their relationship to the indecomposable representation of H (6.3).

$$
\begin{aligned}
\rho(a^\dagger)|m) &= (m+1)\sqrt{2}|m+1), \quad m \in \mathbb{N}, \\
\rho(a^\dagger)|-n) &= -\sqrt{2}|-n+1), \quad n \geqslant 2, \\
\rho(a^\dagger)|-1) &= a^\dagger a = 0, \\
\rho(a)|m) &= \sqrt{2}|m-1), \quad m \in \mathbb{N}^+, \\
\rho(a)|0) &= |-1), \\
\rho(a)|-n) &= (1/\sqrt{2})n|-n-1), \quad n \in \mathbb{N}^+.
\end{aligned} \tag{6.3}
$$

This shows the equivalence of extended (6.1) and (3.9). The subspace with basis $\{|m), m \in \mathbb{N}\}$ carries an irreducible representation of the su(1,1) algebra $l_3 = \frac{1}{2} + a^\dagger a$, $l_+ = a^\dagger$, $l_- = \frac{1}{2}(a + a^\dagger a^2)$,

$$
\begin{aligned}
\rho(l_3)|m) &= (\tfrac{1}{2} + m)|m), \\
\rho(l_+)|m) &= (1/\sqrt{2})(m+1)|m+1), \\
\rho(l_-)|m) &= (1/\sqrt{2})m|m-1), \quad m \in \mathbb{N},
\end{aligned} \tag{6.4}
$$

and the subspace with basis $\{|-n), n \in \mathbb{N}^+\}$ carries an irreducible representation of su(1,1), $l'_3 = \frac{1}{2} + a^\dagger a$, $l'_+ = \frac{1}{2}(a^\dagger + a^{\dagger 2} a)$, $l'_- = a$,

$$
\begin{aligned}
\rho(l'_3)|-n) &= (\tfrac{1}{2} - n)|-n), \\
\rho(l'_+)|-n) &= (1/\sqrt{2})(n-1)|-n+1), \\
\rho(l'_-)|-n) &= (1/\sqrt{2})n|-n-1), \quad n \in \mathbb{N}^+.
\end{aligned} \tag{6.5}
$$

The two realizations of su(1,1) are related by the map η introduced in Sec. II. (See Fig. 2.) Here we have used the ideal K_3 generated by $(E-\mu)\mathbf{1}$, $a^\dagger a\mathbf{1}$, $\mu = 1$. Note that $\rho(a)\mathbf{1} = a$. Thus both the H-invariant subspace W of $\mathrm{U}(H)/K_3$, and the quotient space of $H(H)/K_3$ modulo the H-invariant subspace W carry equivalent Hermitian representations of su(1,1) (with respect to S_1), while the Heisenberg algebra on $\mathrm{U}(H)/K_3$ leads to an indecomposable representation. The Heisenberg algebra, however, has been used to build the two su(1,1) realizations.

ACKNOWLEDGMENTS

Two of us (B.G. and M.L.) wish to thank Professor M. Moshinsky for stimulating discussions on this subject. B.G. and H.D.D. wish to thank Professor F. Zaccaria for the hospitality extended to them at the Dipartimento di Fisica Teorica of the Universitá di Napoli where this paper was completed.

[1]B. Gruber and A. U. Klimyk, J. Math. Phys. **25**, 755 (1984).
[2]B. Gruber and A. U. Klimyk, J. Math. Phys. **19**, 2009 (1978).
[3]B. Gruber, A. U. Klimyk, and Yu. F. Smirnov, Nuovo Cimento A **69**, 97 (1982); B. Gruber and R. Lenczewski, "Finite dimensional indecomposable representations of the Poincaré algebra," *Symmetries in Science II*, edited by B. Gruber and R. Lenczewski (Plenum, New York, 1987); R. Lenczewski and B. Gruber, J. Phys. A **19**, 1 (1986); B. Gruber and R. Lenczewski, *ibid.* **16**, 3703 (1983); B. Gruber, Proc. R. Irish Acad. A **82**, 13 (1982).
[4]Harish-Chandra, Am. J. Math. 77, 743 (1955).
[5]B. Gruber, R. Lenczewski, and M. Lorente, "Induced scalar products and unitarization of representations," preprint, Southern Illinois University, 1988.
[6]B. Gruber, H. D. Doebner, and Ph. Feinsilver, KINAM **4**, 241 (1982). To be more specific, (3,2) has been obtained with the help of (2.1), (2.2), and (2,5). All representations of the algebra H that are discussed in this paper are induced by $\hat{\rho}$. The representations $\hat{\rho}(E)$ are not diagonal.
[7]This procedure in the case of the Heisenberg algebra is as follows: Define $A(m,n,r) = (t^m/m!)(u^m/n!)(v^r/r!)$, with t, u, v commuting variables. For $B(m,n,r) = a^{\dagger m}a^n E^r$, $m,n,r \in \mathrm{N}$, define the bilinear product $\langle A,B \rangle = \Sigma A(m,n,r)B(m,n,r)$ of two vectors A,B. The contravariant representation of the algebra H is defined by the relation $\langle A, gB \rangle = \langle g^*A,B \rangle$ for $g \in H$. This contravariant representation given in terms of differential operators can be translated to the boson realization by the usual correspondence. See also Ph. Feinsilver, "Special functions and representations of su(2)," in *Symmetries in Science II*, edited by B. Gruber and R. Lenczewski (Plenum, New York, 1987).
[8]The construction of the representations of the compact Lie algebra su(2) by creation and annihilation operators is due to Schwinger, see *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965), pp. 225–279. The noncompact case was discussed by A. O. Barut and R. Raczka, see "Representations of the (2 + 1) Lorentz group," Proc. R. Soc. London Ser. A **287**, 532 (1965). L. C. Biedenharn and J. D. Louck have listed other boson realizations of the angular momentum operators, see *Encyclopedia of Mathematics and its Applications*, edited by G. C. Rota, *Vol. 9: The Racah–Wigner Algebra in Quantum Theory* (Addison–Wesley, Reading, 1981). In this volume, they describe the boson realization of the su(2) generators, in terms of pairs corresponding to the first and second fundamental Wigner operators (p. 124); the sympleton realization in terms of a boson operator and its conjugate (p. 245); the Holstein–Primakoff realization of an irreducible representation of the generators of su(2). Recently, L. C. Biedenharn and D. E. Flath have constructed tensor operators in SU(3) with the help of boson polynomials [Commun. Math. Phys. **93**, 143 (1984)] and with the same boson operators they have constructed irreducible representations of sl(3) and sl(4) [Can. J. Math. **37**, 710 (1985)].
[9]Y. Alhassid, F. Gürsey, and F. Iachello, Ann. Phys. (NY) **148**, 346 (1983).
[10]M. Moshinsky, "Boson realization of symplectic algebras and the matrix representation of their generators," in *Symmetries in Science II*, edited by B. Gruber and R. Lenczewski (Plenum, New York, 1987); M. Moshinsky, O. Castaños, E. Chacón, and C. Quesne, J. Math. Phys. **26**, 2107 (1985); O. Castaños, P. Kramer, and M. Moshinsky, *ibid.* **27**, 924 (1986); O. Castaños and M. Moshinsky, J. Phys. A **20**, 513 (1987).
[11]O. Castaños and M. Moshinsky, "Generating kernel for the boson formulation of the symplectic collective model," preprint UNAM, Mexico City.

Constraints on the total and kinetic energy of ground states in a class of potential models

A. K. Common
Mathematics Institute, University of Kent, Canterbury CT2 7NF, United Kingdom
A. Martin
CERN, CH-1211 Geneva 23, Switzerland

(Received 20 July 1988; accepted for publication 2 November 1988)

It is shown that the ground state energy in a central potential is a concave (convex) function of angular momentum l when this potential is a concave (convex) function of r^2. A result stronger than concavity is obtained when the Laplacian of the potential is negative. New bounds in the kinetic energy of these states are derived in terms of energy differences and for s states in terms of the wave function at the origin. These bounds can be used to give good estimates of the ground state energies for power law potentials.

I. INTRODUCTION

In recent work[1] we have obtained constraints on states bound by a central potential $V(r)$ belonging to a class for which energy-level ordering theorems may be proved. We will consider here in particular the energy $E(0,l)$ of the ground state of angular momentum l. In Sec. II we will demonstrate a previous result[2] that $E(0,l)$ is a convex (concave) function of l when the potential $V(r)$ is a convex (concave) function of r^2, and strengthen it when the Laplacian of $V(r)$ is negative for all $r>0$.

An integrated form of the concavity condition is

$$E(0,l+1) - E(0,l) < E(0,l) - E(0,l-1). \qquad (1.1)$$

This inequality can also be tightened when $V(r)$ has negative Laplacian, as will be demonstrated in Sec. II using a variational method applied by Martin in an earlier proof[3] of the concavity of $E(0,l)$.

Bounds on the kinetic energy $\langle T\rangle_{0,l}$ and mean square radii $\langle r^2\rangle_{0,l}$ for the above ground states of angular momentum l, have been obtained in earlier work by Martin in collaboration with Bertlmann.[4] They are in terms of energy differences between adjacent states, and in some cases only hold for integer (physical) values of l as they have been derived using sum rules. In Sec. III we demonstrate how they may be extended to general non-negative l and tightened for certain classes of potentials.

Finally in Sec. IV, improved upper and lower bounds on the wave function at the origin of the S-wave ground state are obtained that depend only on the expectation value of the kinetic energy for this state. As demonstrated by Common previously,[5] this type of bound can be useful in determining quark mass differences using experimental estimates of the above value of the wave function from the leptonic decay rates of quarkonium via the Van Royen–Weisskopf formula.[6]

There are several reasons why it is of interest to study the above types of constraints. First they enable one to obtain a general impression of the distribution of energy levels for given classes of potentials and vice versa, e.g., as discussed in Ref. 2, the distribution of the lowest energy levels indicate that the quark–antiquark potential is *not* a convex function of r^2. Also for specific potentials such as simple powers of r, precise estimates can be given for $E(0,l)$ without detailed numerical evaluation. Some results were presented in Ref. 1 and some improved estimates are given in Sec. III.

Finally, the bounds we have obtained have the merit of being rigorous. We feel that they should be published so that future investigations of potential models may have them to hand.

II. THE CONVEXITY AND CONCAVITY OF $E(0,l)$

In a previous work[1] we considered the two following classes of potentials.

Set A: $\forall r>0$, one of the following conditions hold:

(i) $D_\alpha V(r)>0,\quad 1\leqslant\alpha\leqslant 2,\ V(r)\geqslant 0;$

(ii) $D_\alpha V(r)\geqslant 0,\quad \alpha<1,\ V(r)\leqslant 0;$

(iii) $\left[r^2\dfrac{d^2}{dr^2}+(3-2\alpha)\dfrac{d}{dr}\right]V(r)\geqslant 0,$
$1<\alpha<2,\quad \dfrac{dV}{dr}\geqslant 0;$

where

$$D_\alpha V(r) \equiv \frac{d^2V}{dr^2} + (5-3\alpha)\frac{1}{r}\frac{dV}{dr} + \frac{2(1-\alpha)(2-\alpha)V(r)}{r^2}$$

(the special cases $\alpha=1$, $\alpha=2$ correspond to potentials convex in $1/r$ and r^2, respectively.)

Set B: $\forall r>0$, one of the following conditions hold:

(i) $D_\alpha V(r)\leqslant 0,\quad \alpha>2$ or $<1,\ V(r)\geqslant 0;$

(ii) $D_\alpha V(r)\leqslant 0,\quad 1\leqslant\alpha\leqslant 2,\quad V(r)\leqslant 0;$

(iii) $\left[r^2\dfrac{d^2}{dr^2}+(3-2\alpha)\dfrac{d}{dr}\right]V(r)\leqslant 0,\quad \alpha\geqslant 2.$

We showed that if $V(r)$ belongs to set $\left[{}^{\rm A}_{\rm B}\right]$ for $\alpha=\beta$, then for $l\geqslant 0$ and $n>k$ such that all integrals concerned exist,

$$\begin{aligned}&[2l+1+\beta+k\beta]\langle r^{2(\beta-1)+n\beta}\rangle_{0,l}\langle r^{2(\beta-1)+(k-1)\beta}\rangle_{0,l}\\ &\quad\lesseqgtr[2l+1+\beta+n\beta]\langle r^{2(\beta-1)+k\beta}\rangle_{0,l}\\ &\quad\times\langle r^{2(\beta-1)+(n-1)\beta}\rangle_{0,l},\end{aligned}\qquad (2.1)$$

where

$$\langle r^k \rangle_{0,l} = \int_0^\infty u_{0,l}^2(r) r^k \, dr,$$

and $u_{0,l}(r)$ is the reduced wave function for the ground state of angular momentum l.

In the following we will need to use a limiting form of (2.1) obtained by setting $n = k + \delta$ and letting $\delta \to 0_+$. This gives

$$\frac{\langle r^{2(\beta-1)+k\beta} \log r \rangle_{0,l}}{\langle r^{2(\beta-1)+k\beta} \rangle_{0,l}} \lessgtr \frac{\langle r^{2(\beta-1)+(k-1)\beta} \log r \rangle_{0,l}}{\langle r^{2(\beta-1)+(k-1)\beta} \rangle_{0,l}} + [2l+1+\beta+k\beta]^{-1} \tag{2.2}$$

when $V(r)$ is in the set $\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]$ for $\alpha = \beta$. The moment inequalities follow from the inequalities,[7]

$$-\left[\frac{w_{0,\lambda}'(z)}{w_{0,\lambda}(z)}\right]' \gtrless \frac{\lambda+1}{z^2}, \tag{2.3}$$

which hold for $V(r)$ in the set $\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]$ where $z \equiv r^\alpha$, $\lambda \equiv (2l - \alpha + 1)/2\alpha$, and $w_{0,\lambda}(z) \equiv [\alpha r^{\alpha-1}]^{1/2} u_{0,l}(r)$. In the original variables, (2.3) takes the form

$$-\left[\frac{u_{0,l}'(r)}{r^{\alpha-1} u_{0,l}}\right]' \gtrless \frac{(l+1)\alpha}{r^{1+\alpha}}, \tag{2.4}$$

when $V(r)$ is in the set $\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]$ for a given value for α.

The inequalities (2.4) hold for all $r, l \geqslant 0$ and will now be used to prove a lemma to the main theorem of this paper.

Lemma (2.1): If $V(r)$ is in the set $\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]$ for a given value of $\alpha > 1$, then

$$J(\alpha) \gtrless -\frac{2l+3-\alpha}{\alpha-1} \int_0^\infty u_{0,l}(r) \frac{\partial u_{0,l}(r)}{\partial l} \frac{dr}{r^\alpha}, \tag{2.5}$$

where

$$J(\alpha) = \frac{1}{\alpha-1} \int_0^\infty r^{1-\alpha} \int_0^r \left[\frac{2l+1}{r'^2} - \frac{\partial E(0,l)}{\partial l}\right] \times u_{0,l}^2(r') dr' \, dr. \tag{2.6}$$

Proof: The reduced wave function satisfies

$$-u_{0,l}'' + \frac{l(l+1)}{r^2} u_{0,l} + V(r) u_{0,l} = E(0,l) u_{0,l}. \tag{2.7}$$

Taking the limit of the Wronskian of two solutions of (2.7) for neighboring values of l gives

$$u_{0,l}(r) \frac{\partial u_{0,l}'(r)}{\partial l} - u_{0,l}'(r) \frac{\partial u_{0,l}(r)}{\partial l} = \int_0^r \left[\frac{2l+1}{r'^2} - \frac{\partial E(0,l)}{\partial l}\right] u_{0,l}^2(r') dr'. \tag{2.8}$$

Since the right-hand side is positive, $[(\partial u_{0,l}(r)/\partial l)/u_{0,l}(r)]$ is an increasing function of r and has one zero in $(0, \infty)$. As

$$\int_0^\infty u_{0,l}(r) \frac{\partial u_{0,l}}{\partial l}(r) dr = 0,$$

therefore

$$\int_0^r u_{0,l}(r') \frac{\partial u_{0,l}(r')}{\partial l} dr' < 0, \quad r > 0. \tag{2.9}$$

Using (2.4),

$$\begin{aligned} -\int_0^\infty u_{0,l}'(r) \frac{\partial u_{0,l}(r)}{\partial l} \frac{dr}{r^{\alpha-1}} &= \int_0^\infty \frac{d}{dr}\left[\frac{u_{0,l}'(r)}{r^{\alpha-1} u_{0,l}(r)}\right]\left[\int_0^r u_{0,l}(r') \frac{\partial u_{0,l}(r')}{\partial l} dr'\right] dr \\ &\gtrless -\int_0^\infty \frac{(l+1)\alpha}{r^{1+\alpha}} \left[\int_0^r u_{0,l}(r') \frac{\partial u_{0,l}(r')}{\partial l} dr'\right] dr \\ &= -(l+1) \int_0^\infty u_{0,l}(r) \frac{\partial u_{0,l}(r)}{\partial l} \frac{dr}{r^\alpha} \end{aligned} \tag{2.10}$$

for potentials $V(r)$ in set $\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]$ for given α. Using (2.8) after integrating by parts,

$$\int_0^\infty u_{0,l}(r) \frac{\partial u_{0,l}(r)}{\partial l} \frac{dr}{r^\alpha} = \frac{2}{\alpha-1} \int_0^\infty u_{0,l}'(r) \frac{\partial u_{0,l}(r)}{\partial l} \frac{dr}{r^{\alpha-1}} + J(\alpha), \tag{2.11}$$

where $\alpha > 1$ and $J(\alpha)$ is defined in (2.5). Using (2.10) in (2.11) immediately gives the required result.

We are now in a position to prove the following theorem.

Theorem 2.1: (a) If

$$\frac{d}{dr}\left[\frac{1}{r} \frac{dV}{dr}\right] \gtrless 0, \quad \forall r > 0,$$

then

$$\frac{\partial^2 E(0,l)}{\partial l^2} \gtrless 0, \; l \geqslant -\frac{1}{2}. \tag{2.12}$$

(b) If

$$\frac{d}{dr}\left[r^2 \frac{dV}{dr}\right] \leqslant 0, \quad \frac{dV}{dr} \geqslant 0, \quad \forall r > 0,$$

then

$$\frac{\partial^2 E(0,l)}{\partial l^2} + \frac{1}{l+1} \frac{\partial E(0,l)}{\partial l} \leqslant 0, \quad l \geqslant 0. \tag{2.13}$$

Proof: Integrating the definition of $J(2)$ in (2.6) by parts gives

$$J(2) = -(2l+1)[\langle r^{-2} \log r \rangle_{0,l} \langle 1 \rangle_{0,l} - \langle r^{-2} \rangle_{0,l} \langle \log r \rangle_{0,l}]. \tag{2.14}$$

The potentials satisfying the conditions (a) are in the set $\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]$ for $\alpha = 2$. Therefore, from (2.5),

$$(2l+1) \int_0^\infty u_{0,l}(r) \frac{\partial u_{0,l}(r)}{\partial l} \frac{dr}{r^2} \gtrless -J(2) \gtrless -\langle r^{-2} \rangle_{0,l}. \tag{2.15}$$

The right-hand inequality follows from (2.14) and (2.2) with $\beta = 2$, $k = -1$. However,

$$\frac{\partial^2 E(0,l)}{\partial l^2} = 2 \int_0^\infty u_{0,l}^2(r) \frac{dr}{r^2} + 2(2l+1) \int_0^\infty u_{0,l}(r) \frac{\partial u_{0,l}(r)}{\partial l} \frac{dr}{r^2}. \tag{2.16}$$

The result (2.12) then follows on using (2.15) in (2.16).

The potentials satisfying the conditions in (b) are in set

B for $\alpha=2$ and $\alpha=1$. Now (2.2) may be used twice with $\beta=1$ and first $k=0$ and then $k=-1$ to give

$$J(2)\geqslant\langle r^{-2}\rangle_{0,l}[1+(2l+1)/(2l+2)]. \tag{2.17}$$

In this case, using (2.16),

$$\frac{\partial^2E(0,l)}{\partial l^2}\leqslant 2\langle r^{-2}\rangle_{0,l}-2J(2)$$

$$\leqslant-\frac{2l+1}{l+1}\langle r^{-2}\rangle_{0,l}=-\frac{1}{l+1}\frac{\partial E(0,l)}{\partial l},$$

and the result (2.13) follows.

The inequalities (2.12) are equivalent to the convexity (concavity) of $E(0,l)$ for potentials $V(r)$ that are convex (concave) in r^2. They *cannot* be improved for the class of potentials considered in (a) since the inequalities are saturated by the harmonic oscillator potential for which $d/dr[(1/r)(dV/dr)]=0$, $\forall r>0$. The Coulomb potential $V(r)=-1/r$ satisfies the conditions for case (b) and we know for this potential that

$$\frac{\partial^2E(0,l)}{\partial l^2}+\frac{3}{2(l+1)}\frac{\partial E(0,l)}{\partial l}=0. \tag{2.18}$$

Therefore (2.13) is not saturated and there may be still scope for improvement in case (b).

An integrated form of concavity of $E(0,l)$ in l is given in (1.1) and holds for potentials concave in r^2. We now modify a previous argument[3] to show that (1.1) can be improved for potentials with negative Laplacian.

Theorem 2.3: If

$$\frac{d}{dr}\left[r^2\frac{dV}{dr}\right]\leqslant 0,\quad \forall r>0,$$

then

$$[E(0,l+1)-E(0,l)]$$
$$\leqslant[l/(l+2)][E(0,l)-E(0,l-1)]. \tag{2.19}$$

Proof: Following the discussion in Ref. 3, we use[7] $v(r)=ru_{0,l}(r)$ as a trial wave function in the variational estimate of $E(0,l+1)$ and obtain the inequality

$$E(0,l+1)$$
$$\leqslant\int_0^\infty\left[v'^2+\frac{(l+1)(l+2)v^2}{r^2}+V(r)v^2\right]dr\left(\int_0^\infty v^2\,dr\right)^{-1}$$
$$=\int_0^\infty[r^2u_{0,l}'^2+(l+1)(l+2)u_{0,l}^2+V(r)r^2u_{0,l}^2]dr(\langle r^2\rangle_{0,l})^{-1}$$
$$=\left\{\int_0^\infty r^2u_{0,l}\left[-u_{0,l}''+\frac{l(l+1)}{r^2}u_{0,l}+V(r)u_{0,l}\right]dr+(2l+3)\int_0^\infty u_{0,l}^2\,dr\right\}(\langle r^2\rangle_{0,l})^{-1}$$
$$=E(0,l)+(2l+3)/\langle r^2\rangle_{0,l}. \tag{2.20}$$

Notice that this inequality is valid for arbitrary potentials. Similarly by taking $v=u_{0,l}(r)/r$ as a trial wave function to estimate $E_{0,l}$ we obtain the inequality

$$E(0,l-1)\leqslant E(0,l)-(2l-1)\langle r^{-4}\rangle_{0,l}/\langle r^{-2}\rangle_{0,l}. \tag{2.21}$$

Previously the lower inequality (2.1) with $\beta=2$ was used to prove from (2.20) and (2.21) that (1.1) holds if $V(r)$ is concave in r^2. However, if $d/dr[r^2(dV/dr)]\leqslant 0$, $\forall r>0$, the lower moment inequalities (2.1) can be used with $\beta=1$. Take this value of β and $n=2$, $k=-3$, then this lower inequality gives

$$\langle r^2\rangle_{0,l}\langle r^{-4}\rangle_{0,l}\geqslant\langle r\rangle_{0,l}\langle r^{-3}\rangle_{0,l}\frac{2l+4}{2l-1}, \tag{2.22}$$

while with $n=1$, $k=-2$

$$\langle r\rangle_{0,l}\langle r^{-3}\rangle_{0,l}\geqslant\langle 1\rangle_{0,l}\langle r^{-2}\rangle_{0,l}\frac{2l+3}{2l}. \tag{2.23}$$

We take λ to be a positive parameter to be adjusted later and use (2.20)–(2.23) to prove the sequence of inequalities

$$[E(0,l-1)-E(0,l)]+\lambda[E(0,l+1)-E(0,l)]$$
$$\leqslant-(2l-1)\langle r^{-4}\rangle_{0,l}/\langle r^{-2}\rangle_{0,l}+\lambda(2l+3)\langle 1\rangle_{0,l}/\langle r^2\rangle_{0,l}$$
$$\leqslant\frac{\langle 1\rangle_{0,l}}{\langle r^2\rangle_{0,l}}\left\{-(2l-1)\frac{2l+4}{2l-1}\frac{2l+3}{2l}+\lambda(2l+3)\right\}. \tag{2.24}$$

By appropriate choice of λ, the right-hand side can be set equal to zero and (2.19) follows.

Inequality (2.19) is an appreciable improvement on (1.1) especially for small l. However, (2.19) is again *not* "saturated" for the Coulomb potential where, for example, when $l=1$ the ratio of the left-hand side to right-hand side is $\frac{5}{9}$. Here again there is the possibility for improvement.

III. BOUNDS ON $\langle T\rangle_{0,l}$ and $\langle r^2\rangle_{0,l}$

In a previous work Martin and Bertlmann[4] obtained the following lower bounds on the average value of the kinetic energy for the ground state S and P waves:

$$\langle T\rangle_{0,0}\geqslant\tfrac{3}{4}[E(0,1)-E(0,0)], \tag{3.1a}$$
$$\langle T\rangle_{0,1}\geqslant\tfrac{5}{4}[E(0,2)-E(0,1)]. \tag{3.1b}$$

These inequalities hold for general central potentials. The method used to derive them involved sum rules over physical states, and therefore could not be generalized to noninteger values of l. We demonstrate now how they can indeed be generalized to all $l\geqslant 0$.

Theorem 3.1: For general potentials $V(r)$ and all $l\geqslant 0$,

$$\langle T\rangle_{0,l}\geqslant[(2l+3)/4][E(0,l+1)-E(0,l)]. \tag{3.2}$$

Proof: We have with the usual normalization $\int_0^\infty u_{0,l}^2\,dr=1$,

$$\langle T\rangle_{0,l}=\int_0^\infty\left[u_{0,l}'^2+\frac{l(l+1)}{r^2}u_{0,l}^2\right]dr$$
$$=\int_0^\infty\left[u_{0,l}'-\frac{l+1}{r}u_{0,l}\right]^2dr. \tag{3.3}$$

Now consider

$$I(\lambda)=\int_0^\infty\left[u'_{0,l}-(l+1)\frac{u_{0,l}}{r}-\lambda r u_{0,l}\right]^2dr$$

$$=\int_0^\infty\left[u'_{0,l}-(l+1)\frac{u_{0,l}}{r}\right]^2dr-2\lambda\int_0^\infty[ru_{0,l}u'_{0,l}$$

$$-(l+1)u^2_{0,l}]dr+\lambda^2\int_0^\infty r^2u^2_{0,l}\,dr \tag{3.4}$$

$$=\langle T\rangle_{0,l}+\lambda(2l+3)+\lambda^2\langle r^2\rangle_{0,l}. \tag{3.5}$$

Since $I(\lambda)\geqslant 0$ for all real λ,

$$4\langle T\rangle_{0,l}\langle r^2\rangle_{0,l}\geqslant(2l+3)^2. \tag{3.6}$$

Then by using (2.20) in the form

$$E(0,l+1)-E(0,l)<(2l+3)/\langle r^2\rangle_{0,l} \tag{3.7}$$

we obtain the required inequality. It cannot be improved for general potentials since it becomes an equality for the harmonic oscillator potential.

Corresponding *upper* bounds to $\langle T\rangle_{0,l}$ in terms of energy differences can be obtained for certain classes of potentials as in the following theorem.

Theorem 3.2: If

$$\frac{d}{dr}\left[\frac{1}{r}\frac{dV}{dr}\right]<0,\quad \forall r>0,$$

then, for all $l\geqslant 1$,

$$\langle T\rangle_{0,l}\leqslant[(2l+3)/4][E(0,l)-E(0,l-1)]. \tag{3.8}$$

Proof: The above class of potentials are in set B defined in Sec. II for $\alpha=2$. Therefore from inequality (3.6) of Ref. 1

$$\langle T\rangle_{0,l}\leqslant\frac{2l+3}{4}\frac{\partial E(0,l)}{\partial l} \tag{3.9}$$

$$\leqslant\frac{2l+3}{4}[E(0,l)-E(0,l-1)], \tag{3.10}$$

where in the last step we have used the concavity of $E(0,l)$.

This upper bound complements the lower bound (3.2) and is again saturated by the harmonic oscillator potential. The factor multiplying the energy differences is the same in the upper and lower bounds. By restricting the class of potentials further, upper and lower bounds with the same energy difference but different multiplying factors may be obtained. An example is given by the following result.

Corollary 3.1: If

$$\frac{d}{dr}\left[\frac{1}{r}\frac{dV}{dr}\right]\leqslant 0\quad\text{and}\quad\frac{d}{dr}\left[r^2\frac{dV}{dr}\right]>0,\quad\forall r>0,$$

then

$$\langle T\rangle_{0,l}\leqslant[(2l+9)/4][E(0,l+1)-E(0,l)]. \tag{3.11}$$

Proof: As discussed earlier in Sec. II,

$$\int_0^r\frac{\partial u_{0,l}(r')}{\partial l}u_{0,l}(r')dr'<0.$$

Therefore

$$\frac{\partial E(0,l)}{\partial l}+\frac{\partial}{\partial l}\langle T\rangle_{0,l}$$

$$=2\int_0^\infty\left[r\frac{dV}{dr}+V(r)\right]u_{0,l}\frac{\partial u_{0,l}}{\partial l}dr$$

$$=-2\int_0^\infty\frac{d}{dr}\left[r\frac{dV}{dr}+V(r)\right]\int_0^r u_{0,l}(r')\frac{\partial u_{0,l}(r')dr'}{\partial l}$$

$$\geqslant 0. \tag{3.12}$$

Integrating (3.12) over $(l,l+1)$,

$$\langle T\rangle_{0,l}<E(0,l+1)-E(0,l)+\langle T\rangle_{0,l+1}. \tag{3.13}$$

Using (3.8) $l\to l+1$ gives immediately (3.11). Unfortunately this inequality is not saturated by the Coulomb potential for small l. An improved inequality may be obtained by further restricting the class of potential.

Corollary 3.1a: If

$$\frac{d}{dr}\left[\frac{1}{r}\frac{dV}{dr}\right]<0\quad\text{and}\quad\frac{d}{dr}\left[r\frac{dV}{dr}\right]>0,$$

then

$$\langle T\rangle_{0,l}\leqslant\frac{2l+5}{4}[E(0,l+1)-E(0,l)]. \tag{3.14}$$

The proof is similar to that of the previous corollary.

We now obtain associated bounds on $\langle r^2\rangle_{0,l}$.

Theorem 3.3: If

$$\frac{d}{dr}\left[\frac{1}{r}\frac{dV}{dr}\right]<0,\quad\forall r>0,$$

then

$$\frac{2l+3}{E(0,l+1)-E(0,l)}\geqslant\langle r^2\rangle_{0,l}\geqslant\frac{2l+3}{E(0,l)-E(0,l-1)} \tag{3.15}$$

and alternatively,

$$\langle r^2\rangle_{0,l}\geqslant\frac{(2l+3)^2}{(2l+9)[E(0,l+1)-E(0,l)]}, \tag{3.16}$$

if in addition

$$\frac{d}{dr}r^2\frac{dV}{dr}\geqslant 0.$$

These inequalities hold for all $l\geqslant 0$ except that in the right-hand inequality in (3.15), $l\geqslant 1$.

Proof: Using (3.6), which holds for all potentials, and the upper bounds on $\langle T\rangle_{0,l}$ given by (3.8) and (3.11), the right-hand sides of (3.15) and (3.16) are obtained immediately. The left-hand side of (3.15) is just (3.7). The inequalities (3.15) are again saturated in the case of the harmonic oscillator potential. As shown previously, one has upper and lower bounds that differ only by the energy difference used or by the multiplying factor of the inverse of this difference.

In Sec. IV of Ref. 1, we used bounds on $\langle T\rangle_{0,l}$ depending on $\partial E(0,l)/\partial l$ to obtain for power law potentials $V(r)=r^\nu$, $\nu>0$, both upper and lower bounds on $E(0,L)/E(0,l)$ with $L>l\geqslant 0$ and also for $E(0,l)$ itself. For $0<\nu\leqslant 2$, we found

$$\left[\frac{l+\frac{5}{2}}{l+\frac{3}{2}}\right]^{2\nu/(\nu+2)} \leqslant \frac{E(0,l+1)}{E(0,l)} \leqslant \left[\frac{l+1+(\nu+4)/4}{l+(\nu+4)/4}\right]^{2\nu/(\nu+2)}, \quad (3.17)$$

while for $\nu \geqslant 2$ one may similarly prove that

$$\left[\frac{l+1+(\nu+4)/4}{l+(\nu+4)/4}\right]^{2\nu/(\nu+2)} \leqslant \frac{E(0,l+1)}{E(0,l)} \leqslant \left[\frac{l+\frac{5}{2}}{l+\frac{3}{2}}\right]^{2\nu/(\nu+2)}. \quad (3.18)$$

Then using

$$\lim_{l\to\infty} \frac{E(0,l)}{l^{2\nu/(\nu+2)}} = \left(\frac{\nu}{2}\right)^{2/(\nu+2)} \frac{\nu+2}{\nu},$$

one can obtain from (3.17) and (3.18) the following bounds on $E(0,l)$ itself:

$$\left(\frac{\nu+2}{\nu}\right)\left(\frac{\nu}{2}\right)^{2/(\nu+2)}\left(l+\frac{\nu+4}{4}\right)^{2\nu/(\nu+2)} \leqslant E(0,l) \leqslant \left(\frac{\nu+2}{\nu}\right)\left(\frac{\nu}{2}\right)^{2/(\nu+2)}\left(l+\frac{3}{2}\right)^{2\nu/(\nu+2)},$$

$$\text{for} \quad 0 \leqslant \nu \leqslant 2, \quad (3.19)$$

and

$$\left(\frac{\nu+2}{\nu}\right)\left(\frac{\nu}{2}\right)^{2/(\nu+2)}(l+3/2)^{2\nu/(\nu+2)} \leqslant E(0,l) \leqslant \left(\frac{\nu+2}{\nu}\right)\left(\frac{\nu}{2}\right)^{2/(\nu+2)}\left(l+\frac{\nu+4}{4}\right)^{2\nu/(\nu+2)},$$

$$\text{for} \quad \nu \geqslant 2. \quad (3.20)$$

However, improved upper bounds on $E(0,l+1)/E(0,l)$ and lower bounds on $E(0,l)$ can be obtained from inequality (3.2) as we will now demonstrate. For $V(r) = r^\nu$ we have from the Virial theorem that

$$\langle T\rangle_{0,l} = [\nu/(2+\nu)]E(0,l) \quad (3.21)$$

so that inserting

$$(\nu/(2+\nu))E(0,l) \geqslant [(2l+3)/4][E(0,l+1) - E(0,l)] \quad (3.22)$$

in (3.2) and rearranging, we obtain

$$\frac{E(0,l+1)}{E(0,l)} \leqslant \frac{(2l+3)/4 + \nu/(2+\nu)}{(2l+3)/4}. \quad (3.23)$$

By repeating application of this bound, when $L \geqslant l$,

$$\frac{E(0,L)}{E(0,l)} \leqslant \frac{\Gamma(L+\frac{3}{2}+2\nu/(\nu+2))\Gamma(l+\frac{3}{2})}{\Gamma(l+\frac{3}{2}+2\nu/(\nu+2))\Gamma(L+\frac{3}{2})}. \quad (3.24)$$

By taking the limit $L\to\infty$,

$$E(0,l) \geqslant \left(\frac{\nu}{2}\right)^{2/(\nu+2)}\left(\frac{\nu+2}{\nu}\right)\frac{\Gamma(l+\frac{3}{2}+2\nu/(\nu+2))}{\Gamma(l+\frac{3}{2})}. \quad (3.25)$$

We will now give numerical examples that show that (3.24) and (3.25) give better bounds than do (3.17) and (3.20).

Consider the case when $l=0$. Set $\nu=4$ in (3.19)

$$1.718 = (\tfrac{3}{2})^{4/3} \leqslant E(0,1)/E(0,0) \leqslant (\tfrac{5}{3})^{4/3} \cong 1.975, \quad (3.26)$$

while the "new" upper bound (3.23) gives

$$E(0,1)/E(0,0) \leqslant \tfrac{17}{9} \simeq 1.778. \quad (3.27)$$

Similarly when $\nu=1$ in (3.17),

$$1.405 = (\tfrac{5}{3})^{2/3} \leqslant E(0,1)/E(0,0) \leqslant (\tfrac{9}{5})^{2/3} = 1.482 \quad (3.28a)$$

and the new upper bound gives

$$E(0,1)/E(0,0) \leqslant \tfrac{13}{9} \sim 1.444. \quad (3.28b)$$

We see that the new upper bound is a substantial improvement for $\nu=4$ and a small improvement for $\nu=1$. When it is combined with the "old" lower bound, estimates of the above ratio accurate within a few percent are obtained. Similarly in Ref. 1 we used the lower bounds in (3.18) and (3.19) and variational upper bounds to give, for $\nu=1$,

$$2.19301 \leqslant E(0,0) \simeq 2.338\,107\,41\cdots \leqslant 2.380 \quad (3.29a)$$

and, for $\nu=4$

$$3.24 \leqslant E(0,0) \simeq 3.8 \leqslant 3.91. \quad (3.29b)$$

Our new lower bound (3.26) can be used to give, for $\nu=1$,

$$E(0,0) \geqslant 2.3079 \quad (3.30a)$$

and, for $\nu=4$,

$$E(0,0) \geqslant 3.6777, \quad (3.30b)$$

which improve on those given in (3.29). The old upper bounds with the new lower bounds give good estimates for $E(0,0)$ to within a few percent error.

IV. BOUNDS ON THE WAVE FUNCTION AT THE ORIGIN IN TERMS OF THE KINETIC ENERGY

In Ref. 5 we obtained bounds on $|u'_{0,0}|^2$ in terms of $\langle T\rangle_{0,0}$. Here $|u'_{0,0}|^2$, for instance, enters in the expression of leptonic decay rates of quarkonium via the Van Royen–Weisskopf formula.[7] The bounds were given by

$$|u'_{0,0}(0)|^2 > \frac{1}{4}\langle T\rangle_{0,0}^{3/2}, \quad \text{if} \quad \frac{dV}{dr} > 0, \quad \forall r > 0, \quad (4.1)$$

$$|u'_{0,0}(0)|^2 < 3\langle T\rangle_{0,0}^{3/2}, \quad \text{if} \quad \frac{d}{dr} r \frac{dV}{dr} \geqslant 0 \quad \text{and} \quad \frac{dV}{dr} \geqslant 0. \quad (4.2)$$

The upper and lower bounds are in the ratio 12:1, and so rather weak. Here we shall derive tighter bounds on $u'_{0,0}(0)$ in terms of the kinetic energy and, in some cases these bounds will be valid for $n \neq 0$, i.e., arbitrary radial excitations.

First we derive a series of upper bounds on $|u'(0)|^2$, increasingly tighter, with more and more restricted classes of potentials.

Theorem 4.1 (boundary case Coulomb potential):

$$|u'_{0,0}(0)|^2 \leqslant 4\langle T\rangle_{0,0}^{3/2} \quad (4.3)$$

if

$$\Delta V(r) \geqslant 0, \quad \forall r > 0, \quad \text{i.e.,} \quad \frac{d}{dr} r^2 \frac{dV}{dr} > 0.$$

Proof: We have

$$|u'(0)|^2 = \int u^2 \frac{dV}{dr}\, dr.$$

Now if

$$\frac{d}{dr} r^2 \frac{dV}{dr} \geqslant 0,$$

$$\left\langle r\frac{dV}{dr}\right\rangle\left\langle\frac{1}{r^2}\right\rangle - \left\langle\frac{dV}{dr}\right\rangle\left\langle\frac{1}{r}\right\rangle$$

$$= \frac{1}{2}\int \frac{dx\,dy}{xy} u^2(x)u^2(y) \times [W(x) - W(y)][x-y] > 0, \tag{4.4}$$

with

$$W(x) = x^2 \frac{dV}{dx},$$

then

$$|u'_{0,0}(0)|^2 < 2\langle T\rangle_{0,0}\langle 1/r^2\rangle/\langle 1/r\rangle. \tag{4.5}$$

But, by using the inequalities of Ref. 1, we have

$$\langle 1/r^2\rangle < 2(\langle 1/r\rangle)^2 \tag{4.6}$$

and

$$\langle 1/r^2\rangle < 2\langle T\rangle_{0,0}, \tag{4.7}$$

if $\Delta V > 0$, which gives (4.3) by substitution in (4.5). Inequality (4.1) is optimal in the sense that it is saturated for the case of a pure Coulomb potential.

Corollary: If $\Delta V \geqslant 0$ and if $V(r) \to 0$ for $r \to \infty$,

$$|u'_{0,0}(0)|^2 < 4|E(0,0)|^{3/2}. \tag{4.8}$$

Proof: if $(d/dr)r^2(dV/dr) > 0$ and if $V(\infty) = 0$, it is easy to get by integration

$$r\frac{dV}{dr} + V(r) < 0, \tag{4.9}$$

and hence

$$2\langle T\rangle + \langle V\rangle = E + \langle T\rangle < 0.$$

An alternative derivation of this result consists in integrating an inequality recently obtained by Benguria[8]:

$$u'_{0,0} + \sqrt{-E}\,u_{0,0} - u_{0,0}/r > 0. \tag{4.10}$$

It is possible that Theorem 4.1 or at least the corollary applies to radial excitations, but we have no proof of this.

Theorem 4.2 (boundary case, $\log r$ potential): If

$$\frac{d}{dr} r\frac{dV}{dr} > 0,$$

then

$$|u'_{n,0}(0)|^2 < 2\langle T\rangle_{n,0}^{3/2}. \tag{4.11}$$

Here the inequality holds for *any* radial excitations.

Proof: Following the same technique as in (4.4), we have

$$\left\langle r\frac{dV}{dr}\right\rangle\left\langle\frac{1}{r}\right\rangle - \left\langle\frac{dV}{dr}\right\rangle\langle 1\rangle \geqslant 0$$

if

$$\frac{d}{dr} r\frac{dV}{dr} \geqslant 0.$$

Hence

$$|u'_{n,0}(0)|^2 < 2\langle T\rangle_{n,0}\frac{\int (u^2/r)dr}{\int u^2\,dr}. \tag{4.12}$$

However, consider the quantity

$$X = \frac{\int u'^2\,dr \int u^2\,dr}{(\int (u^2/r)dr)^2}. \tag{4.13}$$

We have, for arbitrary $Z > 0$,

$$\int u'^2\,dr - Z\int \frac{u^2}{r}\,dr > -\frac{Z^2}{4}\int u^2\,dr,$$

by application of the variational principle to the Hamiltonian

$$-\frac{d^2}{dr^2} - \frac{Z}{r}. \tag{4.14}$$

Hence

$$X \geqslant \frac{1}{x^2}\left(Zx - \frac{Z^2}{4}\right), \quad \forall Z > 0,$$

where

$$x = \int (u^2/r)dr\left(\int u^2\,dr\right)^{-1},$$

and choosing $Z = 2x$, we get $X \geqslant 1$. For $n = 0$ the inequality is almost, but not exactly saturated:

$$|u'(0)|^2/\langle T\rangle^{3/2} = 1.9513..., \quad \text{for} \quad V = \log r.$$

This is because the wave function of (4.14) is not that of a logarithmic potential.

Theorem 4.3 (boundary case linear potential): If $V'' > 0$, then

$$|u'_{n,0}(0)|^2 < (3/E_0)^{3/2}\langle T\rangle_{n,0}^{3/2} < 1.455\langle T\rangle_{n,0}^{3/2}, \tag{4.15}$$

where E_0 is the first zero of the Airy function.

Proof: By using the same technique,

$$\left\langle r\frac{dV}{dr}\right\rangle\langle 1\rangle - \left\langle\frac{dV}{dr}\right\rangle\langle r\rangle > 0$$

and

$$|u'_{n,0}(0)|^2 < 2\langle T\rangle_{n,0}\frac{\int u^2\,dr}{\int ru^2\,dr}.$$

Now we look for the minimum of

$$X' = \frac{\int u'^2\,dr(\int ru^2\,dr)^2}{(\int u^2)^3}$$

using the same type of variational argument; we get

$$\int u'^2\,dr + \mu^3\int ru^2\,dr > E_0\mu^2\int u^2\,dr,$$

where E_0 is the ground state energy of the Hamiltonian

$$-\frac{d^2}{dr^2} + r,$$

$$E_0 \cong 2.3381....$$

After optimization with respect to μ, (4.15) follows trivially. For $n = 0$ the inequality is saturated in the case of a linear potential. For $n > 0$ it is obviously not saturated since the wave function at the origin is independent of the excitation.

Finally, we give the following theorem.

Theorem 4.4 (limit case, harmonic oscillator): If

$$\frac{d}{dr}\frac{1}{r}\frac{dV}{dr} \geqslant 0,$$

then

$$|u'_{n,0}(0)|^2 < \tfrac{4}{3}\langle T\rangle^{3/2}_{n,0}. \tag{4.16}$$

Proof: By the same kind of argument

$$|u'_{n,0}(0)|^2 < 2\langle T\rangle_{n,0}\frac{\int ru^2\,dr}{\int r^2u^2\,dr}, \tag{4.17}$$

and by the Schwarz inequality

$$|u'_{n,0}(0)|^2 < 2\langle T\rangle_{n,0}\left(\frac{\int u^2\,dr}{\int r^2u^2\,dr}\right)^{1/2}.$$

Then

$$X'' = \frac{\int u'^2dr \int r^2u^2\,dr}{(\int u^2\,dr)^2}$$

is minimized by an harmonic oscillator wave function and we get

$$|u'_{n,0}(0)|^2 < \tfrac{4}{3}\langle T\rangle^{3/2}_{n,0}. \tag{4.18}$$

In the special case of $n = 0$, this can be slightly improved by noticing, as was done some years ago by Common,[5] that

$$f_k = k\int r^{k-3}u^2\,dr$$

form a Stieljes sequence if $dV/dr > 0$, i.e.,

$$f_k \leqslant (f_{k+\delta}f_{k-\delta})^{1/2},\ldots .$$

This allows us to improve the Schwarz inequality. So, if

$$\frac{d}{dr}\frac{1}{r}\frac{dV}{dr} > 0$$

and

$$\frac{dV}{dr} > 0,$$

then

$$|u'_{0,0}(0)|^2 < \sqrt{\tfrac{5}{3}}\langle T\rangle^{3/2}_{0,0}. \tag{4.19}$$

For a pure harmonic oscillator

$$|u'_{0,0}(0)|^2/\langle T\rangle^{3/2}_{0,0} = (4/\sqrt{\pi})(\tfrac{2}{3})^{3/2} = 1.257,$$

while (4.19) gives 1.291, very close indeed.

We could continue this game with tighter and tighter conditions, but it is only for some specific cases, like the Coulomb potential and linear potential, that we can get optimal results for the ground state wave function.

We turn now to lower bounds of $u'_{0,0}(0)$, i.e., to improvements of (4.1). We have obtained the following theorem.

Theorem 4.5: If $dV/dr \geqslant 0$ and if $V(r)$ is in the set B for $\alpha = \beta_1$,

$$|u'_{0,0}(0)|^2 \geqslant [8/\sqrt{3}(1+\beta_1)^{3/2}]\langle T\rangle^{3/2}_{0,0}. \tag{4.20}$$

In the special case $V = r^2$ we can take $\beta_1 = 2$ so that the coefficient in (4.20) becomes $\frac{8}{9} \simeq 0.889$ not too far from the upper bound.

Proof: We already said that the

$$f_k = k\int r^{k-3}u^2\,dr$$

form a Stieljes sequence. Now

$$\int_0^\infty \frac{u^2\,dr}{r^{3-\epsilon}} \simeq \frac{1}{\epsilon}|u'(0)|^2, \quad \text{for} \quad \epsilon\to 0,$$

so that

$$f_0 = |u'(0)|^2.$$

So[9]

$$\langle r^{-2}\rangle_{0,0} = f_1 < (f_3)^{1/3}(f_0)^{2/3} = (3)^{1/3}|u'_{0,0}(0)|^{4/3}, \tag{4.21}$$

but from Ref. 1

$$\langle T\rangle_{0,0} \leqslant \tfrac{1}{4}(1+\beta_1)\langle r^{-2}\rangle_{0,0}. \tag{4.22}$$

By combining (4.21) and (4.22) we get (4.20).

Theorem 4.6: The lower bound (4.1) can easily be generalized to arbitrary radial excitations in the form

$$|u'_{n,0}(0)|^2 \geqslant \frac{1}{4(n+1)}\langle T\rangle^{3/2}_{n,0}, \quad \text{if} \quad \frac{dV}{dr} \geqslant 0. \tag{4.23}$$

Proof: The proof is a carbon copy of the one given in Ref. 5. We notice that if $dV/dr > 0$ it is still true that

$$|u'_{n,0}(r)|^2 < |u'_{n,0}(0)|^2, \tag{4.24}$$

because $u'^2 + (E-V)u^2$ is decreasing and hence $u'^2(r)$ is less than $u'^2(0)$ as long as $E - V$ is positive, and, for $(E-V) < 0$, $|u'|$ decreases. Therefore, by calling $r_1,\ldots,r_n$ the zeros of $u_{n,0}$ and $R_1,\ldots,R_{n+1}$ the maxima of $|u_{n,0}|$, we have

$$\begin{aligned}\int_{r_p}^{R_{p+1}} u'^2\,dr &< |u'(0)|\int_{r_p}^{R_{p+1}} |u'|dr \\ &= |u'(0)|\left[\int_{r_p}^{R_{p+1}} 2|uu'|dr\right]^{1/2} \\ &< \sqrt{2}|u'(0)|\left(\int_{r_p}^{R_{p+1}} u'^2\,dr\right)^{1/4}\left(\int_{r_p}^{R_{p+1}} u^2\,dr\right)^{1/4}.\end{aligned} \tag{4.25}$$

Hence

$$\left(\int_0^{R_1} u'^2\,dr\right)^3 + \left(\int_{R_1}^{r_1} u'^2\,dr\right)^3 + \cdots + \left(\int_{R_{n+1}}^{\infty} u'^2\,dr\right)^3 < 4|u'(0)|^4\int_0^\infty u^2\,dr. \tag{4.26}$$

Then by using the fact that, for positive a's,

$$a_1^3 + a_2^3 + \cdots + a_p^3 > (1/p^2)(a_1 + a_2 + \cdots + a_p)^3,$$

we easily get (4.23).

Notice that for positive power potentials this lower bound has the correct qualitative behavior for large n. For instance, for a linear potential $E(n,0) \sim \langle T\rangle_{n,0} \sim n^{2/3}$, so that the lower bound is a positive constant for large n.

ACKNOWLEDGMENT

One of us (A.M.) wishes to thank R. Benguria for stimulating correspondence.

[1]A. K. Common and A. Martin, J. Phys. A: Math. Gen. **20**, 424 (1987).
[2]A. K. Common and A. Martin, Europhys. Lett. **4**, 1349 (1987).
[3]A. Martin, Lectures at the XXVI Internationale Universitäts Wochen for Kernphysik, Schladming 1987, to be published.
[4]R. A. Bertlmann and A. Martin, Nucl. Phys. B **168**, 111 (1980).
[5]A. K. Common, Nucl. Phys. B **224**, 229 (1983).
[6]R. P. Van Royen and V. F. Weisskopf, Nuovo Cimento A **50**, 617 (1967).
[7]This method was developed from a suggestion by T. Regge.
[8]M. S. Ashbaugh and R. Benguria, Phys. Lett. A **131**, 275 (1988).
[9]G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities* (Cambridge U.P., Cambridge, 1952), 2nd ed., p. 140.

Coexistence of symmetric periodic points in the standard map

Kiyotaka Tanikawa
National Astronomical Observatory, Mizusawa, Iwate 023, Japan
Yoshihiro Yamaguchi
Liberal Arts and Science, Teikyo University of Technology, Ichihara, Chiba, Japan

(Received 5 November 1987; accepted for publication 16 November 1988)

The problem of the coexistence of symmetric periodic points of the standard map has been investigated. First, it is shown that there is another reversibility in addition to the one already known. The dominant axis of this reversibility is the remaining coordinate axis. Using these reversibilities, simple recursion formulas that determine the positions of symmetric periodic points have been derived. Second, bifurcation structures of symmetric periodic points from the fixed point at the origin have been investigated. It is shown that periodic points bifurcate along both the dominant axes of two reversibilities every time the rotation number of the fixed point passes through a rational number. Third, the problem of the coexistence of symmetric periodic points has been numerically investigated with the aid of the recursion formulas. A devil's staircase form of the evolution of the outermost Kolmogorov–Arnold–Moser (KAM) curve of the fixed point with the change of the perturbation parameter is observed. This indicates that bifurcated periodic points do not disappear by inverse bifurcation as long as they are in the invariant region of the fixed point.

I. INTRODUCTION

In the present paper, we consider the problem of the coexistence of symmetric periodic points for the standard map[1] that is given by

$$T\colon\ x_{n+1} = x_n + h(\theta_n) \ \text{ and } \ \theta_{n+1} = \theta_n + x_{n+1} \pmod 1, \tag{1}$$

where

$$h(\theta_n) = -(K/2\pi)\sin 2\pi\theta_n. \tag{2}$$

The phase space of the map is a torus. We take the square $-\frac{1}{2} < \theta \leqslant \frac{1}{2}$, $-\frac{1}{2} < x \leqslant \frac{1}{2}$ as the fundamental domain of T.

In the preceding paper,[2] we have shown that periodic points are dense in the phase space of the map if the value of the perturbation parameter K goes to infinity. Trivially, periodic points are dense when the system is integrable at $K = 0$. Then, it is natural to ask whether or not periodic points are dense for the values of K in between. In the preceding paper,[2] we have shown that periodic points of *any* ($\geqslant 2$) period are dense. In particular, we have shown that two-cycles lie on a curve and that the curve itself densely fills the phase space. On the other hand, in the integrable case, periodic points lie on different lines and a countable number of lines densely fill the phase space. The situation of the problem for $0 < K < \infty$ is expected to be somewhere between the cases $K = 0$ and $K \to \infty$.

The purpose of the present paper is to investigate the problem of the coexistence of periodic orbits on the coordinate axes to obtain evidence of the denseness of the distribution of periodic points. Greene *et al.*[3] suggested that the symmetry axis is a preferred position of periodic points. This will be proved in Sec. III.

In Sec. II, we investigate the coexistence of periodic points on the coordinate axes utilizing reversibility. It is found that the standard map has another reversibility, in addition to the one already known. This suggests that if periodic points are dense in one of the coordinate axes, then they should be dense in the other. We derive recursion formulas that determine the positions of symmetric periodic points. These formulas turn out to be very useful.

In Sec. III, the structure of bifurcations of periodic points from the origin is investigated. It is shown that periodic points bifurcate on the coordinate axes. This is strong evidence that periodic points should be dense in the coordinate axes.

In Sec. IV, the coexistence of periodic points is investigated numerically. In particular, the evolution of the outermost KAM curve of the fixed point with the increase of the perturbation parameter is investigated in relation to the stability of bifurcated periodic points. It is conjectured that bifurcated periodic points of even periods are unstable from the beginning, while bifurcated periodic points of odd periods become unstable after they get out of the invariant region of the fixed point. A devil's staircase form of the evolution indicates that daughter periodic points of the fixed point do not disappear with the increase of the value of K.

In the following sections, we denote the Gauss notation by [ ].

II. REVERSIBILITY AND PERIODIC POINTS ON THE COORDINATE AXES

The standard map is reversible.[4,5] The map is written as a product of two involutions

$$T = I_2 I_1 \tag{3}$$

with

$$I_1\begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} x + h(\theta) \\ -\theta \end{bmatrix}, \quad I_2\begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} x \\ x - \theta \end{bmatrix}, \tag{4}$$

where $I_1^2 = I_2^2 = 1$ and $\det I_1 = \det I_2 = -1$. The dominant axes are $\theta = 0$ and $\theta = \frac{1}{2}$, while the complementary axes

 0022-2488/89/030608-09$02.50

are $\theta = (x-1)/2$, $\theta = x/2$, and $\theta = (x+1)/2$.

The map has another reversibility. In fact, it can be written as a product of two involutions

$$T = J_2 J_1 \tag{5}$$

with

$$J_1\begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} -x \\ \theta - x \end{bmatrix}, \quad J_2\begin{bmatrix} x \\ \theta \end{bmatrix} = \begin{bmatrix} -x + h(\theta - x) \\ \theta - 2x + h(\theta - x) \end{bmatrix}, \tag{6}$$

where $J_1^2 = J_2^2 = 1$ and $\det J_1 = \det J_2 = -1$. The matrices J_1 and J_2 are explicitly given by

$$J_1 = \begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix}, \quad J_2 = \begin{bmatrix} -1-\alpha & \alpha \\ -2-\alpha & 1+\alpha \end{bmatrix}, \tag{7}$$

with $\alpha = h(\theta - x)/(\theta - x)$. The dominant and complementary axes of this reversibility are $x = 0$ and $2x = h(\theta - x)$, respectively. The latter changes form as the parameter K changes. Actually, it is not a straight line. The complementary axis together with other symmetry axes are shown in Fig. 1 for several values of K.

We consider periodic points of the x and θ axes, i.e., the dominant axes of two reversibilities. Let us review important properties of these points.[3,6]

Property 1: A point P on the dominant axis is a fixed point of T^{2n} if and only if the point $T^n P$ is on the same dominant axis.

Property 2: A point P on the dominant axis is a fixed point of T^{2n+1} if and only if the point $T^{n+1}P$ is on the corresponding complementary axis.

Periodic points having Property 1 or 2 are called symmetric. There are symmetric periodic points that have two points on the complementary axis. We do not consider these points.

Let us formulate the conditions for symmetric periodic points. Let a sequence of points (x_1,θ_1), (x_2,θ_2), (x_3,θ_3), ..., be such that each point is an image of the preceding point under the map. Suppose first that the point $P = (x_1,\theta_1)$ is on the x axis, i.e., $\theta_1 = 0$. By Properties 1 and 2, the equalities $\theta_{n+1} = m/2$ and $2\theta_{n+2} - x_{n+2} = m$, respectively, are the conditions of fixed points of T^{2n} and T^{2n+1}. Here, m is an arbitrary integer. Suppose next that the point P is on the θ axis, i.e., $x_1 = 0$. By Properties 1 and 2, the equalities $x_{n+1} = m$ and $2x_{n+2} - h(\theta_{n+2} - x_{n+2}) = m$, respectively, are the conditions of fixed points of T^{2n} and T^{2n+1}. Here, an integer m is arbitrary.

Let us define functions G_q and H_q $(q \geqslant 1)$ by

$$\begin{aligned} G_{2n} &= 2\theta_{n+1}, & G_{2n+1} &= 2\theta_{n+2} - x_{n+2}, \\ H_{2n} &= x_{n+1}, & H_{2n+1} &= 2x_{n+2} - h(\theta_{n+2} - x_{n+2}). \end{aligned} \tag{8}$$

Then, the conditions for a fixed point of T^q on the x axis and on the θ axis, respectively, become $G_q = m$ and $H_q = m$, where m is an arbitrary integer. With the aid of Eq. (1), functions G_q and H_q can be expressed as

$$G_q = q x_1 + \sum_{i=1}^{[(q-1)/2]} (q - 2i) h(\theta_{i+1}),$$

$$H_q = \begin{cases} \displaystyle\sum_{i=1}^{q/2} h(\theta_i), & \text{for an even } q, \\ \displaystyle 2\sum_{i=1}^{(q-1)/2} h(\theta_i) + h(\theta_{(q+1)/2}), & \text{for an odd } q. \end{cases} \tag{9}$$

In the first equation of Eq. (9), θ_i are functions of x_1. In the second equation, $\theta_2, \theta_3, \ldots$, are functions of θ_1.

Functions G_q and H_q are also given by recursion formulas as

$$\begin{aligned} &G_1 = x, \quad G_2 = 2x, \\ &G_{2q+1} = 2G_{2q} - G_{2q-1} + h(G_{2q}/2), \\ &G_{2q+2} = 2G_{2q+1} - G_{2q}, \end{aligned} \tag{10}$$

and

$$\begin{aligned} &H_1 = 0, \quad H_2 = h(\theta), \\ &H_{2q+1} = 2H_{2q} + h\left(\theta + \sum_{i=1}^{q} H_{2i}\right), \\ &H_{2q+2} = H_{2q+1} - H_{2q}, \end{aligned} \tag{11}$$

where we omit suffixes of x and θ for brevity.

It is apparent that functions G_q and H_q are odd functions of their arguments. Consequently, if a point $(x,0)$ is a periodic point then the point $(-x,0)$ is a periodic point of the same period. The same is true for the θ axis.

Double reversibilities give us more information on periodic points.

Proposition: If the orbit of a point has points both on the x and θ axes then the point is a periodic point of period $4n-2$ for $n = 2, 3, \ldots$.

Proof: Let $P = (x,0)$ be a point on the x axis and suppose that $T^n P$ is on the θ axis, i.e., $P = I_1 P$ and $T^n P = J_1 T^n P$. Then, taking into account the relations $T^k I_1 = I_1 T^{-k}$ and $T^k J_1 = J_1 T^{-k}$ for $k = 1, 2, \ldots$, we obtain $T^{2n-1}P = J_1 T P = J_1(x,x) = (-x,0) = I_1 T^{2n-1} P$. Thus the point P is a periodic point of period $4n-2$. A similar argument applies to a point initially on the θ axis. Q.E.D.

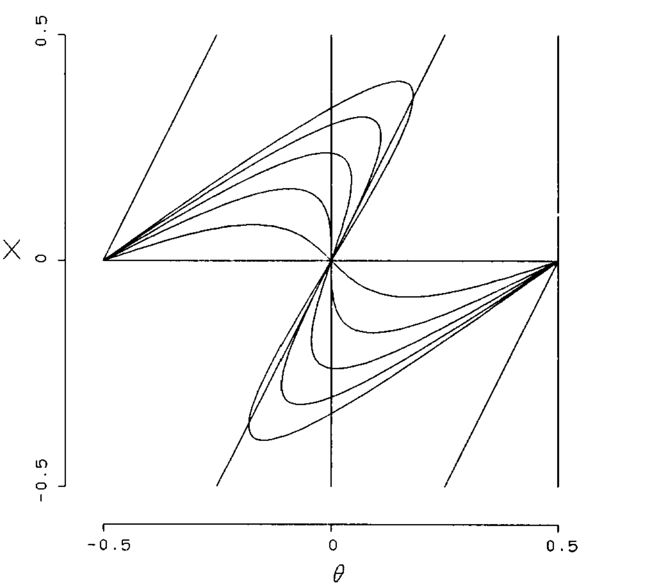


FIG. 1. The symmetry axes of the standard map. The dominant axes of the first reversibility are the x axis and $\theta = \frac{1}{2}$. The complementary axes are $\theta = (x-1)/2$, $\theta = x/2$, and $\theta = (x+1)/2$. The dominant axis of the second reversibility is the θ axis. The complementary axis is given by $2x = -(K/2\pi)\sin 2\pi(\theta - x)$. This axis is shown for $K = 1, 2, 3, 4$, and 5.

III. BIFURCATION OF PERIODIC POINTS FROM THE ORIGIN

Let us consider the problem of bifurcation of periodic points from the origin. According to Golubitsky and Schaeffer,[7] in any bifurcation problem $g(x,\lambda)$, the number of periodic points increases from one to three (when we increase λ) at a specific point (x_0,λ_0) if the conditions

$$g = g_x = g_{xx} = g_\lambda = 0, \quad g_{xxx} g_{\lambda x} < 0$$

are satisfied. In particular, if the inequalities

$$g_{xxx} > 0, \quad g_{\lambda x} < 0$$

are satisfied in addition, then the bifurcation problem is equivalent to the normal form $x^3 - \lambda x = 0$ for the pitchfork bifurcation.

Let us define functions Q_q and R_q for $q = 1, 2, ...,$ by

$$Q_q(K) = \left.\frac{\partial G_q}{\partial x}\right|_{x=0}, \text{ and } R_q(K) = \left.\frac{\partial H_q}{\partial \theta}\right|_{\theta=0}. \quad (12)$$

Then, zeros of these functions are the points in the parameter space of bifurcations of the fixed point of T^q (see Lemma 3 in the following). The functions Q_q are given recursively as

$$\begin{aligned} &Q_1 = 1, \quad Q_2 = 2, \\ &Q_{2q+1} = \tfrac{1}{2}(4-K)Q_{2q} - Q_{2q-1}, \quad (13)\\ &Q_{2q+2} = 2Q_{2q+1} - Q_{2q}, \quad q = 1, 2, \ldots . \end{aligned}$$

The bifurcation structures of the x and θ axes are the same. In fact, if we define functions $\widetilde{R}_q$ by

$$\widetilde{R}_{2q} = -2R_{2q}/K, \quad \widetilde{R}_{2q+1} = -R_{2q+1}/K,$$

then these functions satisfy the same recursion formulas as functions Q_q do. So we consider the functions Q_q in the following. These functions have interesting properties. Proofs of the following lemmas are given in the Appendix.

Lemma 1: The function Q_q $(1 \leqslant q)$ has $[(q-1)/2]$ simple zeros in $0 < K < 4$.

Let us denote these zeros by $K_{q,1}, K_{q,2}, \ldots, K_{q,[(q-1)/2]}$ in the ascending order.

Lemma 2: The number of different zeros of functions $Q_3, Q_4, \ldots, Q_q$ on the K axis $(0 < K < 4)$ is equal to the number of fractions of the Farey sequence $\mathscr{F}_q$ of order q of the interval $(0,\frac{1}{2})$. In addition, if $K_{m,n} \leqslant K_{m',n'}$ then $n/m \leqslant n'/m'$ and vice versa where $K_{m,n} = K_{m',n'}$ if and only if $n/m = n'/m'$.

Lemma 3: $G_q = G_{q;x} = G_{q;xx} = G_{q;K} = 0$, $G_{q;xxx} G_{q;Kx} < 0$ for $x = 0$ and $K = K_{q,s}$ $(1 \leqslant s \leqslant [(q-1)/2])$. Moreover, $G_{q;xxx} > 0$ and $G_{q;Kx} < 0$ for an even s, and $G_{q;xxx} < 0$ and $G_{q;Kx} > 0$ for an odd s.

Thus even zeros of the function Q_q correspond to pitchfork bifurcations in the sense of Golubitsky and Schaeffer.

Lemma 4: Zeros of the function Q_q densely fill the interval $0 \leqslant K \leqslant 4$ as q tends to infinity.

Let w be the rotation number around the fixed point. We obtain the relation $K = 2(1 - \cos 2\pi w)$ through the analysis of linear stability of the fixed point at the origin. Every time w takes a rational number, periodic points bifurcate from the origin. Let $K^*_{q,s}$ be the value of K when $w = s/q$. We confirmed numerically that $K_{q,s}$ and $K^*_{q,s}$ coincide. Thus every time when periodic points bifurcate from the origin, some of them are always borne on the x and θ axes. If q is odd, four different periodic points bifurcate along the coordinate axes. If $q = 4n$, two different periodic points bifurcate. If $q = 4n - 2$, only one periodic point bifurcates along the coordinate axes. Due to the oddness of the functions G_q and H_q, these bifurcated periodic points are symmetrically located along the respective axes.

IV. NUMERICAL INVESTIGATIONS

Let us numerically investigate the problem of the coexistence of symmetric periodic points of the x axis with the aid of Eq. (10).

Let us introduce the rotation number W along the θ axis. This is defined by

$$W = \lim_{q \to \infty} (\theta_{q+1} - \theta_1)/q.$$

Let $W_{q,m}$ be the rotation number of a periodic point satisfying $G_q = m$. Then, we have $W_{q,m} = m/q$ from Eq. (8). Let us introduce functions G^*_q by $G^*_q = G_q/q$. Then, conditions for periodic points become $G^*_q = m/q = W_{q,m}$. We have $G^*_q(0) = 0$ and $G^*_q(\frac{1}{2}) = \frac{1}{2}$ for all $q \geqslant 1$. The value of the function G^*_q takes every rational number between 0 and $\frac{1}{2}$ in the interval $0 \leqslant x \leqslant \frac{1}{2}$ as $q \to \infty$. It follows that the number of periodic points on the x axis is at least equal to the number of rational numbers between 0 and $\frac{1}{2}$ irrespective of the value of K. Figures 2(a) and 2(b) show the graph of the functions G_q and G^*_q ($q = 31$, 113, 197, 271, and 359) for $K = 0.5$.

Let G^*_∞ be the limit (or one of the limits) of the functions G^*_q as $q \to \infty$. Let us discuss briefly the relation between the possible dense distribution of periodic points and the form of the function G^*_∞. Periodic points are dense on the x axis if the function G^*_∞ is analytic, because a nonconstant analytic function is not constant for a finite nonzero interval and arguments for which the value of the function is rational are dense in the domain of the definition of the function. Thus periodic points are dense in the integrable case where $G^*_q = x$ for all q, and hence $G^*_\infty = x$.

The function G^*_∞ is not analytic as long as there is a stable periodic point on the x axis. In fact, rotation numbers of the bifurcated periodic points are equal to that of the mother periodic point. Therefore, the amplitude of the oscillations of the function G_q is less than unity in the invariant region. This means that the function G^*_∞ is constant in the interval corresponding to the invariant region. Thus the function G^*_∞ is not analytic for $K > 0$ because there are always stable periodic points. The step structure is seen in Figs. 2(b) and 2(c). The limit function G^*_∞ has a form of a devil's staircase. It is to be noted here that there is a possibility that the value of the function G^*_∞ is equal to an irrational number for a finite nonzero interval. In this case, there is no periodic point in the step.

The sizes of invariant regions of periodic points decrease as we increase K. They become infinitesimal in the limit of $K \to \infty$ (this is a consequence of the result of the preceding paper[2]). The function G^*_∞ becomes more and more oscillatory as K increases. It becomes a nowhere differentiable function in the limit since periodic points of period q for any $q \geqslant 3$ become dense. Figure 2(c) shows the graph of the function G^*_q ($q = 359$) for $K = 1.0$. It is seen that there are rapid

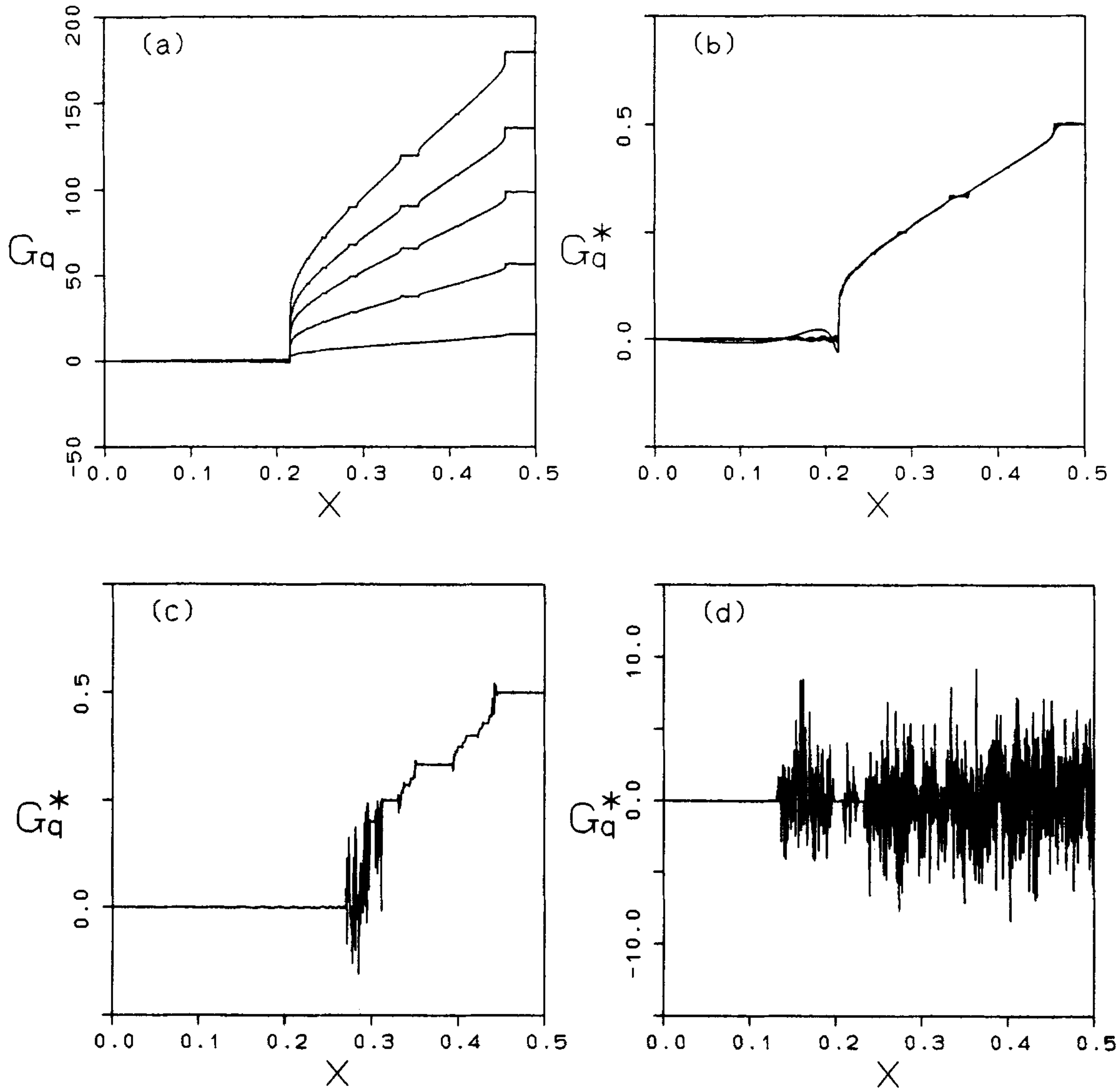


FIG. 2. (a) The functions G_q ($q = 31, 113, 197, 271$, and 359) for $K = 0.5$. (b) The normalized functions G_q^* for the same values of q and K as in (a). (c) The normalized function G_q^* ($q = 359$) for $K = 1.0$. (d) The normalized function G_q^* ($q = 359$) for $K = 4.0$.

oscillations with large amplitudes outside the invariant regions of stable periodic points. In these regions, periodic points are borne through tangent bifurcations and the number of periodic points increases rapidly. Figure 2(d) shows a similar graph for $K = 4.0$. Almost all parts of the x axis except the stable region around the origin are filled with rapid oscillations.

Now, the step structures of functions G_q^* and G_∞^* indicate that the bifurcation structures of the periodic points which exist down to the integrable case are similar to that of the fixed point at the origin. This suggests that if the periodic points are dense in the stable region of the fixed point at the origin, then they should be dense in the other stable region. So, as a representative case, we numerically calculate the position of periodic points in the invariant region of the fixed point in the following.

Let us numerically determine the position of the outermost KAM curve of the fixed point at the origin. Let x_q be the smallest positive value of x at which the absolute value of the function G_q becomes unity. Then, we have a simple property.

Proposition: $x_{q+1} < x_q$ for $q \geqslant 3$.

Proof: We shall prove the inequality by induction on q. For $q = 3$, one can easily confirm the inequality. We assume that the proposition is true for $q = 3, 4, \ldots, n-1$. We have

$$G_{n+1}(x_n) = 2G_n(x_n) - G_{n-1}(x_n) + h(G_n(x_n)/2)$$
$$= \pm 2 - G_{n-1}(x_n),$$

for an even n, and

$$G_{n+1}(x_n) = 2G_n(x_n) - G_{n-1}(x_n)$$
$$= \pm 2 - G_{n-1}(x_n),$$

for an odd n. In both the cases, the absolute value of the function G_{n+1} at x_n is greater than unity because the abso-

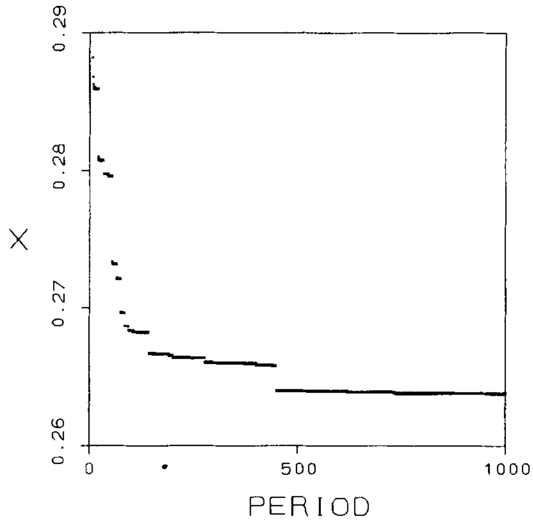


FIG. 3. Convergence of x_q against q ($7 \leqslant q \leqslant 1000$) for $K = 1$. The value x_q is the first x at which the absolute value of the function G_q becomes unity.

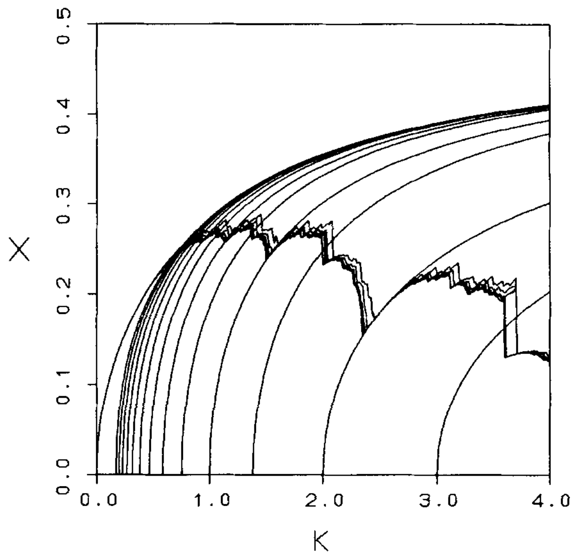


FIG. 4. Convergence of approximate outermost KAM curves. The values x_q are plotted against K for $q = 100, 200, 400, 600, 800$, and 1000. Characteristics of periodic orbits starting at $K_{q,1}$ ($q = 3, 4, ..., 15$) are plotted for reference.

lute value of G_{n-1} is smaller than unity by assumption. Q.E.D.

A periodic point at x_q has the rotation number $+1/q$ or $-1/q$ and is clearly outside the invariant region of the fixed point at the origin whose rotation number is zero. The sequence $\{x_q\}$ is a decreasing sequence, hence it converges. Let x_{bf} be the limit. Let $x_{\rm KAM}$ be the position of the outermost KAM curve on the positive x axis. Then, we have $x_{\rm KAM} \leqslant x_{bf}$. There is no proof that $x_{\rm KAM} = x_{bf}$. However, it is highly probable that this equality holds. We assume that this equality holds.

We determined x_{bf} approximately. Figure 3 shows the value of x_q against q ($7 \leqslant q \leqslant 1000$) for $K = 1$. Figure 4 shows the overall convergence of the approximate outermost KAM curves (they start at the origin). We plotted x_q against K for $q = 100, 200, 400, 600, 800$, and 1000. Characteristics of periodic points starting at $K_{q,1}$ ($q = 3, 4, ..., 15$) are plotted for reference. It is apparent that the convergence is along the characteristics. For this reason, our conclusion that follows that will not be so much affected by the badness of the convergence.

The result is shown in Fig. 5. In this figure, characteristics of periodic points starting at $K_{q,1}$ ($q = 3, 4, ..., 15$), those starting at $K_{q,2}$ ($q = 5, 7, 9, 11, 13, 15$), and those starting at $K_{q,3}$ ($q = 7, 8, 10, 11, 13, 14$), are plotted. Characteristics

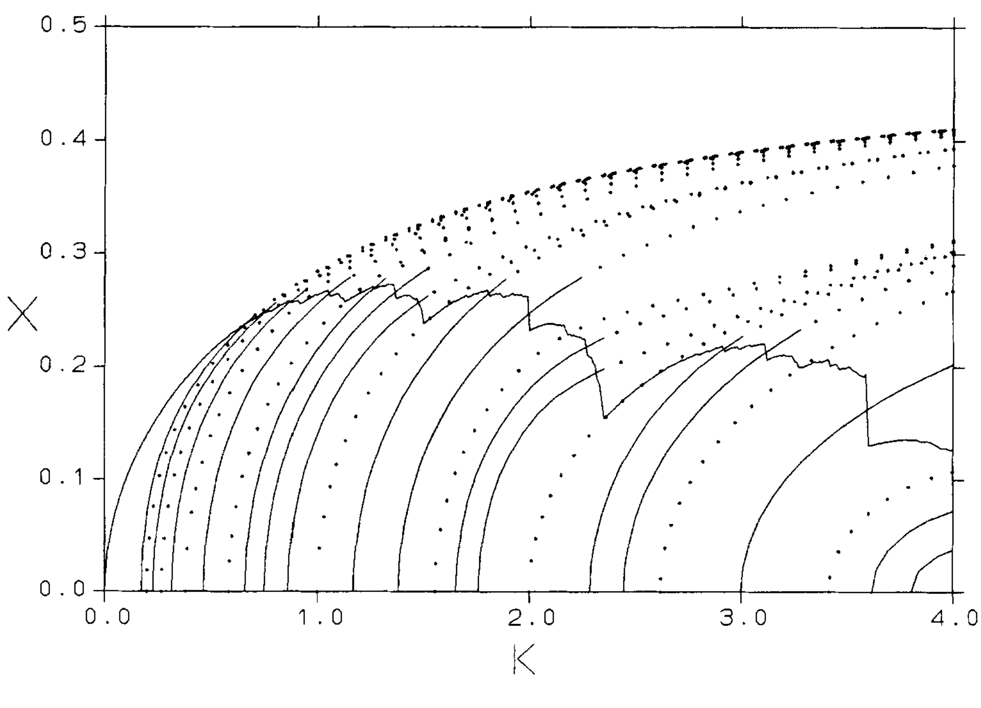


FIG. 5. Approximate outermost KAM curve. Also plotted are characteristics of periodic orbits starting at $K_{q,1}$ ($q = 3, 4, ..., 15$), those starting at $K_{q,2}$ ($q = 5, 7, 9, 11, 13, 15$), and those starting at $K_{q,3}$ ($q = 7, 8, 10, 11, 13, 14$). Characteristics are drawn with a solid line if the corresponding periodic points are stable, while they are drawn with a dotted line if unstable.

are drawn as solid lines if the corresponding periodic points are stable, while they are drawn as dotted lines if they are unstable. The curve starting at the origin is the approximate outermost KAM curve. We note sudden contractions and gradual expansions of the outermost KAM curve with the increase of K. These are closely related to the stability of periodic points. Each sudden contraction occurs when a stable daughter periodic point together with all its bifurcated points gets out of the invariant region. The length of contraction indicates the scale of the territory of the periodic point. Gradual expansions are related to general outward movements of periodic points.

In our calculation, periodic points of even periods are unstable from the beginning, while those of odd periods are stable even after they get out of the outermost KAM curve. A deformed devil's staircase evolution of the outermost KAM curve with the increasing perturbation parameter K indicates that this property should hold for higher periodic points. Here we conjecture that periodic points of even periods are unstable from the beginning and that periodic points of odd periods become unstable after they get out of the outermost KAM curve. Numerical calculations show that the residue[4] of a periodic point of a higher odd period remains positive and close to zero until its characteristic approaches the outermost KAM curve and that the residue quickly increases when the characteristic crosses the outermost KAM curve. This means that the stability of this kind of periodic point is almost neutral and invariant regions around them are small. Thus even a dense distribution of stable periodic points in a finite interval is possible if the residue tends to zero as the period tends to infinity.

A devil's staircase form of the evolution of the outermost KAM curve also suggests that daughter periodic points bifurcated from the origin do not disappear by inverse bifurcation as long as they are in the invariant region.

APPENDIX: PROOFS OF LEMMAS OF SEC. III

For the proofs of the Lemmas, let us give a series of properties of the functions Q_q.

Proposition 1: (a) Functions Q_q are polynomials of order $[(q-1)/2]$ in K. (b) Coefficients of the highest-order terms of functions Q_q are -1, -2, 1, and 2, respectively, for $q=4n-1$, $4n$, $4n+1$, $4n+2$. (c) $Q_{4q-3}>0$, $Q_{4q-2}>0$, $Q_{4q-1}<0$, and $Q_{4q}<0$ for $K\geqslant 4$ and $q=1,2,\ldots$. (d) $Q_q=q$ for $K=0$.

Proposition 2: (a) $Q_{2q+1}=Q_{2i+1}Q_{2(q-i)+1}-\frac{1}{2}(Q_{2i+1}+Q_{2i-1})Q_{2(q-i)}$ for $i=1,2,\ldots,q-1$, (b) $Q_{2q}=Q_{i+1}Q_{2q-i}-Q_iQ_{2q-i-1}$ for $i=1,2,\ldots,q$.

The proofs of the above two propositions are easy, so we omit them. Lemma 1 is a direct consequence of Proposition 1(a) and Proposition 3(a) that follows.

Proposition 3: (a) Functions Q_q, $q=1,2,\ldots$, have $[(q-1)/2]$ zeros in $0<K<4$. (b) There are the following relations among zeros $K_{q,1}, K_{q,2},\ldots,K_{q,[(q-1)/2]}$ of functions Q_q for $q=3,4,\ldots$: $K_{q+2,i}<K_{q+1,i}<K_{q,i}<K_{q+2,i+1}$ for $i=1,2,\ldots,[(q-1)/2]$ and for an odd q. If q is even, we have in addition $K_{q+2,[(q+1)/2]}<K_{q+1,[(q+1)/2]}$.

Proof: We shall prove (a) and (b) at the same time. Let us put $q=2k-1$, $2k$, $2k+1$, $2k+2$. The proof proceeds by induction on the number k. It is easily confirmed that our assertion is true for $k=2$. We assume that the assertion is true for $k=n$, and shall show that the assertion is also true for $k=n+1$.

From Proposition 1(a) and the induction hypothesis, zeros of the functions Q_{2n+1} and Q_{2n+2} are all simple. The proofs in the following are almost the same for even and odd n, so we assume n as even. Then, we obtain the following estimates of the function Q_{2n+3}:

$$
\begin{aligned}
&Q_{2n+3}(0)=2n+3>0,\\
&Q_{2n+3}(K_{2n+2,1})=-Q_{2n+1}(K_{2n+2,1})<0,\\
&Q_{2n+3}(K_{2n+1,1})=\tfrac{1}{2}(4-K_{2n+1,1})\\
&\qquad\qquad\times Q_{2n+2}(K_{2n+1,1})<0,\\
&Q_{2n+3}(K_{2n+2,2})=-Q_{2n+1}(K_{2n+2,2})>0,\\
&\vdots\\
&Q_{2n+3}(K_{2n+1,n-1})=\tfrac{1}{2}(4-K_{2n+1,n-1})\\
&\qquad\qquad\times Q_{2n+2}(K_{2n+1,n-1})<0,\\
&Q_{2n+3}(K_{2n+2,n})=-Q_{2n+1}(K_{2n+2,n})>0,\\
&Q_{2n+3}(K_{2n+1,n})=\tfrac{1}{2}(4-K_{2n+1,n})\\
&\qquad\qquad\times Q_{2n+2}(K_{2n+1,n})>0,\\
&Q_{2n+3}(4)<0.
\end{aligned}
$$

Therefore, the function Q_{2n+3} has one simple zero in each of the $n+1$ intervals $(0,K_{2n+2,1})$, $(K_{2n+1,1},K_{2n+2,2})$, ..., $(K_{2n+1,n-1},K_{2n+2,n})$, $(K_{2n+1,n},4)$, since the function is of order $n+1$ in K by Proposition 1(a). If we denote these zeros in the ascending order by

$$K_{2n+3,1}, K_{2n+3,2}, \ldots, K_{2n+3,n+1},$$

then we have the following inequalities:

$$K_{2n+3,i}<K_{2n+2,i}<K_{2n+1,i}<K_{2n+3,i+1},\quad \text{for } i=1,2,\ldots,n.$$

Similarly we obtain the following estimates of the function Q_{2n+4}:

$$
\begin{aligned}
&Q_{2n+4}(0)=2n+4>0,\\
&Q_{2n+4}(K_{2n+3,1})=-Q_{2n+2}(K_{2n+3,1})<0,\\
&Q_{2n+4}(K_{2n+2,1})=2Q_{2n+3}(K_{2n+2,1})<0,\\
&Q_{2n+4}(K_{2n+3,2})=-Q_{2n+2}(K_{2n+3,2})>0,\\
&\vdots\\
&Q_{2n+4}(K_{2n+2,n-1})=2Q_{2n+3}(K_{2n+2,n-1})<0,\\
&Q_{2n+4}(K_{2n+3,n})=-Q_{2n+2}(K_{2n+3,n})>0,\\
&Q_{2n+4}(K_{2n+2,n})=2Q_{2n+3}(K_{2n+2,n})>0,\\
&Q_{2n+4}(K_{2n+3,n+1})=-Q_{2n+2}(K_{2n+3,n+1})<0.
\end{aligned}
$$

By a similar reasoning as above, the function Q_{2n+4} has one simple zero in each of the $n+1$ intervals $(0,K_{2n+3,1})$, $(K_{2n+2,1},K_{2n+3,2})$, ..., $(K_{2n+2,n-1},K_{2n+3,n})$, $(K_{2n+2,n},K_{2n+3,n+1})$. If we denote these zeros in the ascending order as

$$K_{2n+4,1}, K_{2n+4,2}, \ldots, K_{2n+4,n+1},$$

then we have the following inequalities:

$$K_{2n+4,i} < K_{2n+3,i} < K_{2n+2,i} < K_{2n+4,i+1}, \quad \text{for } i = 1,2,\ldots,n,$$

and

$$K_{2n+4,n+1} < K_{2n+3,n+1}.$$

Q.E.D.

Proposition 4: Take a positive integer $m = rn$, where $r \geqslant 3$ is a prime number and $n \geqslant 1$ an integer. Then the function Q_m has Q_r as a factor.

We omit the proof.

Let us introduce notations for convenience. Let $\mathscr{F}_q^*$ be the set of all the fractions constructed by two suffixes of the zeros of the functions Q_n for $3 \leqslant n \leqslant q$. If we identify elements of $\mathscr{F}_q^*$ that have the same value, then we obtain the Farey sequence $\mathscr{F}_q$ of order q in the interval $(0,\frac{1}{2})$. Let $\mathscr{K}_q^*$ be the set of all zeros of the functions Q_n for $3 \leqslant n \leqslant q$.

Proof of Lemma 2: We shall prove that for any n/m $n'/m' \in \mathscr{F}_q^*$, we have

$$n/m = n'/m' \Leftrightarrow K_{m,n} = K_{m',n'},$$

$$n/m < n'm' \quad \Leftrightarrow K_{m,n} < K_{m',n'}.$$

The proof proceeds by induction on q. It is easily confirmed that our assertion is true for $q = 3$, 4, 5, and 6. We assume that Lemma 2 is true for $q = k$ and shall show that it is true for $q = k+1$. In order to do this, it is enough to show that the required relations are true among any fraction n/m of the set $\mathscr{F}_{k+1}^* - \mathscr{F}_k^*$ and any fraction n'/m' of the set $\mathscr{F}_k^*$.

Let us first consider two extreme cases. Let $n/m = 1/(k+1)$. We have $n/m < n'/m'$ for any $n'/m' \in \mathscr{F}_k^*$. On the other hand, by Proposition 3, the zero $K_{k+1,1}$ is the smallest element in $\mathscr{K}_{k+1}^*$. Therefore, Lemma 2 is true for this n/m. Next, let $n/m = [k/2]/(k+1)$ when $k+1$ is odd. We have $n'/m' < n/m$ for any $n'/m' \in \mathscr{F}_k^*$. On the other hand, by Proposition 3, the zero $K_{k+1,[k/2]}$ is the largest element in $\mathscr{K}_{k+1}^*$. Therefore, Lemma 2 is true for this n/m.

In general, let $n/m = i/(k+1)$ with $1 < i < [k/2]$ if $k+1$ is odd, and with $1 < i \leqslant [k/2]$ if $k+1$ is even. Let

$$i/(k+1) = s_i/r_i$$

where the fraction s_i/r_i is irreducible. Let us take three successive fractions n_i'/m_i', s_i/r_i, and n_i''/m_i'' of $\mathscr{F}_{k+1}$ containing s_i/r_i at the middle. Evidently two fractions n_i'/m_i', and n_i''/m_i'', are the elements of $\mathscr{F}_k$.

We shall show in the following that the function Q_{k+1} has one zero in each interval $(K_{m'_i,n'_i}, K_{m''_i,n''_i})$. It is evident from the definition of the intervals and relations among zeros of the function Q_{k+1} that this zero is equal to $K_{k+1,i}$. This shows that the proposition is true for $q = k+1$. In the following, we omit the suffix i of r_i, s_i, etc., for brevity.

Consider the case when the fraction $i/(k+1)$ is irreducible. Then, we have

$$\frac{i-1}{k-1} \leqslant \frac{n'}{m'} < \frac{i}{k+1} = \frac{s}{r} < \frac{n''}{m''} \leqslant \frac{i}{k}.$$

By the hypothesis of induction, Propositions 3 and 4, we have $K_{k-1,i-1} < K_{r,s} < K_{k,i}$, $K_{k-1,i-1} < K_{k+1,i} < K_{k,i}$, and $Q_{k+1}(K_{r,s}) = 0$, respectively. Combining these results, we obtain $K_{k+1,i} = K_{r,s}$.

Consider the case when the fraction $i/(k+1)$ is reducible and $k+1$ is even. The fractions n'/m' and n''/m'' are irreducible by definition. Hence both the integers m',m'' are odd, while one of the integers n' and n'' is odd and the other is even. From Proposition 2 and the property of the Farey sequence, we obtain

$$\begin{aligned} Q_{k+1}(K_{m',n'})Q_{k+1}(K_{m'',n''}) &= -Q_{m'+1}(K_{m',n'})Q_{m''}(K_{m',n'}) \\ &\quad \times Q_{m'}(K_{m'',n''})Q_{m''-1}(K_{m'',n''}). \end{aligned}$$

In general, the function $Q_q(K)$ intersects the K axis at $K_{q,r}$ from below with increasing K if the second suffix r of $K_{q,r}$ is even, while it intersects the K axis from above if r is odd. Then, we have $Q_{m'+1}(K_{m',n'}) > 0$, $Q_{m''}(K_{m',n'}) > 0$, $Q_{m'}(K_{m'',n''}) > 0$, and $Q_{m''-1}(K_{m'',n''}) > 0$ if n' is even and n'' is odd, while $Q_{m'+1}(K_{m',n'}) < 0$, $Q_{m''}(K_{m',n'}) < 0$, $Q_{m'}(K_{m'',n''}) < 0$, and $Q_{m''-1}(K_{m'',n''}) < 0$ if n' is odd and n'' is even. In any case, we have $Q_{k+1}(K_{m',n'})Q_{k+1}(K_{m'',n''}) < 0$, which means that the function Q_{k+1} has a zero in the interval $(K_{m',n'}, K_{m'',n''})$.

Consider the case when the fraction $i/(k+1)$ is reducible and $k+1$ is odd. Then, there are four cases to be considered: $(m', m'', n', n'') =$ (odd, even, odd, odd), (odd, even, even, odd), (even, odd, odd, odd), and (even, odd, odd, even). In any of these cases, we obtain $Q_{k+1}(K_{m',n'})$ $Q_{k+1}(K_{m'',n''}) < 0$. Q.E.D.

For the proof of Lemma 3, let us introduce notations. Let

$$T_q = \left.\frac{\partial^2 G_q}{\partial K\, \partial x}\right|_{x=0}, \quad U_q = \left.\frac{\partial^3 G_q}{\partial x^3}\right|_{x=0}.$$

Then, these functions satisfy the following recursion formulas:

$$\begin{aligned} &T_1 = 0, \quad T_2 = 0, \\ &T_{2q+1} = (4-K)T_{2q}/2 - T_{2q-1} - Q_{2q}/2, \\ &T_{2q+2} = 2T_{2q+1} - T_{2q}, \quad q = 1,2,\ldots, \\ &U_1 = 0, \quad U_2 = 0, \\ &U_{2q+1} = (4-K)U_{2q}/2 - U_{2q-1} - (\pi^2 K/2)Q_{2q}^3, \\ &U_{2q+2} = 2U_{2q+1} - U_{2q}, \quad q = 1,2,\ldots. \end{aligned}$$

Proposition 5:

$$T_q = -\left(\frac{1}{2}\right)\sum_{i=1}^{[(q-1)/2]} Q_{q-2i}Q_{2i},$$

$$U_q = \left(\frac{\pi^2 K}{2}\right)\sum_{i=1}^{[(q-1)/2]} Q_{q-2i}Q_{2i}^3, \text{ for } q = 3,4,\ldots.$$

We can readily prove the above proposition by induction on the suffix q, so we omit it.

Proposition 6: The inequality $U_q T_q < 0$ holds for each $K_{q,s}$ $(1 \leqslant s \leqslant [(q-1)/2])$. Moreover, we have $U_q > 0$, $T_q < 0$ if s is even, and $U_q < 0$, $T_q > 0$ if s is odd.

Proof: It is enough to show that the following inequalities hold for each term in the expression of the function T_q:

$$Q_{q-2i}Q_{2i} \geqslant 0 \text{ at } K = K_{q,s} \quad \text{for an odd } s,$$

$$Q_{q-2i}Q_{2i} \leqslant 0 \text{ at } K = K_{q,s} \quad \text{for an even } s,$$

where the strict inequality holds for at least one i.

Let us consider the case when the integers $q - 2i$ and $2i$

are mutually prime. Then, the function $Q_{q-2i}Q_{2i}$ has $[(q-1)/2]-1$ simple zeros. Let us denote these zeros in the ascending order by $K_1, K_2, \ldots, K_{[(q-1)/2]-1}$. From Lemma 2, we see that the inequalities $K_{q,1} < K_1 < K_{q,2} < \ldots < K_{[(q-1)/2]-1} < K_{q,[(q-1)/2]}$ hold. These inequalities with $Q_{q-2i}(0)Q_{2i}(0)$ ensure the required inequalities.

Next, consider the case when the integers $q-2i$ and $2i$ are mutually not prime. Let k be the greatest common factor of the integers $q-2i$ and $2i$. Then, the function $Q_{q-2i}Q_{2i}$ has $[(k-1)/2]$ double zeros $K_{k,1}, K_{k,2}, \ldots, K_{k,[(k-1)/2]}$, and the remaining $[(q-1)/2]-[(k-1)/2]-1$ simple zeros. The double zeros of the function $Q_{q-2i}Q_{2i}$ are also the zeros of the function Q_q, since the integer k is a factor of q. At each of these zeros, the function $Q_{q-2i}Q_{2i}$ is tangent to the K axis. If we consider that a point $K_{k,i}$ belongs to both intervals of the form $[K_{q,l}, K_{q,l+1}]$ which contain the point $K_{k,i}$ as a terminal, then each zero of the function $Q_{q-2i}Q_{2i}$ are contained in each interval $[K_{q,l}, K_{q,l+1}]$, as is easily confirmed. Q.E.D.

Proof of Lemma 3: From the oddness of the function G_q the properties $G_{q;xx} = G_{q;K} = 0$ for $x=0$ and $K = K_{q,s}$ $(1 \leqslant s \leqslant [(q-1)/2])$ directly follow.[7] Clearly, $G_q = G_{q;x} = 0$ for $x=0$ and $K = K_{q,s}$ $(1 \leqslant s \leqslant [(q-1)/2])$. Q.E.D.

Proposition 7: (a) The intersection points of the functions Q_{2n+1} and Q_{2n} are odd zeros of the function $Q_{4(n+2)}$. (b) The intersection points of the functions Q_{2n} and Q_{2n-1} are odd zeros of the function $Q_{4(n-2)}$. (c) The intersection points of the functions Q_{2m+1} and Q_{2n+1} $(m>n)$ are either odd zeros of the function $Q_{2(m+n+1)}$ or zeros of the function Q_{m-n}. (d) The intersection points of the functions Q_{2m} and Q_{2n} $(m>n)$ ar either odd zeros of the functioin $Q_{2(m+n)}$ or zeros of the function Q_{m-n}.

Proof: Let us define functions Q_q for $q \leqslant 0$ by $Q_0 = 0$ and $Q_{-q} = -Q_q$ $(q = 1, 2, \ldots)$. Then, the functions Q_q satisfy relations in Proposition 2 for all integers q. We use these extended relations. The proofs of the issues (a) and (b) are easy. We shall prove only the issue (c) for brevity.

By Proposition 2, we have $Q_{2(m+n+1)} = Q_{m+n+1}(Q_{m+n+2} - Q_{m+n})$. Thus a zero of the function $Q_{2(m+n+1)}$ is a zero of either the function Q_{m+n+1} or $Q_{m+n+2} - Q_{m+n}$. On the other hand, we can easily prove equalities $Q_{m+n+2+j} = Q_{m+n-j}$ $(j = 0, 1, \ldots)$ by induction on j using the above extended relations if the relation $Q_{m+n+2} = Q_{m+n}$ holds. Therefore, odd zeros of the function $Q_{2(m+n+1)}$ are zeros of the function $Q_{2m+1} - Q_{2n+1}$.

By Proposition 2, we have $Q_{2(m-n)} = Q_{m-n}(Q_{m-n+1} - Q_{m-n-1})$. Thus a zero of the function $Q_{2(m-n)}$ is a zero of either the function Q_{m-n} or $Q_{m-n+1} - Q_{m-n-1}$. On the other hand, we can easily prove equalities $Q_{m-n+j} = -Q_{m-n-j}$ $(j = 1, 2, \ldots)$ by induction on j using the above extended relations if the relation $Q_{m-n} = 0$ holds. Therefore, zeros of the function Q_{m-n} are zeros of the function $Q_{2m+1} - Q_{2n+1}$.

The number of relevant zeros of the functions $Q_{2(m+n+1)}$ and Q_{m-n} are $m-1$, while the number of zeros of the function $Q_{2m+1} - Q_{2n+1}$ is at most $m-1$, except for the trivial zero at $K=4$. Therefore, the proof is completed if the number of relevant zeros of the functions $Q_{2(m+n+1)}$ and Q_{m-n} are all distinct. In the case when the zeros of the functions $Q_{2(m+n)}$ and Q_{m-n} coincide, the function $Q_{2m+1} - Q_{2n+1}$ has a double zero at this point, i.e., the derivative $T_{2m+1} - T_{2n+1}$ vanishes, as is easily confirmed. Q.E.D.

Proposition 8: We have $K_{q,1} \to 0$ and $K_{2q+1,q} \to 4$ as $q \to \infty$.

Proof: The sequence $\{K_{q,1}\}$ converges. Let K_0 be the limit. We observe that the relations $Q_{2q} > q+1$ $(q = 2, 3, \ldots)$ hold for $K \leqslant K_0$. Since the function Q_{2q} is a polynomial in K with $K_{2q,1}$ as a first zero and $Q_{2q}(0) > 0$, the function itself and the absolute value of its derivative decrease in the interval $[0, K_{2q,1}]$. This means $K_{2q,1} > 2K_0$, which implies $K_0 = 0$.

The second relation can be shown in a similar manner. Q.E.D.

Proposition 9: Any zero $K_{q,s}$ of the function Q_q are the limit of any sequence $\{K_{m,n}\}$ suffixed by a sequence of fractions $\{n/m\} \subset \mathscr{F}_q^*$ that converges to the fraction s/q.

Proof: Take monotone sequences $\{K_{nq,ns-1}\}$ and $\{K_{nq,ns+1}\}$ and let the limits be K_0 and K_1. Evidently we have $K_0 \leqslant K_{q,s} \leqslant K_1$. We shall show that the relations $K_0 = K_{q,s} = K_1$ hold, which prove the Proposition.

Let us give several properties of the functions Q_q without proof. These can be verified easily with the aid of Proposition 7.

(1) $0 < Q_{2q} < Q_{4q} < \ldots < Q_{2nq} < \ldots$ for $K_{q,s} < K \leqslant K_1$,
$0 > Q_{2q} > Q_{4q} > \ldots > Q_{2nq} > \ldots$ for $K_0 \leqslant K < K_{q,s}$;

(2) $Q_{2nq+2} > 0, Q_{2nq+1} > 0, Q_{2nq-1} < 0, Q_{2nq-2} < 0$,
and $Q_{2nq+2} > Q_{2nq+1} > Q_{2nq} > Q_{2nq-1} > Q_{2nq-2}$
for $K_0 \leqslant K \leqslant K_1$;

(3) $Q_{2q+2} > Q_{2nq} > Q_{2q-2}$ for $K_0 \leqslant K \leqslant K_1$;

(4) $Q_{2nq+1} = -Q_{2nq-1} = 1$
and $Q_{2nq+2} = -Q_{2nq-2} = 2$ at $K = K_{q,s}$.

By issue (3), the sequence $\{Q_{2nq}\}$ converges in the interval $[K_0, K_1]$ though the convergence is not necessarily uniform. We obtain

$$Q_{2nq+1} - Q_{2nq-1} = Q_{4nq}/Q_{2nq} \to 1 \text{ as } n \to \infty,$$

in the interval except at $K = K_{q,s}$. We have $Q_{2nq-1} \to 0$ and hence $Q_{2nq+1} \to 1$ as $n \to \infty$ if we approach the point K_1 from inside the interval $[K_0, K_1]$. Therefore, we obtain the estimate $Q_{2nq+1}(K_1) > 1 - \delta$ for sufficiently large n and any $\delta > 0$. Moreover, since the function Q_{2nq+1} is a polynomial of order nq and has nq simple zeros, we have the estimates

$$Q_{2nq+1}(K) > \min\{Q_{2nq+1}(K_{q,s}), Q_{2nq+1}(K_1)\} > 1 - \delta$$

in the interval $[K_{q,s}, K_1]$ for a sufficiently large n. For this n, we have $Q_{2nq-1} > -\delta$ in the interval $(K_{q,s}, \ldots, K_1]$. Then, the function Q_{2nq-1} has two points of inflection in the interval $(K_{q,s}, K_{2nq-1,2ns})$, one in the neighborhood of $K_{q,s}$, and the other in the neighborhood of $K_{2nq-1,2ns}$, since the function is of finite amplitude as seen from the proof of Proposition 6. This contradicts the fact that the function Q_{2nq-1} is a polynomial which has as much zeros as its order. Therefore, we have $K_1 = K_{q,s}$. The equality $K_0 = K_{q,s}$ is obtained similarly. Q.E.D.

Proof of Lemma 4: Take any point K^* $(0 < K^* < 4)$. By Proposition 8, there exist integers q_0 and i_0 for which $K_{q_0,i_0} \leqslant K^* \leqslant K_{q_0,i_0+1}$. Let us fix such q_0 and i_0. By Lemma 2, we have

$$K_{2q_0,2i_0} = K_{q_0,i_0} < K_{2q_0,2i_0+1} < K_{q_0,i_0+1} = K_{2q_0,2i_0+2}.$$

Let $i_1 = 2i_0 + 1$ if $K_{2q_0,2i_0+1} < K^*$, while $i_1 = 2i_0$ if $K_{2q_0,2i_0+1} > K^*$. Then, we obtain the inequality $K_{2q_0,i_1} \leqslant K^* \leqslant K_{2q_0,i_1+1}$. In a similar manner, let us define integers i_n, $n = 2, 3, \ldots$, so as to satisfy the inequalities $K_{2^n q_0,i_n} \leqslant K^* \leqslant K_{2^n q_0,i_n+1}$.

The sequences $\{K_{2^n q_0,i_n}\}$ and $\{K_{2^n q_0,i_n+1}\}$ are monotone, and hence converge. Similarly, the sequences $\{i_n/2^n q_0\}$ and $\{(i_n+1)/2^n q_0\}$ converge to a limit. By Proposition 9, we obtain $|K_{2^n q_0,i_n} - K_{2^n q_0,i_n+1}| \to 0$ as $|i_n/2^n q_0 - (i_n+1)/2^n q_0| \to 0$. Q.E.D.

[1]B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[2]K. Tanikawa and Y. Yamaguchi, J. Math. Phys. **28**, 921 (1987).
[3]J. M. Greene, R. S. Mackay, F. Vivaldi, and M. J. Feigenbaum, Physica (Utrecht) D **3**, 468 (1981).
[4]J. M. Greene, J. Math. Phys. **20**, 1183 (1979).
[5]S. J. Shenker and L. P. Kadanoff, J. Stat. Phys. **27**, 631 (1982).
[6]R. DeVogelaere, *Contributions to the Theory of Nonlinear Oscillations*, edited by S. Lefschetz (Princeton U. P., Princeton, 1958), Vol. IV, p. 53.
[7]M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory* (Springer, New York, 1985).

Scalar concomitants of a metric and a curvature form. II

R. J. Noriega
Departmento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
C. G. Schifini
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina, and CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

(Received 24 May 1988; accepted for publication 19 October 1988)

The general form of a Lagrangian concomitant of a metric and a curvature form is found, improving substantially a previous result.

I. INTRODUCTION

In gauge field theories, Lagrangians of the form[1]

$$L = L(g_{ij};F^{\alpha}_{ij}) \tag{1}$$

are considered to obtain, through the use of variational principles, the field equations for a gauge theory. To establish the uniqueness of such equations, it is important to know the general form of Lagrangians of the type (1). Since $L/\sqrt{g}$ is a scalar, it is enough to find all scalar concomitants of such objects. In a previous paper[2] we have proved that it must be

$$L(g_{ij};F^{\alpha}_{ij}) = \sqrt{g}\, f(\phi^{\alpha\beta};\psi^{\alpha\beta}\phi^{\alpha\beta\gamma};\psi^{\alpha\beta\gamma}), \tag{2}$$

where

$$\phi^{\alpha\beta} = F^{\alpha ij}F^{\beta}{}_{ij}, \quad \psi^{\alpha\beta} = F^{\alpha ij}\, {}^*F^{\beta}{}_{ij}, \tag{3}$$

$$\phi^{\alpha\beta\gamma} = F^{\alpha i}{}_{j}F^{\beta j}{}_{h}F^{\gamma h}{}_{i}, \quad \psi^{\alpha\beta\gamma} = F^{\alpha i}{}_{j}F^{\beta j}{}_{h}\, {}^*F^{\gamma h}{}_{i}, \tag{4}$$

and where

$${}^*F^{\alpha ij} = \eta^{ijhk}F^{\alpha}{}_{hk} = (1/2\sqrt{g})\varepsilon^{ijhk}F^{\alpha}{}_{hk}.$$

Now, expression (2) is hard to handle because of the presence of the terms (4). In this paper we will prove that they can be removed from (2), obtaining an expression very similar to the one obtained in the electromagnetic case.[3] Since the latter was essential for solving the equivariant inverse problem, and so for proving the uniqueness of the Maxwell equations,[4] the result in the present paper could be useful for proving the uniqueness of the Yang–Mills equations, which will be the subject of a forthcoming paper.

II. GENERAL FORM OF THE LAGRANGIAN

Dividing or multiplying by $\sqrt{g}$ we turn densities to scalars and vice versa. So we can consider L given by (1) as a scalar. In this case, the invariance identities[5] are

$$L^{bs}g_{as} + L^{bs}_{\alpha}F^{\alpha}_{as} = 0, \tag{5}$$

where $L^{bs} = \partial L/\partial g_{bs}$, $L^{bs}_{\alpha} = \partial L/\partial F^{\alpha}_{bs}$. Since g_{as} is nonsingular, we deduce

$$L^{bs} = -L^{bt}_{\alpha}F^{\alpha}_{at}g^{as}. \tag{6}$$

Let us suppose for the moment that the Lie group G is three dimensional. Then, as we proved in Ref. 2, the skew-symmetric gauge tensor L^{bt}_{α} can be written as

$$L^{bt}_{\alpha} = a_{\alpha\beta}F^{\beta bt} + b_{\alpha\beta}\, {}^*F^{\beta bt}, \tag{7}$$

where $a_{\alpha\beta}$ and $b_{\alpha\beta}$ are gauge invariant scalar concomitants of g_{ij} and F^{α}_{ij}. In the next step we will prove that they are symmetric in their greek indices, i.e., $a_{\alpha\beta} = a_{\beta\alpha}, b_{\alpha\beta} = b_{\beta\alpha}$. To achieve this, we replace (7) in (6) to obtain

$$L^{bs} = -a_{\alpha\beta}F^{\beta bt}F^{\alpha s}{}_{t} - b_{\alpha\beta}\, {}^*F^{\beta bt}F^{\alpha s}{}_{t}. \tag{8}$$

Taking account of the symmetry of L^{bs} in b, s, we have from (8),

$$a'_{\alpha\beta}F^{\beta bt}F^{\alpha s}{}_{t} + b_{\alpha\beta}({}^*F^{\beta bt}F^{\alpha s}{}_{t} - {}^*F^{\beta st}F^{\alpha b}{}_{t}) = 0, \tag{9}$$

where $a'_{\alpha\beta} = a_{\alpha\beta} - a_{\alpha\beta}$. Multiplying (9) by $-F^{\gamma}_{bs}$ we have

$$a'_{\alpha\beta}\phi^{\alpha\beta\gamma} + 2b_{\alpha\beta}\psi^{\gamma\alpha\beta} = 0, \tag{10}$$

where $\phi^{\alpha\beta\gamma}$ and $\psi^{\gamma\alpha\beta}$ are given by (4). It can be proved easily that $\psi^{\alpha\beta\gamma}$ is skew symmetric in all of its indices, so that (10) can be rewritten as

$$a'_{\alpha\beta}\phi^{\alpha\beta\gamma} + b'_{\alpha\beta}\psi^{\alpha\beta\gamma} = 0, \tag{11}$$

where $b'_{\alpha\beta} = b_{\alpha\beta} - b_{\beta\alpha}$.

Similarly, by multiplying (9) by $-{}^*F^{\gamma}{}_{bs}$ we obtain

$$a'_{\alpha\beta}\psi^{\alpha\beta\gamma} - b'_{\alpha\beta}\phi^{\alpha\beta\gamma} = 0. \tag{12}$$

Now, since $\dim G = 3$, and from the skew symmetry of $\phi^{\alpha\beta\gamma}$ and $\psi^{\alpha\beta\gamma}$, we have

$$\phi^{\alpha\beta\gamma} = \lambda\varepsilon^{\alpha\beta\gamma}, \quad \psi^{\alpha\beta\gamma} = \mu\varepsilon^{\alpha\beta\gamma}, \tag{13}$$

where λ and μ are scalar concomitants of g_{ij} and F^{α}_{ij}. Replacing (13) in (11) and (12) gives us

$$\varepsilon^{\alpha\beta\gamma}[a'_{\alpha\beta}\lambda + b'_{\alpha\beta}\mu] = 0, \tag{14}$$

and

$$\varepsilon^{\alpha\beta\gamma}[a'_{\alpha\beta}\mu - b'_{\alpha\beta}\lambda] = 0. \tag{15}$$

Taking $\gamma = 3$ in (14) and (15) it follows that

$$\lambda a'_{12} + \mu b'_{12} = 0, \tag{16}$$

and

$$\mu a'_{12} - \lambda b'_{12} = 0. \tag{17}$$

Since the determinant of the system given by (16) and (17) is $-(\lambda^2 + \mu^2) \neq 0$, we have

$$a'_{12} = b'_{12} = 0. \tag{18}$$

Similarly, taking $\gamma = 2$ and $\gamma = 1$ in (14) and (15) it follows that

$$a'_{31} = b'_{31} = a'_{23} = b'_{23} = 0. \tag{19}$$

From (18) and (19) we obtain the claimed symmetry of $a_{\alpha\beta}$ and $b_{\alpha\beta}$.

Now, we are in position to prove that L depends only on $\phi^{\alpha\beta}$ and $\psi^{\alpha\beta}$. To obtain this let us suppose

$$\phi^{\alpha\beta} = \text{const}, \quad \psi^{\alpha\beta} = \text{const}. \tag{20}$$

Then

$$0 = d\phi^{\alpha\beta} = \frac{\partial\phi^{\alpha\beta}}{\partial g_{ij}} dg_{ij} + \frac{\partial\phi^{\alpha\beta}}{\partial F^{\gamma}_{ij}} dF^{\gamma}_{ij}$$
$$= -[F^{\alpha i}{}_k F^{\beta kj} + F^{\alpha j}{}_k F^{\beta ki}] dg_{ij}$$
$$+ F^{\beta ji} dF^{\alpha}_{ij} + F^{\alpha ji} dF^{\beta}_{ij}. \tag{21}$$

Similarly

$$0 = d\psi^{\alpha\beta} = -\tfrac{1}{2} g^{ij}\psi^{\alpha\beta} dg_{ij} + {*F}^{\beta ji} dF^{\alpha}_{ij} + {*F}^{\alpha ji} dF^{\beta}_{ij}. \tag{22}$$

Then

$$dL = L^{bs} dg_{bs} + L^{bt}_{\alpha} dF^{\alpha}_{bt} = L^{bt}_{\alpha}(-F^{\alpha}_{at} g^{as} dg_{bs} + dF^{\alpha}_{bt})$$
$$= (a_{\alpha\beta} F^{\beta bt} + b_{\alpha\beta} {*F}^{\beta bt})(-F^{\alpha}_{at} g^{as} dg_{bs} + dF^{\alpha}_{bt})$$
$$= -\tfrac{1}{2} a_{\alpha\beta}(F^{\beta bt} F^{\alpha s}{}_t + F^{\alpha bt} F^{\beta s}{}_t) dg_{bs}$$
$$+ \tfrac{1}{2} a_{\alpha\beta}(F^{\beta bt} dF^{\alpha}_{bt} + F^{\alpha bt} dF^{\beta}_{bt})$$
$$- \tfrac{1}{2} b_{\alpha\beta}({*F}^{\beta bt} F^{\alpha s}{}_t + {*F}^{\alpha bt} F^{\beta s}{}_t) dg_{bs}$$
$$+ \tfrac{1}{2} b_{\alpha\beta}({*F}^{\beta bt} dF^{\alpha}_{bt} + {*F}^{\alpha bt} dF^{\beta}_{bt})$$
$$= \tfrac{1}{2} a_{\alpha\beta}[-(F^{\beta i}{}_k F^{\alpha jk} + F^{\alpha i}{}_k F^{\beta jk}) dg_{ij}$$
$$+ F^{\beta ij} dF^{\alpha}_{ij} + F^{\alpha ij} dF^{\beta}_{ij}]$$
$$+ \tfrac{1}{2} b_{\alpha\beta}[-({*F}^{\beta i}{}_k F^{\alpha jk} + {*F}^{\alpha i}{}_k F^{\beta jk}) dg_{ij}$$
$$+ {*F}^{\beta ij} dF^{\alpha}_{ij} + {*F}^{\alpha ij} dF^{\beta}_{ij}]. \tag{23}$$

Now, the first term in the right-hand side of (23) is zero because of (21), and the second term is zero because of (22) and the identity

$${*F}^{\beta i}{}_k F^{\alpha jk} + {*F}^{\alpha i}{}_k F^{\beta jk} = -\tfrac{1}{2} g^{ij}\psi^{\alpha\beta},$$

which is easy to prove.

In summary, we have obtained that L is a constant when $\phi^{\alpha\beta}$ and $\psi^{\alpha\beta}$ are also constant. Then, when the Lie group is three dimensional we have proved that there is a function f of real variables such that

$$L = f(\phi^{\alpha\beta}; \psi^{\alpha\beta}). \tag{24}$$

Let us suppose now that $\dim G > 3$. Then we have the result (2) from Ref. 2. But, for each α, β, γ fixed, $\phi^{\alpha\beta\gamma}$ and $\psi^{\alpha\beta\gamma}$ are scalar concomitants of a metric tensor and three skew-symmetric tensors, namely, $F^{\alpha}_{ij}, F^{\beta}_{ij}, F^{\gamma}_{ij}$, and so they can be written in the form (24). So the result is also true when $\dim G > 3$.

If $\dim G = 2$, i.e.,

$$L = L(g_{ij}; F^1_{ij}; F^2_{ij}),$$

let F^3_{ij} be an auxiliary and arbitrary skew-symmetric tensor. We can write

$$L = L(g_{ij}; F^1_{ij}; F^2_{ij}; F^3_{ij}),$$

with $\partial L / \partial F^3_{ij} = 0$. Then from (24)

$$L = f(\phi^{\alpha\beta}; \psi^{\alpha\beta}) \quad (1 \leqslant \alpha, \beta \leqslant 3). \tag{25}$$

Differentiating (25) with respect to F^3_{ij} we have

$$0 = \frac{\partial f}{\partial\phi^{13}} F^{1ji} + \frac{\partial f}{\partial\phi^{23}} F^{2ji} + 2\frac{\partial f}{\partial\phi^{33}} F^{3ji} + \frac{\partial f}{\partial\psi^{13}} {*F}^{1ji}$$
$$+ \frac{\partial f}{\partial\psi^{23}} {*F}^{2ji} + 2\frac{\partial f}{\partial\psi^{33}} {*F}^{3ji}.$$

Since $F^1, F^2, F^3, {*F}^1, {*F}^2, {*F}^3$ are linearly independent in a dense subset of the set of the concomitance variables, then

$$\frac{\partial f}{\partial\phi^{\alpha 3}} = \frac{\partial f}{\partial\psi^{\alpha 3}} = 0 \quad (1 \leqslant \alpha \leqslant 3)$$

and so

$$L = h(\phi^{\alpha\beta}; \psi^{\alpha\beta}) \quad (1 \leqslant \alpha, \beta \leqslant 2).$$

In summary, taking account that (2) and (7) are valid in a dense subset of the set of the concomitance variables, we have proved the following theorem.

Theorem: If $L = L(g_{ij}; F^{\alpha}_{ij})$ is a scalar density, then there is a function f of real variables such that

$$L = \sqrt{g}\, f(\phi^{\alpha\beta}; \psi^{\alpha\beta}),$$

in a dense subset of the set of the concomitance variables, where $\phi^{\alpha\beta}, \psi^{\alpha\beta}$ are shown in Eq. (3).

Remark: The result in the theorem is not true in general for the whole set of concomitance variables. Otherwise, every scalar density would be an even function of $F^1, \ldots, F^r$, and $\phi^{\alpha\beta\gamma}$ is a counterexample.

[1]As for any gauge field theory, we are working in a manifold M endowed with the metric tensor g_{ij} and with a G-principal fiber bundle P with base space M. The F^{α}_{ij} are the coefficients of $\sigma^* F$ in some basis of the Lie algebra LG of G, where $\sigma: U \subset M \to P$ is a local section and F is the curvature form. See S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley-Interscience, New York, 1963), Vol. 1.

[2]R. J. Noriega and C. G. Schifini, Gen. Relativ. Gravit. **20**, 337 (1988).
[3]R. J. Noriega and C. G. Schifini, Gen. Relat. Gravit. **16**, 292 (1984).
[4]R. J. Noriega and C. G. Schifini, J. Math. Phys. **28**, 815 (1987).
[5]H. Rund, Abh. Math. Sem. Univ. Hamburg **29**, 243 (1966).

Lorentzian structures of open, spin manifolds

Krystyna Bugajska
Institut für Theoretische Physik, Technische Universität Clausthal, D-3392 Clausthal-Zellerfeld, West Germany

(Received 3 September 1986; accepted for publication 26 October 1988)

The possible set of inequivalent Lorentz structures on an open, connected, spin four-manifold M is investigated. Using the Steenrod correspondence between Lorentz metrics and tangent line bundles, the family of all homotopy classes of possible Lorentz structures on M is obtained. This approach allows one to distinguish explicitly between time-orientable and time-nonorientable Lorentzian metrics. Moreover, it is shown that any two inequivalent, time-orientable Lorentzian structures that correspond to generators of the torsionless part of $H_1(M, Z)$ can be related to each other by using a Morse function with one nondegenerate critical point of index 1 which describes a surgery of type (1,3) in some decomposition of $M = \cup_{i=1}^{\infty} U_i$.

I. INTRODUCTION

Usually we assume that space-time M is a connected, open, differential four-manifold that carries a Lorentzian metric and its associated affine connection and admits a spin structure (i.e., its second Stiefel–Whitney class vanishes).

However, although the notions of differentiable manifold and diffeomorphism go back at least to Poincaré's paper "*Analysic situs*,"[1] a fundamental problem of differential topology, namely, the diffeomorphic classification of manifolds, is not completely solved. Even though the theory of numerical and algebraic invariants, which can distinguish many nondiffeomorphic manifolds, has developed considerably, the simplest problem posed by Poincaré, "Is a three-manifold which is closed and simply connected diffeomorphic to the three-sphere?," has not been answered even now. The understanding of four-manifolds is also not complete; we know by a classical theorem of Markov[2] that the classification of noncompact four-manifolds up to diffeomorphism is impossible. Moreover, Freedman and Donaldson have shown recently[3] that smooth four-manifolds behave in some respects radically different from higher-dimensional manifolds. The most important example of this behavior is the existence of exotic smooth structures on $\mathbb{R}^4$. In other words, there exist smooth manifolds, like the Donaldson–Freedman $\mathbb{R}^4$, that are homeomorphic to $\mathbb{R}^4$, but not diffeomorphic to it. (It is a standard fact that, for $n \neq 4$, exotic $\mathbb{R}^n$'s cannot exist.)

Under these circumstances the classification of all possible Lorentzian structures on a given four-manifold M is perhaps the easiest problem of space-time geometry. In 1959, Finkelstein and Misner classified the Lorentzian metrics of a space-time manifold M that admits a globally hyperbolic splitting $M = S \times \mathbb{R}$ with compactification of S equal to S^3 (Ref. 4). In this case the homotopy classes of time-orientable Lorentzian structures on M are given by $\Pi_3(\mathbb{R}P^3) \cong \mathbb{Z}$. As a matter of fact, for such manifolds, all Lorentzian structures have to be time orientable. The case of a more general manifold M was considered by Shastri *et al.* They investigated the homotopy classes of maps $[M, \mathbb{R}P^3]$ in the case when a parallelizable four-manifold M is a bundle space with a closed, connected, orientable three-manifold S as a base. (We will use the notation $[\cdot, \cdot]$ for the set of homotopy classes of maps.) Now, when we fix a global section s of the bundle LM of linear frames over M then the set $[M, \mathbb{R}P^3] \cong [S, \mathbb{R}P^3]$ numerates the homotopy classes of Lorentzian structures on M. Namely, let s: $M \to LM$ determine a global trivialization of the bundle LM, i.e., $LM \overset{s}{\Leftrightarrow} M \times \mathrm{GL}(4, \mathbb{R})$. Let g denote the Lorentz structure of M related to s, i.e., $g =$ trivial section of the bundle $M \times \mathrm{GL}(4,\mathbb{R}) \bmod \mathrm{SO}(3,1)$. Because any other section of $M \times \mathrm{GL}(4,\mathbb{R}) \bmod \mathrm{SO}(3,1)$ also determines a Lorentzian structure of M, the homotopy classes of these sections are given exactly by $[S, \mathbb{R}P^3]$ [$\mathbb{R}P^3$ is homotopically equivalent to $\mathrm{GL}(4, \mathbb{R})/\mathrm{SO}(3, 1)$]. This set was investigated by the above-mentioned authors in Ref. 5. They use the beautiful and powerful formalism of Postnikov systems and Steenrod squaring operations and they have obtained very remarkable results.

In this paper we present a rather easy approach. Our starting point is the famous Steenrod theorem on the correspondence between Lorentz structures and tangent line bundles[6] and the fact that all Riemannian structures on M are homotopic to each other. It allows us to obtain all homotopy classes of Lorentz structures in the general case of an open, connected, spin four-manifold M. Moreover, in this approach, we can immediately distinguish which "kinks" (i.e., which homotopy classes of Lorentz metrics) correspond to time-orientable structures and which to time-nonorientable ones. Our result is that the set of all homotopy classes of Lorentzian structures on M is given by $H_1^{\infty}(M, Z) \times H^1(M, Z_2)$. Any Lorentzian metric related to a nontrivial component of $H^1(M, Z_2)$ cannot be time orientable. For the case considered by Shastri *et al.*, we get, in number, the same result. However, there is some difference. In Ref. 5 the authors obtain, in addition, a group structure for the set of the homotopy classes of Lorentzian metrics. Namely, they consider the short exact sequence of groups

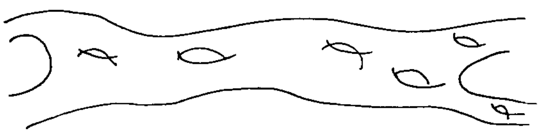
$$0 \to [S,S^3] \xrightarrow{k} [S,\mathbb{R}P^3] \xrightarrow{\mu} [S,\mathbb{R}P^\infty] \to 0 . \tag{1.1}$$

 0022-2488/89/030619-05$02.50

Now time-orientable structures correspond to the image of k. To illustrate this situation, let us consider the case when M is given by R-fibration over $\mathbb{R}P^3$. We have $H_1^\infty(M, Z) \times M'(M, Z_2) = \mathbb{Z} \oplus \mathbb{Z}_2$ whereas $[\mathbb{R}P^3, \mathbb{R}P^3] = \mathbb{Z}$. Although these sets look different there is one to one correspondence between their elements, namely, $\mathbb{Z}$ in the former relation corresponds to $\operatorname{Im} k$ and $\mathbb{Z}_2$ corresponds to $\operatorname{Im} \mu$. Generally, for any space-time manifold M admitting a compact Cauchy surface (i.e., admitting a global hyperbolic splitting $M = S \times \mathbb{R}$) the Shastri *et al.* result is, in some sense, more deep (just because it also defines a group structure of the set of classes of Lorentzian metrics) and our method has only a supplementary character. The reason for this is that the one-point compactification $M^+ = (S \times \mathbb{R})^+$ of M "kills" all nontrivial elements of $H_1(S, Z)$, i.e.,

$$H_1^\infty(S \times \mathbb{R}, \mathbb{Z}) = H_1((S \times \mathbb{R})^+, \mathbb{Z}) = \mathbb{Z}\,.$$

The situation will change when we consider, for example, a manifold M topologically equivalent to $X \times \mathbb{R}_N$, where X is some compact two-manifold and $\mathbb{R}_N$ is an open two-manifold with N "handles":



(N can be countable). In this case, the one-point compactification "cancels" only elements of $H_1(X,\mathbb{Z})$. For this type and for similar types (i.e., when one-point compactification does not cancel the first holonomy) of manifolds we give some relations between some Lorentzian structures on M using the Morse theory. To get this we consider a decomposition of M into an expanding sum of compact-with-boundaries manifolds U_i, $M = \cup_{i=0}^\infty U_i$, such that $U_i \subset U_{i+1}$, U_0 is a four-cell, and U_{i+1} is either a collarlike neighborhood of U_i or U_{i+1} is U_i with a handle of index $\lambda \leqslant n-1$ attached. Next we take into account some Smale–Phillip results that produce a structure theory of differential manifolds by a refined Morse theory.

This paper is organized as follows: In Sec. II we classify Riemannian and Lorentzian structures on $\mathbb{R}^4$. In Sec. III we consider the homotopy classes of Lorentzian structures on an open manifold. We are considering cases with the trivial and nontrivial Stiefel–Whitney classes of corresponding line bundles. In Sec. IV we show that some nonhomotopic trivial line bundles can be related to each other by the use of a Morse function with one nondegenerate critical point of index 1. This Morse function describes a handle of index 1 attached to U_i which cannot be canceled.

II. RIEMANNIAN AND LORENTZIAN STRUCTURES ON $\mathbb{R}^4$

Let us take some concrete, oriented, one-dimensional vector space $X \subset \mathbb{R}^4$ and one of its linear complements, say $N \subset \mathbb{R}^4$,

$$X \oplus N = \mathbb{R}^4\,. \tag{2.1}$$

Now choosing a unit vector of X and some Riemannian structure on N we can determine a concrete Riemannian structure g on $\mathbb{R}^4$. Since we have $GL(3,\mathbb{R})/O(3,\mathbb{R}) = \mathbb{R}^6$ possibilities to fix a Riemannian structure on N and $\mathbb{R}^1$ possibilities to fix a unit vector on X we can construct the set $W = \mathbb{R}^7$ of different Riemannian structures on $\mathbb{R}^4$. To get the remaining set of possible Riemannian structures we have to change the pair (X, N), i.e., we have to change the orthogonality relations that can be defined by this pair. Now, any orthogonal group $O(4, R)$ related to a given Riemannian structure acts transitively on the set of oriented directions of $\mathbb{R}^4$. Hence it is enough to preserve the direction of X and consider only the set of all its linear complements (of course, we can change the role of X and N).

The set of linear complements to X in $\mathbb{R}^4$ is given by $\operatorname{Hom}(\mathbb{R}^3, \mathbb{R}^1) \cong \mathbb{R}^3$. So we see that the whole set $\{\tilde{g}\}$ of inequivalent Riemannian structures on $\mathbb{R}^4$ is equal to the set

$$\{\tilde{g}\} \cong \mathbb{R}^7 \times \mathbb{R}^3 \cong GL(4,\mathbb{R})/O(4,\mathbb{R})\,. \tag{2.2}$$

Let us take one element from this set, say $\tilde{g}$. Now, any line subspace $L \subset \mathbb{R}^4$ uniquely determines some Lorentzian structure g on $\mathbb{R}^4$ according to the formula[7]

$$g(u_1,u_2) = \tilde{g}(u_1,u_2) - 2\tilde{g}(u_1,x)\tilde{g}(u_2,x)/\tilde{g}(x,x)\,,$$
$$\forall u_1,u_2 \in \mathbb{R}^4, \quad x \in L, \quad x \neq 0\,. \tag{2.3}$$

We can easily check that the whole set of different Lorentz structures on $\mathbb{R}^4$ can be given by the set of all pairs $(\tilde{g}, L)$ with $\tilde{g} \in W$ and $L \in \mathbb{R}P^3$. Hence we have

$$\{g\} \cong \mathbb{R}^7 \times \mathbb{R}P^3 \cong GL(4,\mathbb{R})/O(3,1) \tag{2.4}$$

different Lorentz structures on $\mathbb{R}^4$.

Let $\tilde{g} \in W$ and $V \in \mathbb{R}^4$ determine some concrete Lorentz structure g on $\mathbb{R}^4$:

$$g \cong \{\tilde{g}, V\}\,. \tag{2.5}$$

Let V' be any other g-time-like unit vector of $\mathbb{R}^4$. Now if we want to determine the same Lorentz structure g using V' we also have to take another Riemannian metric $\tilde{g}'$, i.e., $g \cong \{g', V'\}$. This metric does not belong to W. Namely, if a is any element of the invariance group of g, $a \in \mathscr{L}_0(g)$, such that

$$V' = aV\,, \tag{2.6}$$

then we can easily see that, $\forall u_1, u_2 \in \mathbb{R}^4$,

$$\tilde{g}'(u_1,u_2) = \tilde{g}(au_1,au_2) \tag{2.7}$$

and that this relation does not depend on the choice of an element a that satisfies (2.6).

III. LINE BUNDLES OVER M

In this section we will look for the homotopy classes of possible Lorentzian metrics on an open manifold M (this means that we cannot pass from one Lorentzian metric to the other by a continuous deformation of a metric tensor g). From our previous considerations we see that we can try to find them using Riemannian structures and tangent line bundles over M. Let us assume that we have some Riemannian metric $\tilde{g}$ on M. Moreover, let us assume that $\tilde{g}$ is given by a section of a bundle η whose fiber at a point $x \in M$ contains the set of all different Riemannian structures on TM_x related to some fixed pair (X_x, N_x). In other words η can be given by a concrete decomposition of TM into line and codimension 1 subbundles of TM. (Since M is open, its Euler class vanishes and such a decomposition always exists.) Now any Lorentzian metric on M can be introduced by a section of η and some tangent line bundle of M uniquely.

From the general theory of obstructions, any two sections of η are homotopic to each other, i.e., the obstruction cocycle for any map $F\colon M\times I\to\eta$ vanishes. Thus to find classes of inequivalent Lorentz structures of M we have to classify tangent line bundles over M. We see that the introduction of a bundle η as well as the decomposition of TM is only auxiliary. We use them only to relate classes of Lorentzian metrics with classes of line bundles.

A. A trivial case

Let L be a trivial tangent line bundle over M. In this case there exists a global nonvanishing section Y of the unit vectors of L. Hence to find all homotopy classes of trivial line bundles we have to consider the homotopy classes of nonvanishing vector fields.

By definition,[8] the homotopy of Y_0 is a map

$$F\colon\ M\times I\to TM\,,$$

such that, for a fixed $t\in I$,

$$Y_t(x):=F(x,t)\in\Gamma(TM)$$

is a nonvanishing vector field on M and $F(x,0)=Y_0$. In this case Y_0 is homotopic to $Y_1=F(x,1)$. In other words if we can construct a nowhere-zero section S of the bundle $p^*(TM)$ (where $p\colon M\times I\to M$ is the projection) then we have homotopy between $S|_{M\times\{0\}}$ and $S|_{M\times\{1\}}$. Hence the obstruction to have homotopy between Y_0 and Y_1 is given exactly by the obstruction of the existence of a nonvanishing section S of $p^*(TM)$ with $S|_{M\times\{0\}}=Y_0$ and $S|_{M\times\{1\}}=Y_1$. Thus the obstruction cocycle belongs to $H^3(M,\Gamma)$, where Γ is the system of coefficients[9] whose value at $x\in M$ is $\Pi_3(S_x^3)\cong\Pi_3(\mathring{T}M_x)$. (Here $\mathring{T}M_x$ denotes the space TM_x with zero deleted.) Now by the generalized Poincaré duality[10] for an open manifold M we have

$$H^3(M;\Gamma)\cong H_1^\infty(M,Z)\,. \tag{3.1}$$

In this way the set of all homotopy classes of nonvanishing vector fields on M is given by $H_1^\infty(M,Z)$. However, we have a well known natural relation between the integral first homology group of M and the fundamental group of M.[11] It yields to the assertion that if M is pathwise connected then the natural homomorphism

$$h\colon\ \Pi_1(M,x)\to H_1(M,Z) \tag{3.2}$$

has the commutator subgroup of $\Pi_1(M,x)$ as its kernel. It means that we have the natural isomorphism between the group $H_1(M,Z)$ and $\Pi_1(M,x)$ made Abelian. Thus for an open manifold M some classes of homotopic sections of the bundle $\mathring{T}M$ can be given by elements of $\mathrm{Ab}[\Pi_1(M,x_0)]$. In our consideration we will be interested in the set of homotopy classes of trivial line bundles over M that correspond to the set of those elements of $\mathrm{Ab}[\Pi_1(M,x_0)]$ related to elements of $H_1(M,Z)$ contained in $H_1^\infty(M,Z)$.

B. A nontrivial case

It is known[12] that any class of isomorphic line bundles over M is determined by a homotopy class of maps from M to the classifying space B_{Z_2} of Z_2-bundles. Since $B_{Z_2}=\mathbb{R}P^\infty$ is an Eilenberg–MacLane space of type $(Z_2,1)$, i.e., $B_{Z_2}\cong K(Z_2,1)$, the relation

$$[M,K(\Pi,q)]\cong H^q(M,\Pi)$$

gives

$$[M,B_{Z_2}]\cong H^1(M,Z_2)\,. \tag{3.3}$$

(Here $[\cdot,\cdot]$ denotes the set of homotopy classes of maps.) In this way any class of isomorphic line bundles is uniquely characterized by its Stiefel–Whitney class $w_1(\cdot)\in H^1(M,Z_2)$. Let ξ be some linear bundle over M characterized by $w_1(\xi)\in H^1(M,Z_2)$. Can ξ be embedded into the tangent bundle TM? In other words does a nowhere vanishing bundle map from ξ to TM over all of M exist? Any such map is equivalent to a nowhere vanishing section of $\mathrm{Hom}(\xi,TM)\cong TM\otimes\xi$. We will denote this bundle by z. Its fiber is isomorphic to $\mathbb{R}^4$. However, we are interested rather in the bundle $\mathring{z}$ of the nonzero elements of z. The obstruction for the existence of a section of $\mathring{z}$ lies in the fourth cohomology group of M in a coefficient bundle $\mathbb{Z}:=\mathring{\eta}(\Pi_3)$.[13] The fiber of this coefficient bundle is given by the third homotopy group of the fiber of $\mathring{z}$. However, for our open manifold M the fourth cohomology group $H^4(M,\sigma)$ vanishes for any coefficient bundle σ. This means that any line bundle over M can be embedded in TM and that for every $\sigma\in H^1(M,Z_2)$ there exists a tangent line bundle $L\subset TM$ such that

$$w_1(L)=\sigma\,.$$

Now let us try to find the homotopy classes of nontrivial tangent line bundles characterized by a given element $\sigma\in H^1(M,Z_2)$.

Again the homotopy of $L\subset TM$ is a map

$$F\colon\ M\times I\to PTM$$

such that $F(x,0)=L_0$, $F(x,1)=L_1$, $w_1(L_0)=w_1(L_1)=\sigma$. (Here PTM is the projective tangent bundle.) Since the map F is equivalent to a section of the projective bundle $P(p^*TM)$, i.e.,

$$S\colon\ M\times I\to P_{p^*}TM,\quad S|_{M\times\{i\}}=L_i,\quad i=0,1\,,$$

the obstruction cocycle of S lies in $H^3(M,PTM(\Pi_3))$ with a system of coefficients having, at $x\in M$, the value

$$\Pi_3(PTM_x)\cong\Pi_3(\mathbb{R}P^3)=\Pi_3(S^3)\cong\Pi_3(S_x^3)\,.$$

Thus we have obtained the result that the set of homotopic classes of tangent line bundles characterized by a given element $\sigma\in H^1(M,Z_2)$ can be enumerated by the elements of $H_1^\infty(M,Z)$. So the set of all homotopy classes of Lorentzian structures on M is given by

$$A=H_1^\infty(M,Z)\times H^1(M,Z_2)\,.$$

IV. RELATIONS BETWEEN CLASSES OF INEQUIVALENT "TRIVIAL" LORENTZ STRUCTURES

Since all Riemannian structures on M are homotopic we can introduce the classes of Lorentzian structures on M by the classes of homotopic tangent line bundles of M. Let us consider a case where a Lorentz structure g can be given by a section $\tilde{g}$ of η and a trivial line bundle L,

$$w_1(L) = 0 . \tag{4.1}$$

Now let us consider a realization of M by an expanding union of compact manifolds $\{U_i\}$ with boundaries

$$M = \bigcup_{i/0}^{\infty} U_i , \tag{4.2}$$

such that $U_i \subset U_{i+1}$, U_0 is a four-cell, and either U_{i+1} is a collarlike neighborhood of U_i or U_{i+1} is U_i with a handle of index $\lambda \leqslant n-1$ attached.[10] In the latter case we can say that ∂U_{i+1} can be obtained from ∂U_i by a surgery of type $(\lambda, n-\lambda)$. Let us consider some concrete $U_i \subset M$ such that ∂U_{i+1} is obtained from ∂U_i by a surgery of type (1.3) or equivalently that we can get U_{i+1} by attaching a handle of index 1 to U_i. Moreover, let us assume that U_{i+2} cannot be related to U_{i+1} by attaching a handle of index 2 since in this case these two handles cancel each other.[14]

Let D_n denote an n-disk and let $\varphi_i: S^0 \times D_3 \to \partial U_i$ be a characteristic embedding of our surgery. Then $\partial U_i \cup D_1/\sim$ is a deformation retract of $U_{i+1} - \operatorname{Int} U_i$. The relation $\sim$ is given by the identification of $\partial D_1 = S^0$ with its image under a map φ. Let f_j be a Morse function of a smooth triad $(U_{j+1} - \operatorname{Int} U_j, \partial U_j, \partial U_{j+1})$, i.e., $\forall_j$,

$$f_j:(U_{j+1} - \operatorname{Int} U_j, \partial U_j, \partial U_{j+1}) \to ([0,1],0,1) . \tag{4.3}$$

This means that for the i mentioned above, f_i has one nondegenerate critical point p of index 1; $p \in U_{i+1} - U_i$. Since there always exists an $\epsilon > 0$ such that $f_{i-1}^{-1}(1-\epsilon, 1]$ has no critical points, we have a well defined homotopy class $[\lambda]$ of loops in

$$(U_{i+1} - \operatorname{Int} U_i) \cup f_{i-1}^{-1}(1-\epsilon, 1] , \tag{4.4}$$

which close D^1 in a handle equivalent to

$$U_i \cup (D_1 \times D_3)/\varphi . \tag{4.5}$$

Let us take $D_1 = [a, b]$, $a < b < \infty$, $\partial D_1 = S^0 = \{a, b\}$. Let

$$\varphi(a) = x_0 \in \partial U_i . \tag{4.6}$$

Then λ is a loop at x_0 that represents the just-mentioned homotopy class of loops given by the effective attachment of a handle of index 1 to U_i.

Since $N_i = M - (U_{i+1} - \operatorname{Int} U_i)$ is open we can fit the Morse function f_i together with some Morse function without any critical point on N. We get a proper Morse function f on M with only one nondegenerate critical point at p. This function is related to the adding of a handle of index 1 to U_i as well as to the homotopy class of λ at $x_0 \in \partial U_i$.

Proposition: Let L_0 and L_1 be two nonhomotopic, trivial, tangent line bundles over M. Let Y_0 and Y_1 be their sections of unit vectors with respect to some Riemannian structure $\tilde{g}$ on M. Let Y_0 represent a homotopy class of nonvanishing vector fields on M related to the trivial element of $H_1^{\infty}(M, Z)$. Analogously let Y_1 be related to an element $\sigma \neq 0$, $\sigma \in H_1^{\infty}(M, Z)$. If $\sigma \in H_1(M, Z)$ is a generator of the torsionless part of $H_1(M, Z)$ which corresponds to the homotopy class of λ under the isomorphism h (3.2), then there exists a section S of the bundle p^*TM that has the following properties: (i) $\forall t \in [0,1]$, $t \neq t_0$, $S(x, t)$ is a nowhere-zero vector field on M; (ii) $S(x, 0) = Y_0(x)$, $S(x, 1) = Y_1(x)$; and (iii) $S(x, t_0)$ has an isolated zero at a point p of index $(-1)^1$. Moreover $S(x, t_0)$ is homotopic to a gradient field of a Morse function f described above.

Proof: It is known[10] that on an open manifold any nonzero vector field Y is homotopic (through nonzero vector fields) to a nowhere-zero gradient field of some function ϕ. In local coordinates $\{u^i\}$ a gradient field can be given by

$$\operatorname{grad} \phi = \Sigma\, \tilde{g}_{ij} \frac{\partial \phi}{\partial u^i} \frac{\partial}{\partial u^j} , \tag{4.7}$$

or equivalently by

$$\langle X, \operatorname{grad} \phi \rangle = X(\phi), \quad \forall X \in \Gamma(TM) . \tag{4.7'}$$

Thus we can assume that $Y_0 = \operatorname{grad} \phi_0$ and $Y_1 = \operatorname{grad} \phi_1$. However, since

$$\Pi_0(\operatorname{Sub} M, \mathbb{R}^1) = \Pi_0(\Gamma(\mathring{T}M)) , \tag{4.8}$$

the functions ϕ_0 and ϕ_1 belong to different components of $\Pi_0(\operatorname{Sub} M, \mathbb{R}^1)$. Moreover, we see that the obstruction cocycle for Y_0, Y_1 to be homotopic is exactly the same as for submersions ϕ_0, ϕ_1 to be homotopic. In other words, there exists a function

$$F\colon M \times I \to \mathbb{R} \tag{4.9}$$

that has the properties $F(x, 0) = \phi_0(x)$, $F(x, 1) = \phi_1(x)$, and there exists only one element $t_0 \in (0, 1)$ such that the obstruction cocycle (here we are using Steenrod's notation[15])

$$c(\sigma_3, F_{t_0}) \in H^3(M, \Gamma) \tag{4.10}$$

is a nontrivial one $[F_{t_0} = F(\ , t_0)]$.

Now using the generalized Poincaré duality (3.1) and the natural isomorphism h (3.2) we obtain immediately that F is a Morse function of $M \times I$ with one critical point of index 1 at (p, t_0) related to a surgery of type (1.3) between $\partial U_i \times \{t_0\}$ and $\partial U_{i+1} \times \{t_0\}$. Now taking the field grad F we obtain our section S: $M \times I \to p^*TM$.

Corollary: Let L_0 and L_1 be the two line bundles considered above. Let g_0 and g_1 be the two Lorentz structures on M determined by $\tilde{g}$ and L_0 and L_1, respectively, i.e.,

$$g_0 \cong (\tilde{g}, L_0)$$

and

$$g_1 \cong (\tilde{g}, L_1) .$$

Then we can pass from g_0 to g_1 only when we use a Morse function F with one critical point of index 1 as described above. It means that we can pass from g_0 to g_1 only when we recognize a handle of index 1 in a decomposition $\cup_{i/0}^{\infty} U_i$ of M.

Now let us take into account nonisomorphic classes of tangent line bundles. Since $H^1(M, Z_2)$ (which enumerates these classes) can be considered as $\operatorname{Hom}(H_1(M, Z), Z_2)$, we cannot relate two line bundles with different Stiefel–Whitney classes in a similar way. First, if $w_1(L) \neq 0$, then we have no global nowhere-zero section of L, i.e., we have no relation between L and any nonvanishing gradient field. Besides we cannot repeat the above construction even for nonhomotopic tangent line bundles of the same nontrivial Stiefel–Whitney class. Namely, if $w_1(L_0) = w_1(L_1) \neq 0$ and L_0 is not homotopic to L_1, then, although the obstruction cocycle for this homotopy can be exactly the same as for some appropriate trivial line bundles, we have a situation completely different from the previous one. Of course, we can say that a surgery of type (1, 3) is responsible for the lack of homotopy between L_0 and L_1, but any Morse function (or equivalently any gradient field) that determines this surgery cannot be incorporated to connect L_0 and L_1.

V. SUMMARY

In the classical theory of matter we begin with a blank differentiable, open, orientable, spin four-manifold M. Next we add a Lorentzian metric g and any other physical fields. We assume that these fields obey equations expressed as relations between tensors on M and all derivatives are covariant derivatives with respect to the Levi-Civita connection defined by the metric g. Thus it is natural to ask how many homotopy classes of Lorentzian metric structures can be carried by M and whether there exist some relations between some of them. Because all Riemannian metrics are homotopic to each other we can distinguish different classes of Lorentzian structures by investigation of tangent line bundles over M.

First we should consider time-orientable metric structures. They correspond to trivial line bundles. Since we can fix some Riemannian metric, say $\tilde{g}$, without any consequences for our classification, we can represent any trivial tangent line bundle by its section of unit vectors. Now, owing to the general topological properties of M [orientability $\Rightarrow w_1(TM) = 0$; spin manifold $\Rightarrow w_2(TM) = 0$; open $\Rightarrow w_4(TM) = 0$; four-manifold $\Rightarrow w_3(TM) = 0$], we can introduce the bundle $\mathscr{K}$ of unit vectors on M in such a way that any section of $\mathscr{K}$ is given by the unit section of some line bundle over M. So, to find the homotopy classes of tangent line bundles over M we have to find the homotopy classes of sections of $\mathscr{K}$. Since the obstruction cocycle for such homotopy belongs to $H^3(M;\Gamma)$ (introduced in Sec. III A) we obtain $H_1^{\infty}(M, \mathbb{Z})$ different classes of inequivalent, time-orientable Lorentzian structures on M. Moreover, any two time-orientable Lorentzian structures L_i, $i = 0, 1$, that correspond to elements $\sigma_i \in H_1^{\infty}(M, \mathbb{Z})$, such that $\sigma_0 = 0$ and σ_1 determines a generator of the torsionless part of $H_1(M, \mathbb{Z})$, can be related to each other by the gradient field of a Morse function with one nondegenerate critical point of index 1 which describes a handle of index 1 in some decomposition of $M = \cup_{i/0}^{\infty} U_i$.

In the case of a time-nonorientable Lorentzian structure a corresponding tangent line bundle has to have the nontrivial Stiefel–Whitney class. Now, since for open manifolds every line bundle can be embedded into the tangent bundle TM (see Sec. III B) we obtain $H^1(M, \mathbb{Z}_2)$ classes of nonisomorphic tangent line bundles of M. The classification of tangent line bundles with a given nonvanishing Stiefel–Whitney class $\sigma \in H^1(M, \mathbb{Z}_2)$ up to homotopy is given by $H^3(M, PTM(\Pi_3))$. Hence we can have again $H_1^{\infty}(M, \mathbb{Z})$ homotopic classes of tangent line bundles characterized by a given element σ.

In this way an open, spin four-manifold can carry

$$H^1(M,\mathbb{Z}_2) \times H_1^{\infty}(M,\mathbb{Z})$$

inequivalent Lorentz structures. Here, by inequivalent structures we understand nonhomotopic ones. The possible physical consequences of these investigations are contained in Refs. 16 and 17.

ACKNOWLEDGMENTS

I wish to thank Professor H. Doebner for his kind hospitality at Clausthal University as well as for his help and many valuable discussions.

[1]H. Poincaré, *Collected Works*, Vol. 6 (Gauthier-Villars, Paris, 1953).
[2]A. Markov, Dokl. Ak. Nauk SSSR **121**, 218 (1958).
[3]M. H. Freedman, Ann. Math. **110**, 177 (1979).
[4]S. Donaldson, J. Differ. Geom. **18**, 279 (1981); D. Finkelstein and C. W. Misner, Ann. Phys. (NY) **6**, 230 (1959).
[5]A. Shastri, J. Williams, and P. Zvengrowski, Int. J. Theor. Phys. **19**, 1 (1980).
[6]G. Whiston, Int. J. Theor. Phys. **12**, 225 (1975).
[7]S. W. Hawking and G. F. R. Ellis, *The Large Scale of Space-Time* (Cambridge U.P., Cambridge, 1973).
[8]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, NJ, 1951), p. 167.
[9]See Ref. 8, p. 151.
[10]A. Phillips, Topology **6**, 171 (1966).
[11]S. T. Hu, *Homotopy Theory* (Academic, New York, 1959), p. 44.
[12]D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York).
[13]See Ref. 8, p. 154.
[14]V. Poenaru, "On the geometry of differential manifolds," in *Studies in Mathematics*, Vol. 5 (Prentice-Hall, Englewood Cliffs, NJ).
[15]See Ref. 8, p. 167.
[16]K. Bugajska, Int. J. Theor. Phys. **26**, 623 (1987).
[17]K. Bugajska, submitted to Math. Z.

A gauge condition for orthonormal three-frames

James M. Nester
Department of Physics, National Central University, Chung-li, Taiwan 32054, Republic of China

(Received 21 June 1988; accepted for publication 26 October 1988)

A set of natural gauge conditions for an orthonormal frame on a three-dimensional Riemannian manifold is discussed. The conditions determine a teleparallel geometry. They may be regarded as a nonlinear elliptic system for a rotation. Existence and uniqueness of solutions are discussed using the linearized system. Applications to Einstein gravity are noted.

I. INTRODUCTION

Orthonormal frames frequently provide a convenient way to work with Riemannian geometries. The choice of orthonormal frame, however, is not unique; consequently, at times a gauge condition fixing the rotational freedom of the frame is desired. Algebraic conditions are sometimes used, e.g., symmetry of the components with respect to an extra fixed coordinate basis.[1] For many applications, however, more geometric gauge conditions are preferable.

Here we present new, purely geometric, natural gauge conditions for the choice of orthonormal frame on a three-dimensional Riemannian manifold. Up to an overall constant rotation, the conditions select a preferred orthonormal frame and hence a preferred parallelism, i.e., a teleparallel geometry. The gauge conditions may be regarded as a nonlinear elliptic system for a rotation. Via the linearization of this elliptic system, unique solutions are shown to exist for geometries in a neighborhood of Euclidean space.

Our concern in this work is with our local gauge conditions and not with global topology. In particular, we want certain closed one-forms to be exact. Hence we shall simply assume that the first cohomology vanishes. Only future investigations can reveal to what extent our conditions inherently depend on this assumption and, consequently, whether they can be extended to more general topologies.

The new gauge conditions have already proved their worth in applications to asymptotically flat solutions for Einstein's theory of gravity. We first recognized the value of these conditions when we realized that they allow a locally positive representation for the Hamiltonian density and thereby a new proof[2] of the positivity of total energy. Moreover, as we shall discuss, they show further promise for applications to the initial value problem.

II. RIEMANNIAN AND TELEPARALLEL GEOMETRY

Before going into the gauge conditions, we review certain ideas of Riemannian and teleparallel geometry. Our discussion is conveniently done in terms of differential forms. Let θ^a for $a = 1,2,3$ be three linearly independent one-forms that may thus be used as a coframe basis. They are dual to a vector basis denoted here by e_a. We assume that our manifold is connected and orientable; via Ref. 3 we have learned that every orientable three-manifold is parallelizable,[4] so our frames may be assumed to be globally defined.

A linear connection is determined by the connection one-form coefficients $\omega^a{}_b := \theta^a(\nabla e_b)$. The connection determines both the *curvature* two-form

$$\Omega^a{}_b := d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b = \tfrac{1}{2} R^a{}_{bkm}\theta^k \wedge \theta^m , \qquad (1)$$

and the *torsion* two-form

$$\Theta^a := \tfrac{1}{2} Q^a{}_{bc}\theta^b \wedge \theta^c = D\theta^a = d\theta^a + \omega^a{}_b \wedge \theta^b . \qquad (2)$$

Each coframe may also be used to determine a special associated metric: namely, the one in which this particular frame is *orthonormal*, $g = g_{ab}\theta^a \otimes \theta^b = \theta^1 \otimes \theta^1 + \theta^2 \otimes \theta^2 + \theta^3 \otimes \theta^3$. A metric, of course, determines a Riemannian geometry with a connection that is both metric compatible and torsion-free. In an orthonormal frame, the metric compatibility condition is just antisymmetry of the connection one-forms, $\omega_{ab} = -\omega_{ba}$, while the torsion-free condition is $d\theta^a = -\omega^a{}_b \wedge \theta^b$. These two conditions uniquely determine the connection one-form via the components $\omega_{abc} = \omega_{ab}(e_c) = \tfrac{1}{2}(C_{cab} - C_{bca} - C_{abc})$, where $C^a{}_{bc} = -d\theta^a(e_b, e_c)$.

While the orthonormal coframe determines the metric and thereby the Riemannian geometry, the converse is not the case. A given Riemannian geometry determines only an equivalence class of orthonormal frames. From one frame θ^a, orthonormal with respect to g, we may obtain many others by applying a position-dependent (local) rotation

$$\theta^{a'} = R^{a'}{}_b \theta^b . \qquad (3)$$

Under such a transformation, the connection one-form components transform inhomogeneously,

$$\omega^{a'}{}_{b'} = R_k{}^{a'}(\omega^k{}_m R^m{}_{b'} + dR^k{}_{b'}) , \qquad (4)$$

while the torsion and curvature are tensorial.

The usual viewpoint is that the metric does not favor any special orthonormal frame; that any such selection is purely an ungeometric gauge choice. The gauge conditions we propose here, however, are constructed purely out of the above outlined Riemannian geometry. They use only this Riemannian geometry to select a certain preferred orthonormal frame. They can be regarded as first-order differential conditions on the connection coefficients $\omega^a{}_b$, or alternately on $C^a{}_{bc}$. There is some resemblance to the Lorentz and Coulomb gauge conditions used in electromagnetism.

The discussion of the gauge conditions could quite nicely be carried out strictly in terms of Riemannian geometry. The significance of the ideas can be better appreciated, however, in terms of a different language: *teleparallel* or *Weitzenböck* geometry.[5] This type of geometry is quite suited to our goal: the selection of a favored global orthonormal frame. A preferred orthonormal frame θ^a may be used to define a new parallelism. The new parallel transport is simply accomplished by keeping the coefficients of an object constant with

 0022-2488/89/030624-03$02.50

respect to this preferred frame. In other words, the new connection coefficients *vanish* in this special frame, which is referred to as *orthoteleparallel* (OT). Of course, the new connection coefficients are generally nonvanishing in any other orthonormal frame. However, since the new *curvature* two-form vanishes in this special frame, it *vanishes* also in every other orthonormal frame. This is tied in with the fact that the new parallel transport is path independent.

Conversely, if parallel transport is path independent, the curvature vanishes. There then exists a special orthonormal frame in which the connection one-forms vanish. This OT frame is uniquely determined up to constant (global) rotations. It may be constructed by choosing any orthonormal frame at a point and parallel transporting it to the other points. Although the curvature vanishes, the parallel transport is not at all trivial. It is characterized by the torsion. In the OT frame, the torsion two-form is simply

$$\Theta^a = \tfrac{1}{2}Q^a{}_{bc}\theta^b\wedge\theta^c = d\theta^a = -\tfrac{1}{2}C^a{}_{bc}\theta^b\wedge\theta^c. \qquad (5)$$

Teleparallel geometry is thus a kind of opposite to Riemannian geometry; curvature vanishes and torsion characterizes the geometry. Somewhat surprisingly, teleparallel geometry is much less restrictive than Riemannian geometry. Indeed there are many ways to construct a teleparallel geometry from a given Riemannian geometry. One need merely choose any orthonormal frame and declare it to be an OT frame. Thus one Riemannian geometry corresponds to a whole equivalence class of teleparallel *geometries*.

In these terms, our gauge conditions select a certain preferred representative from each of these equivalence classes of geometries. Stated another way, they determine a single special teleparallel geometry for each Riemannian geometry.

Some further notation will be needed. We use i_a as shorthand for the interior product $i_{e_a}\beta = \beta(e_a,...)$ which takes p-forms to $p-1$ forms. It is convenient to use also the dual basis for the Grassmann algebra

$$\begin{aligned}\zeta &= (1/3!)\zeta_{abc}\theta^a\wedge\theta^b\wedge\theta^c = \theta^1\wedge\theta^2\wedge\theta^3 = *1,\\ \zeta_a &= i_a\zeta = \tfrac{1}{2}\zeta_{abc}\theta^b\wedge\theta^c = *\theta_a, \qquad (6)\\ \zeta_{ab} &= i_b\zeta_a = \zeta_{abc}\theta^c = *(\theta_a\wedge\theta_b),\end{aligned}$$

where ζ_{abc} is the three-dimensional, totally antisymmetric Levi-Civita tensor with $\zeta_{123} = +1$ and $*$ is the Hodge dual.

III. THE GAUGE CONDITIONS

The gauge conditions are discussed here in terms of the teleparallel geometric quantities. The discussion may easily be transcribed into Riemannian or strictly frame terminology via $Q^a{}_{bc} = -C^a{}_{bc} = \omega^a{}_{bc} - \omega^a{}_{cb}$. The conditions are explicitly stated in terms of two quantities constructed from θ^a regarded as an OT frame: a one-form

$$\tilde{q} = q_b\theta^b = Q^a{}_{ab}\theta^b = i_a\Theta^a = i_a\, d\theta^a, \qquad (7)$$

and a function $*q$ dual to the three-form

$$q = \theta_a\wedge\Theta^a = g_{ab}\theta^b\wedge d\theta^a = \tfrac{1}{2}Q_{abc}\theta^a\wedge\theta^b\wedge\theta^c. \qquad (8)$$

The gauge conditions are simply that the forms $*q$, $\tilde{q}$ are closed,

$$d*q = 0 = d\tilde{q}. \qquad (9)$$

Note that this is the correct number of constraints on the orthonormal frame; the first condition says that $*q$ is constant (our manifold is connected) and the second condition provides two more restrictions.

To fix a solution, appropriate boundary conditions must be chosen. For example, on $\mathbb{R}^3$ the Cartesian frame satisfies the conditions (9), but so does the spherical frame (except, of course, at the origin). We are particularly interested in applications to manifolds which are asymptotically flat in the sense that, with respect to suitable asymptotic coordinates, the metric coefficients differ from the Euclidean metric by terms of order $\mathbb{O}(1/r)$, and the first m derivatives fall off at order $\mathbb{O}(1/r^{m+1})$, for $m\leqslant 3$. On asymptotically flat manifolds, the boundary condition is that q vanishes; whereas on the three-sphere, q must be nonvanishing. Note that as $\tilde{q}$ is closed, it is locally exact. If $\tilde{q}$ is globally exact (in particular when the first cohomology vanishes as we assume), it determines a function up to a constant which may be suitably normalized, e.g., at infinity.

Our attention was first attracted to these conditions when we realized that they would allow a locally positive representation for the Hamiltonian density and thereby permit a new, strictly tensorial (in contrast to the Witten[6] spinor method) proof[2] of positive energy for Einstein's theory of gravity. We have since noticed that they show promise for other applications to Einstein gravity. In particular, they mesh nicely with the standard initial value problem analysis.[7] The conformal change of three-metric $g\to\psi^4 g$ has proved to be very useful in the initial value problem of general relativity. Under the corresponding conformal change of three-frame $\theta^a\to\psi^2\theta^a$, the quantities q, $\tilde{q}$ become $\psi^4 q$ and $\tilde{q} - 2d\ln\psi^2$, respectively. Consequently, on an asymptotically flat space where q vanishes, the gauge conditions (9) are conformally invariant. Moreover, as already noted, since $\tilde{q}$ is exact, it defines (modulo a constant) a special function. This construction may provide the best definition of the generalization of the Newtonian potential: the scale factor which satisfies the super-Hamiltonian constraint equation.

IV. EXISTENCE AND UNIQUENESS

The restrictions (9) are "good" gauge conditions if for any three-metric there exists a "unique" (i.e., up to a constant rotation) orthonormal frame $\theta^{a'}$ such that they are satisfied. Rather than analyze the gauge conditions as restrictions on the choice of orthonormal frame, we have found it easier to study them as conditions on a local rotation. Thus given *any* orthonormal frame θ^a, we want to show that there exists a unique local rotation such that $\theta^{a'} = R^{a'}{}_b\theta^b$ satisfies

$$d*q' = 0 = d\tilde{q}'. \qquad (9')$$

Explicitly we find

$$q' = \theta_{a'}\wedge d\theta^{a'} = q + \theta_m\wedge R^m{}_{b'}\, dR^{b'}{}_c\wedge\theta^c, \qquad (10)$$

$$\tilde{q}' = i_{a'}\, d\theta^{a'} = \tilde{q} + R^b{}_{a'}R^{a'}{}_{c,b}\theta^c. \qquad (11)$$

Given θ^a, the conditions obtained by substituting the expressions (10) and (11) into (9′) constitute a nonlinear second-order elliptic system for the rotation $R^{a'}{}_b$.

For nonlinear systems, existence and uniqueness are difficult questions. We examine the linearized conditions. Infinitesimally, $R^{b'}{}_a = \delta^b{}_a + \zeta^b{}_{ac}A^c$, so

$$q' = q - dA^c \wedge \zeta_{cab}\theta^a \wedge \theta^b = q - 2A^c{}_{,c}\zeta$$
$$= q - 2\{d(A_c\zeta^c) + A_c q^c \zeta\}$$
$$= q - 2\{d*A + A \wedge *\tilde{q}\} \tag{12}$$

and

$$\tilde{q}' = \tilde{q} + A_{b,c}\zeta^{bc}{}_k\theta^k = \tilde{q} - *(dA_b \wedge \theta^b)$$
$$= \tilde{q} - *(dA - A_b\, d\theta^b), \tag{13}$$

where $A = A_b\theta^b$.

Via these linearized equations, we shall show that the gauge conditions (9) have unique solutions for geometries within a neighborhood of Euclidean space. Let us first assume that conditions (9) can be satisfied for a given geometry. A nearby geometry will certainly have an OT frame θ^a for which $d*q$, $d\tilde{q}$ are small. An infinitesimal rotation of amount A will produce a $\theta^{a'}$ satisfying the desired gauge conditions (9′) if

$$d*d*A + d*(A \wedge *\tilde{q}) = d(\tfrac{1}{2}*q), \tag{14}$$
$$d*dA - d*(A_b\, d\theta^b) = d\tilde{q}. \tag{15}$$

Applying $*$ and using the codifferential $\delta = (-1)^p *d*$ yields

$$d\delta A - d*(A \wedge *\tilde{q}) = d(-\tfrac{1}{2}*q), \tag{16}$$
$$\delta\, dA - \delta(A_b\, d\theta^b) = \delta(*\tilde{q}). \tag{17}$$

Adding gives a linear elliptic equation for the one-form A,

$$LA := \Delta A - \delta(A_b\, d\theta^b) - d*(A \wedge *\tilde{q})$$
$$= \delta(*\tilde{q}) + d(-\tfrac{1}{2}*q), \tag{18}$$

with the elliptic operator L being a perturbation of the nicest possible operator: the Laplace–Beltrami operator, $\Delta = d\delta + \delta d$. Given appropriate boundary conditions, $\Delta A = \rho$ has unique solutions; Eq. (18) should also have unique solutions, at least as long as $d\theta^a$ is not too large.

Formally, solutions to Eq. (18) may be constructed by iteration. Thus consider the sequence of elliptic equations

$$\Delta A^0 = \delta(*\tilde{q}) + d(-\tfrac{1}{2}*q), \tag{19}$$
$$\Delta A^{m+1} = \delta(A^m_b\, d\theta^b) + d*(A^m \wedge *\tilde{q}). \tag{20}$$

Any solution to each of these equations is unique, since our vanishing first cohomology assumption means that there are no harmonic one-forms. The operator Δ is self-adjoint. For self-adjoint elliptic operators, uniqueness implies existence. Consequently, each of the linear elliptic equations [(19) and (20)] then has unique solutions. The sum $A = \Sigma_{m=0}^{\infty} A^m$ formally solves Eq. (18). This sum will certainly converge if $d\theta^b$ is sufficiently small. Consider, in particular, the case where there is a one-parameter curve in the space of geometries connecting our geometry with Euclidean space. If $d\theta$ is of order ϵ, then A^m is of order ϵ^{m+1}. Moreover, on an asymptotically flat space, the boundary conditions are stated in terms of the fall off rates $d\theta$, q, $\tilde{q} \to O(1/r^2)$, and $A^m \to O(1/r^{m+1})$. Consequently, the sum will certainly converge asymptotically.

A more technical argument can be based on the work of Choquet-Bruhat and Christodoulou[8] for elliptic operators on manifolds which are Euclidean at infinity. Briefly, with $d\theta^a$ in the Sobolev space, $H_{s,\delta}$ with differentiability $s = 2$, and fall off parameter $\delta = 0$, the conditions of the theorems in Sec. VI are met and we conclude that $L: H_{3,1} \to H_{1,-1}$ is an isomorphism for geometries in a neighborhood of Euclidean space.

Note that a solution to Eq. (18) also solves Eqs. (16) and (17) separately, as they are mutually orthogonal with respect to the inner product $(\alpha,\beta) = \int *\alpha \wedge \beta$. Retracing our deduction of conditions (16) and (17) leads us to conclude that the perturbed geometry also satisfies the desired gauge conditions (9′). We conclude that the gauge conditions (9) are good. They certainly have unique solutions, at least for geometries which are in a neighborhood of Euclidean space.

V. DISCUSSION

Our aim was to introduce and encourage investigation of a certain set of natural gauge conditions for orthonormal frames on a three-dimensional Riemannian manifold. The gauge conditions (9) should determine a unique teleparallel geometry, i.e., an orthonormal frame up to a constant rotation.

The gauge conditions (9) are elliptic but nonlinear. We have argued that they have unique solutions for geometries that are in a neighborhood of Euclidean space. We know of no counterexample and conjecture that they are of quite general validity. Hopefully, a proof of existence and uniqueness that would apply to much more general geometries will be produced. We hope also that the global topological restriction on the vanishing of the first cohomology will be investigated.

The gauge conditions (9) have already proved valuable in permitting a new proof of positive energy and show promise for applications to the initial value problem of general relativity. Only future explorations can reveal their true worth.

ACKNOWLEDGMENT

Partially supported by the National Science Council of the Republic of China under the Contract No. NSC77-0208-M008-20.

[1] T. W. B. Kibble, J. Math. Phys. **4**, 1433 (1963).
[2] J. M. Nester, to be published in Int. J. Mod. Phys. A.
[3] R. Geroch, J. Math. Phys. **9**, 1739 (1968).
[4] E. Stiefel, Commun. Math. Helv. **8**, 3 (1936).
[5] J. A. Schouten, *Ricci Calculus* (Springer, Berlin, 1954), Second Ed.
[6] E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[7] Y. Choquet-Bruhat and J. W. York, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. I, pp. 99–172.
[8] Y. Choquet-Bruhat and D. Christodoulou, Acta Math. **146**, 129 (1981).

Multivariable biorthogonal Hahn polynomials

M. V. Tratnik
Center for Nonlinear Studies and Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico, 87545

(Received 17 March 1988; accepted for publication 21 September 1988)

A multivariable biorthogonal generalization of the discrete Hahn polynomials, a $p+1$ complex parameter family, where p is the number of variables, is presented. It is shown that the polynomials are orthogonal with respect to subspaces of lower degree and biorthogonal within a given subspace. These properties are over the discrete simplex $0 \leqslant x_1 + x_2 + \cdots + x_p \leqslant \Delta$, where $x_1, x_2, \ldots, x_p$ and Δ are non-negative integers. Some further properties of the closely related multivariable continuous Hahn polynomials are also discussed.

I. INTRODUCTION

The discretized Jacobi polynomials that commonly bear the name of Hahn[1] were actually introduced much earlier by Tchebychef.[2,3] Since their reintroduction by Hahn the polynomials have found applications in many areas of theoretical and mathematical physics, including genetics, group representation theory, computational physics and techniques, as well as others. The main utility of the polynomials lies in that they satisfy an orthogonality relation over a discrete set of points. Other special properties are discussed by Weber and Erdelyi,[4] Karlin and McGregor,[5] Bartko,[6] Levit,[7] Lee,[8] Wilson,[9] Gasper,[10] Koornwinder,[11] and Nikiforov *et al.*[12] Furthermore, Karlin and McGregor[13] have presented a multivariable *orthogonal* (as opposed to biorthogonal) generalization of the Hahn polynomials in the context of linear growth models. In this paper we present a distinct and not simply related family, the multivariable *biorthogonal* Hahn polynomials: These are a $(p+1)$ complex parameter family, where p denotes the number of variables. It is shown that the polynomials are orthogonal with respect to subspaces of lower degree and biorthogonal within a given subspace. These properties are defined on a discrete set of points $x_1, x_2, \ldots, x_p$ which take non-negative integer values on the discrete simplex $0 \leqslant x_1 + x_2 + \cdots + x_p \leqslant \Delta$, where Δ denotes the "discrete parameter," also a non-negative integer. (Our notation for the discrete parameter Δ differs from the customary use of N, which in this paper will denote the degree of a multivariable polynomial.) In the special case $p=1$ these two families both reduce to the familiar single-variable Hahn polynomials.

Before we introduce the discrete polynomials we discuss some further properties of a closely related continuous family. Recently, Atakishiyev and Suslov[14] constructed a continuous analog of the Hahn polynomials by analytically continuing the discrete variable and parameter into the complex plane; these were extended soon after by Askey.[15] In a previous paper,[16] the present author introduced a multivariable biorthogonal generalization of the continuous Hahn polynomials. In Sec. II we deduce some symmetries of this continuous family and show that in a special case they are pure real. These polynomials are conveniently expressed in terms of multiple Gaussian hypergeometric series[17] as follows:

$$F \equiv F^{1:2;\ldots;2}_{1:1;\ldots;1},$$

$$P^{a,b,c,d}_{n_1 n_2 \cdots n_p}(x_1,x_2,\ldots,x_p) = i^N (A+d)_N \left[\prod_{k=1}^{p} \frac{(a_k+b_k)_{n_k}}{n_k!}\right] F\begin{pmatrix} N+A+B+c+d-1: -n_1, a_1+ix_1; \ldots; -n_p, a_p+ix_p \\ A+d: a_1+b_1; \ldots; a_p+b_p \end{pmatrix} \tag{1.1}$$

and also their counterparts

$$Q^{a,b,c,d}_{n_1 n_2 \cdots n_p}(x_1,x_2,\ldots,x_p) = (-i)^N (c+iX)_N \left[\prod_{k=1}^{p} \frac{(a_k+b_k)_{n_k}}{n_k!}\right] \times F\begin{pmatrix} -N-c-d+1: -n_1, b_1-ix_1; \ldots; -n_p, b_p-ix_p \\ -N-c-iX+1: a_1+b_1; \ldots; a_p+b_p \end{pmatrix}, \tag{1.2}$$

where F is defined as

$$F\begin{pmatrix} \eta: \xi_1, \zeta_1; \ldots; \xi_p, \zeta_p; \\ \gamma: \delta_1; \ldots; \delta_p \end{pmatrix} z_1 \cdots z_p = \sum_{\{j_k\}} \frac{(\eta)_J}{(\gamma)_J} \prod_{k=1}^{p} \frac{(\xi_k)_{j_k} (\zeta_k)_{j_k}}{(\delta_k)_{j_k} j_k!} z_k^{j_k}, \tag{1.3}$$

 0022-2488/89/030627-08$02.50

and we are employing the familiar Pochhammer symbol $(\alpha)_n \equiv \Gamma(n+\alpha)/\Gamma(\alpha)$. Here, $\{j_k\}$ denotes the summation indices $j_1, j_2,..., j_p$ that run over all non-negative integers and we are using the following shorthand notation:

$$X \equiv \sum_{k=1}^{p} x_k, \quad N \equiv \sum_{k=1}^{p} n_k, \quad J \equiv \sum_{k=1}^{p} j_k, \quad A \equiv \sum_{k=1}^{p} a_k, \quad B \equiv \sum_{k=1}^{p} b_k, \tag{1.4}$$

and the convention that $1/\Gamma(-n) = 0$, $n = 0,1,2,...$. If the arguments of F are unspecified, it is to be understood that they are unity: $z_1 = z_2 = \cdots = z_p = 1$.

These polynomials form a biorthogonal system with the weight function

$$w^{a,b,c,d}(x_1,x_2,...,x_p) = \left[\prod_{k=1}^{p} \Gamma(a_k + ix_k)\Gamma(b_k - ix_k)\right] \times \Gamma(c+iX)\Gamma(d-iX), \tag{1.5}$$

where the $(2p+2)$ complex parameters a_k, b_k, c, d, $k = 1,2,...,p$, identify a particular family of polynomials and their weight function, whereas the set of p non-negative integers $n_1,n_2,...,n_p$ label the members of a given family. The degree of a polynomial is simply given by N and, as usual, i denotes the square root of -1. The discrete parameter Δ should not be confused with N, and when no ambiguity arises we simply write $P_n(x)$, $Q_n(x)$, and $w(x)$ for the polynomials and weight function, respectively. The polynomial $P_n(x)$ is of degree n_k in the variable x_k, $Q_n(x)$ is of degree N in x_k, and both polynomials are of total degree N.

The family $P_n(x)$ is orthogonal with respect to the degree, that is,

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, P_n(x) P_m(x) w(x) = 0, \quad \text{if } N \neq M, \tag{1.6}$$

and similarly for $Q_n(x)$,

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, Q_n(x) Q_m(x) w(x) = 0, \quad \text{if } N \neq M, \tag{1.7}$$

while in general these two families are biorthogonal:

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, P_n(x) Q_m(x) w(x) = h_n \prod_{k=1}^{p} \delta_{n_k m_k}, \tag{1.8}$$

where h_n is a normalization constant. In Sec. II we obtain alternate representations for these polynomials, from which we deduce some symmetries. In subsequent sections we discuss the analog of relations (1.6)–(1.8) for the discrete family of Hahn polynomials.

II. SYMMETRIES OF THE CONTINUOUS HAHN POLYNOMIALS

We begin by recalling several identities that will be of use. The first,

$$\Gamma(\xi+1)/\Gamma(\xi \pm j + 1) = (-1)^j [\Gamma(-\xi \mp j)/\Gamma(-\xi)], \tag{2.1}$$

where ξ is some constant and j is an integer, can be verified by inspection. The second is the Chu–Vandermonde formula[18]

$$\sum_{j=0}^{n} \binom{n}{j} \frac{\Gamma(\eta)}{\Gamma(\eta - j)} \frac{\Gamma(\xi)}{\Gamma(\xi - n + j)} = \frac{\Gamma(\eta + \xi - 1)}{\Gamma(\eta + \xi - n - 1)}, \tag{2.2}$$

while the third is the multiple summation theorem[19]

$$F_D^{(p)}(\eta,\xi_1,...,\xi_p;\gamma;1,...,1) = \Gamma(\gamma)\Gamma(\gamma - \eta - \xi_1 - \cdots - \xi_p)/\Gamma(\gamma - \eta) \times \Gamma(\gamma - \xi_1 - \cdots - \xi_p), \tag{2.3}$$

where $F_D^{(p)}$ is the p-variable Lauricella hypergeometric series defined as

$$F_D^{(p)}(\eta,\xi_1,...,\xi_p;\gamma;x_1,...,x_p) = \sum_{\{j_k\}} \left[\prod_{k=1}^{p} (\xi_k)_{j_k} \frac{x_k^{j_k}}{j_k!}\right] \frac{(\eta)_J}{(\gamma)_J}, \tag{2.4}$$

and the $\{j_k\}$ sum is over all non-negative integers.

To find an alternate representation for the continuous Hahn polynomials $P_n(x)$ we begin by setting $\eta = 1 - N - B - c$; $\xi_1 = -j_1,...,\xi_p = -j_p$; and $\gamma = A + d$ in (2.3). Then we use (2.1) to obtain

$$\frac{\Gamma(N+J+A+B+c+d-1)}{\Gamma(N+A+B+c+d-1)} = \sum_{\{l_k\}} \left[\prod_{k=1}^{p} \binom{j_k}{l_k}\right] \frac{\Gamma(J+A+d)}{\Gamma(L+A+d)} \times \frac{\Gamma(N+B+c)}{\Gamma(N-L+B+c)}, \tag{2.5}$$

which, when substituted into (1.1), gives

$$P_n(x) = i^N \sum_{\{j_k\}} \sum_{\{l_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k}\binom{j_k}{l_k} \frac{\Gamma(n_k + a_k + b_k)}{n_k!\Gamma(j_k + a_k + b_k)} \times \frac{\Gamma(j_k + a_k + ix_k)}{\Gamma(a_k + ix_k)}\right] \times \frac{\Gamma(N+A+d)}{\Gamma(L+A+d)} \frac{\Gamma(N+B+c)}{\Gamma(N-L+B+c)} (-1)^J. \tag{2.6}$$

If we then interchange the order of the sums and translate the summation indices $j_k \to j_k + l_k$, we obtain

$$P_n(x) = i^N \sum_{\{l_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{l_k} \frac{1}{n_k!}\right] \times \frac{\Gamma(N+A+d)}{\Gamma(L+A+d)} \frac{\Gamma(N+B+c)}{\Gamma(N-L+B+c)} (-1)^L \times \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k - l_k}{j_k} \frac{\Gamma(n_k + a_k + b_k)}{\Gamma(j_k + l_k + a_k + b_k)} \times \frac{\Gamma(j_k + l_k + a_k + ix_k)}{\Gamma(a_k + ix_k)}\right] (-1)^J. \tag{2.7}$$

Using identity (2.1) to write

$$\frac{\Gamma(j_k+l_k+a_k+ix_k)}{\Gamma(a_k+ix_k)}$$
$$=(-1)^{j_k+l_k}\frac{\Gamma(-a_k-ix_k+1)}{\Gamma(-a_k-ix_k+1-j_k-l_k)}$$
$$=(-1)^{j_k+l_k}\frac{\Gamma(-a_k-ix_k+1)}{\Gamma(-a_k-ix_k+1-l_k)}\times\frac{\Gamma(-a_k-ix_k+1-l_k)}{\Gamma(-a_k-ix_k+1-j_k-l_k)}$$
$$=(-1)^{j_k}\frac{\Gamma(l_k+a_k+ix_k)}{\Gamma(a_k+ix_k)}\times\frac{\Gamma(-a_k-ix_k+1-l_k)}{\Gamma(-a_k-ix_k+1-l_k-j_k)} \tag{2.8}$$

and substituting into (2.7) yields

$$P_n(x)=i^N\sum_{\{l_k\}}\left[\prod_{k=1}^p\binom{n_k}{l_k}\frac{\Gamma(l_k+a_k+ix_k)}{n_k!\Gamma(a_k+ix_k)}\right]\times\frac{\Gamma(N+A+d)}{\Gamma(L+A+d)}\frac{\Gamma(N+B+c)}{\Gamma(N-L+B+c)}(-1)^L\times\sum_{\{j_k\}}\left[\prod_{k=1}^p\binom{n_k-l_k}{j_k}\frac{\Gamma(n_k+a_k+b_k)}{\Gamma(j_k+l_k+a_k+b_k)}\times\frac{\Gamma(-a_k-ix_k+1-l_k)}{\Gamma(-a_k-ix_k+1-l_k-j_k)}\right]. \tag{2.9}$$

Then, performing the $\{j_k\}$ summations by using the Chu–Vandermonde theorem (2.2) gives the alternate representation

$$P_n(x)=i^N(A+d)_N\left[\prod_{k=1}^p\frac{(b_k-ix_k)_{n_k}}{n_k!}\right]F\left(\begin{matrix}-N-B-c+1:-n_1,a_1+ix_1;\ldots;-n_p,a_p+ix_p\\A+d:-n_1-b_1+ix_1+1;\ldots;-n_p-b_p+ix_p+1\end{matrix}\right). \tag{2.10}$$

For the biorthogonal counterparts $Q_n(x)$ we set $\eta=d-iX$, $\xi_1=-j_1,\ldots,\xi_p=-j_p$, and $\gamma=1-N-c-iX$ in (2.3) to obtain, in place of (2.5),

$$\frac{\Gamma(N+c+d)}{\Gamma(N-J+c+d)}=\sum_{\{l_k\}}\left[\prod_{k=1}^p\binom{j_k}{l_k}\right]\frac{\Gamma(N-L+c+iX)}{\Gamma(N-J+c+iX)}\frac{\Gamma(L+d-iX)}{\Gamma(d-iX)}. \tag{2.11}$$

Then, proceeding in a completely analogous manner we obtain the following alternate representation:

$$Q_n(x)=(-i)^N(c+iX)_N\left[\prod_{k=1}^p\frac{(a_k+ix_k)_{n_k}}{n_k!}\right]\times F\left(\begin{matrix}d-iX:-n_1,b_1-ix_1;\ldots;-n_p,b_p-ix_p\\-N-c-iX+1:-n_1-a_1-ix_1+1;\ldots;-n_p-a_p-ix_p+1\end{matrix}\right). \tag{2.12}$$

From (2.10) and (2.12) we immediately deduce the symmetries

$$\begin{aligned}P^{a,b,c,d}_{n_1n_2\cdots n_p}(x_1,x_2,\ldots,x_p)&=(-1)^N P^{b,a,d,c}_{n_1n_2\cdots n_p}(-x_1,-x_2,\ldots,-x_p),\\Q^{a,b,c,d}_{n_1n_2\cdots n_p}(x_1,x_2,\ldots,x_p)&=(-1)^N Q^{b,a,d,c}_{n_1n_2\cdots n_p}(-x_1,-x_2,\ldots,-x_p),\end{aligned} \tag{2.13}$$

which might have been anticipated since the weight function (1.5) is invariant under the transformation $a_k\leftrightarrow b_k$, $c\leftrightarrow d$, $x_k\to -x_k$, $k=1,2,\ldots,p$. From (2.13) we see that if $a_k=b_k$, $k=1,2,\ldots,p$, and $c=d$, then the polynomials have parity:

$$P_n(-x)=(-1)^N P_n(x),\quad Q_n(-x)=(-1)^N Q_n(x). \tag{2.14}$$

We also find from representations (2.10) and (2.12) that if $a_k=b_k^*$, $k=1,2,\ldots,p$, and $c=d^*$, then

$$P_n(x)=P_n^*(x^*),\quad Q_n(x)=Q_n^*(x^*); \tag{2.15}$$

thus for this special case the polynomials have pure real coefficients. If in addition $\mathrm{Re}(a_k)$, $\mathrm{Re}(b_k)$, $\mathrm{Re}(c)$, $\mathrm{Re}(d)>0$, $k=1,2,\ldots,p$, then the variables x_k and the weight function are also pure real.[16]

If we define a finite difference operator δ_k as

$$\delta_k f(x_1\cdots x_p)\equiv f(x_1\cdots x_k+i/2\cdots x_p)-f(x_1\cdots x_k-i/2\cdots x_p), \tag{2.16}$$

then it is a relatively simple matter to deduce Rodrigues-type formulas from representations (2.10) and (2.12). For the first family $P_n(x)$ we find the following unconventional form:

$$P_n(x)=i^N\frac{\Gamma(N+A+d)\Gamma(N+B+c)}{\Pi_{k=1}^p n_k!\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)}\left[\prod_{k=1}^p\delta_k^{n_k}\right]\times\left[\frac{\Pi_{k=1}^p\Gamma(n_k/2+a_k+ix_k)\Gamma(n_k/2+b_k-ix_k)}{\Gamma(N/2+A+d+iX-iY)\Gamma(N/2+B+c-iX+iY)}\right], \tag{2.17}$$

where Y is held fixed during the differencing and is set equal to X afterwards. For the second family we obtain the more conventional Rodrigues formula

$$Q_n(x) = i^N \left[\Gamma(c+iX)\Gamma(d-iX) \prod_{k=1}^{p} n_k!\Gamma(a_k+ix_k)\Gamma(b_k-ix_k)\right]^{-1} \left[\prod_{k=1}^{p} \delta_k^{n_k}\right]$$
$$\times\left[\Gamma\left(\frac{N}{2}+c+iX\right)\Gamma\left(\frac{N}{2}+d-iX\right) \prod_{k=1}^{p} \Gamma\left(\frac{n_k}{2}+a_k+ix_k\right)\Gamma\left(\frac{n_k}{2}+b_k-ix_k\right)\right]. \tag{2.18}$$

Applying the symmetries (2.13) to the original representations (1.1) and (1.2) gives a further representation for each family, which we do not write down here.

In the special case of a single variable $p=1$, Eqs. (1.6) and (1.7) imply that both $P_n(x)$ and $Q_n(x)$ form an orthogonal system with the same weight function. These two families must then be identical apart from a possible change in normalization. This can be demonstrated explicitly by using the transformation formula

$${}_3F_2\left(\begin{matrix}-n,\eta,\xi\\ \gamma,\delta\end{matrix};1\right) = \frac{\Gamma(n+\delta-\xi)\Gamma(\delta)}{\Gamma(n+\delta)\Gamma(\delta-\xi)}\, {}_3F_2\left(\begin{matrix}-n,\gamma-\eta,\xi\\ \gamma,1+\xi-\delta-n\end{matrix};1\right) \tag{2.19}$$

derived from Thomae[20] [or see Bailey[18] (p. 21)]. Setting $\eta = n+a+b+c+d-1$, $\xi = a+ix$, $\gamma = a+b$, and $\delta = a+d$ in (2.19) shows that $P_n(x) = Q_n(x)$ in the special case of a single variable.

The weight function for a single variable has additional symmetries compared to the multivariable case: It is invariant under the interchange of a and c and also under the interchange of b and d. Then the single-variable continuous Hahn polynomials are also invariant under these transformations apart from a possible change in normalization. In fact, from representation (2.12) it is clear that the normalization does not change and thus

$$P_n^{a,b,c,d}(x) = P_n^{c,b,a,d}(x) = P_n^{a,d,c,b}(x) \quad (p=1). \tag{2.20}$$

III. MULTIVARIABLE DISCRETE HAHN POLYNOMIALS

In this section we present the multivariable biorthogonal generalization of the discrete Hahn polynomials. These are a natural discrete analog of not only the continuous Hahn polynomials, but also of the conformal polynomials of Lam and Tratnik.[21]

The weight function associated with this discrete family is given by

$$w(x) = \left[\prod_{k=1}^{p} \frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(x_k+1)}\right] \frac{\Gamma(\Delta-X+\beta+1)}{\Gamma(\Delta-X+1)}, \tag{3.1}$$

while corresponding to the polynomials $P_n(x)$ we introduce

$$H_{n_1n_2\cdots n_p}^{\alpha_1\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta) = F\left(\begin{matrix}N+\alpha+\beta+p{:}\,-n_1,-x_1;\ldots;-n_p,-x_p\\ -\Delta{:}\,\alpha_1+1;\ldots;\alpha_p+1\end{matrix}\right), \tag{3.2}$$

which also have the equivalent representation

$$H_{n_1n_2\cdots n_p}^{\alpha_1\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta) = \frac{(-1)^N(\Delta+\alpha+\beta+p+1)_N}{(-\Delta)_N}$$
$$\times F\left(\begin{matrix}N+\alpha+\beta+p{:}\,-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\ \Delta+\alpha+\beta+p+1{:}\;\alpha_1+1;\ldots;\alpha_p+1\end{matrix}\right). \tag{3.3}$$

Expressions (3.2) and (3.3) are the discrete analogs of (1.1) and (1.1) transformed by the symmetry (2.13); if we proceed as in (2.5)–(2.10), where we obtained an alternate representation for $P_n(x)$, we find the following corresponding expression for the discrete family:

$$H_{n_1n_2\cdots n_p}^{\alpha_1\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta) = \left[\prod_{k=1}^{p} \frac{(x_k+\alpha_k+1)_{n_k}}{(\alpha_k+1)_{n_k}}\right] F\left(\begin{matrix}-\Delta-N-\alpha-\beta-p{:}\,-n_1,-x_1;\ldots;-n_p,-x_p\\ -\Delta{:}\,-n_1-x_1-\alpha_1;\ldots;-n_p-x_p-\alpha_p\end{matrix}\right). \tag{3.4}$$

Defining a finite difference operator D_k as

$$D_k f(x_1\cdots x_p) \equiv f(x_1\cdots x_k+1\cdots x_p) - f(x_1\cdots x_k\cdots x_p), \tag{3.5}$$

which one can show satisfies the identity

$$D_k^{n_k} f(x_1\cdots x_p) = \sum_{j_k=0}^{n_k} \binom{n_k}{j_k} (-1)^{j_k} f(x_1\cdots x_k+n_k-j_k\cdots x_p), \tag{3.6}$$

it is a simple matter to deduce from (3.4) a discrete analog of the Rodrigues formula (2.17):

$$H_{n_1,n_2\cdots n_p}^{\alpha_1,\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta)=\frac{\Gamma(\Delta+N+\alpha+\beta+p+1)}{\Gamma(\Delta+1)}\left[\prod_{k=1}^{p}\frac{\Gamma(\alpha_k+1)}{\Gamma(n_k+\alpha_k+1)}\frac{\Gamma(x_k+1)}{\Gamma(x_k+\alpha_k+1)}D_k^{n_k}\right]$$
$$\times\left[\frac{\Gamma(\Delta-N+X-Y+1)}{\Gamma(\Delta+X-Y+\alpha+\beta+p+1)}\prod_{k=1}^{p}\frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(x_k-n_k+1)}\right], \tag{3.7}$$

where, as before, Y is kept fixed during the differencing and is set equal to X afterward. Corresponding to representations (1.2) and (1.2) transformed by the symmetry (2.13), we introduce the following discrete analog of the biorthogonal counterparts:

$$G_{n_1,n_2\cdots n_p}^{\alpha_1,\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta)=\frac{(-1)^N(\Delta-X+\beta+1)_N}{(-\Delta)_N}F\begin{pmatrix}-N-\beta:-n_1,x_1+\alpha_1+1;\ldots;-n_p,x_p+\alpha_p+1\\-\Delta-N-\beta+X:\ \alpha_1+1;\ldots;\alpha_p+1\end{pmatrix}, \tag{3.8}$$

with the equivalent representation

$$G_{n_1,n_2\cdots n_p}^{\alpha_1,\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta)=\frac{(-\Delta+X)_N}{(-\Delta)_N}F\begin{pmatrix}-N-\beta:-n_1,-x_1;\ldots;-n_p,-x_p\\\Delta-X-N+1:\ \alpha_1+1;\ldots;\alpha_p+1\end{pmatrix} \tag{3.9}$$

and the analog of (2.12),

$$G_{n_1,n_2\cdots n_p}^{\alpha_1,\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta)=\frac{(-\Delta+X)_N}{(-\Delta)_N}\left[\prod_{k=1}^{p}\frac{(x_k+\alpha_k+1)_{n_k}}{(\alpha_k+1)_{n_k}}\right]$$
$$\times F\begin{pmatrix}\Delta-X+\beta+1:-n_1,-x_1;\ldots;-n_p,-x_p\\\Delta-X-N+1:-n_1-x_1-\alpha_1;\ldots;-n_p-x_p-\alpha_p\end{pmatrix}. \tag{3.10}$$

Then, from (3.6) and (3.10) one deduces the following Rodrigues formula:

$$G_{n_1,n_2\cdots n_p}^{\alpha_1,\alpha_2\cdots\alpha_p,\beta}(x_1,x_2,\ldots,x_p,\Delta)$$
$$=\frac{(\Delta-N)!}{\Delta!}\frac{\Gamma(\Delta-X+1)}{\Gamma(\Delta-X+\beta+1)}$$
$$\times\left[\prod_{k=1}^{p}\frac{\Gamma(\alpha_k+1)}{\Gamma(n_k+\alpha_k+1)}\frac{\Gamma(x_k+1)}{\Gamma(x_k+\alpha_k+1)}D_k^{n_k}\right]$$
$$\times\left[\frac{\Gamma(N+\Delta-X+\beta+1)}{\Gamma(\Delta-X+1)}\right.$$
$$\left.\times\prod_{k=1}^{p}\frac{\Gamma(x_k+\alpha_k+1)}{\Gamma(x_k-n_k+1)}\right], \tag{3.11}$$

the discrete analog of (2.18). These are polynomials of the p variables $x_1,x_2,\ldots,x_p$ that take non-negative integer values on the discrete simplex $0\leqslant x_1+x_2+\cdots+x_p\leqslant\Delta$, where Δ is also a non-negative integer. The discrete parameter Δ, along with the $p+1$ complex parameters $\alpha_1,\alpha_2,\ldots,\alpha_p,\beta$, identify a particular family of polynomials and their weight function. The set of p non-negative integers $n_1,n_2,\ldots,n_p$ label the members of a given family that are restricted to $0\leqslant n_1+n_2+\cdots+n_p\leqslant\Delta$. We are using the shorthand notation N, J, and X as in (1.4) and in addition we have defined $\alpha\equiv\Sigma_{k=1}^{p}\alpha_k$. The $p+1$ complex parameters $\alpha_1,\alpha_2,\ldots,\alpha_p,\beta$ are excluded from taking negative integer values, but are otherwise arbitrary. We shall simply write $H_n(x)$, $G_n(x)$, and $w(x)$ for the polynomials and weight function, respectively. The polynomials $H_n(x)$ and $G_n(x)$ are both of degree N and as already mentioned, $N\leqslant\Delta$.

The continuous Hahn polynomials are related to the discrete family through the relations

$$P_n^{a,b,c,d}(x)$$
$$=i^N(A+d)_N\left[\prod_{k=1}^{p}\frac{(a_k+b_k)_{n_k}}{n_k!}\right]$$
$$\times H_n^{a+b-1,c+d-1}(-a-ix,-A-d),$$
$$Q_n^{a,b,c,d}(x)$$
$$=i^N(A+d)_N\left[\prod_{k=1}^{p}\frac{(a_k+b_k)_{n_k}}{n_k!}\right]$$
$$\times G_n^{a+b-1,c+d-1}(-a-ix,-A-d), \tag{3.12}$$

where in the rhs we are using shorthand notation to denote the transformations $\alpha_k\to a_k+b_k-1$, $\beta\to c+d-1$, $x_k\to-a_k-ix_k$, and $\Delta\to A-d$ for $k=1,2,\ldots,p$. The equivalence of representations (3.2)–(3.4) follows from (3.12), the equality of (1.1) and (2.10), and the symmetry (2.13); we have a similar equivalence for representations (3.8)–(3.10).

One can also recover the conformal polynomials[16,21] from the discrete Hahn family by a limit transition. If we simply redefine the parameters $\alpha_k=\mu_k-1$, $\beta=\mu_{p+1}-1$, and then replace x_k by Δx_k and take the limit $\Delta\to\infty$ one finds

$$\lim_{\Delta\to\infty}H_n(\Delta x)$$
$$=\frac{(-1)^N(N+\mu_{p+1})}{(2N+\mu-1)\Gamma(N+\mu-1)}$$
$$\times\left[\prod_{k=1}^{p}n_k!\Gamma(\mu_k)\right]D_n^{\mu}(x) \tag{3.13}$$

and

$$\lim_{\Delta\to\infty}G_n(\Delta x)=\left[\prod_{k=1}^{p}\frac{(-1)^{n_k}\Gamma(\mu_k)}{\Gamma(n_k+\mu_k)}\right]C_n^{\mu}(x), \tag{3.14}$$

where $C_n^\mu(x)$ and $D_n^\mu(x)$ are the conformal and dual conformal polynomials and the weight function is obtained in the same limit if one also multiplies by $\Delta^{-\alpha-\beta}$:

$$\lim_{\Delta\to\infty}\Delta^{-\alpha-\beta}w(\Delta x)=\left[\prod_{k=1}^{p}x_k^{\mu_k-1}\right](1-X)^{\mu_{p+1}-1}. \tag{3.15}$$

IV. ORTHOGONALITY PROPERTIES

First we demonstrate that the inner product of $H_n(x)$ with another polynomial of the same family $H_m(x)$ vanishes if $N\neq M$. From representations (3.2) and (3.3) and the expression for the weight function (3.1) we have

$$\begin{aligned}
&\sum_{\{x_k\}}H_n(x)H_m(x)w(x)\\
&=\sum_{\{j_k\}}\sum_{\{l_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\binom{m_k}{l_k}\frac{\Gamma(\alpha_k+1)}{\Gamma(j_k+\alpha_k+1)}\right.\\
&\quad\left.\times\frac{\Gamma(\alpha_k+1)}{\Gamma(l_k+\alpha_k+1)}\right]\frac{\Gamma(N+J+\alpha+\beta+p)}{\Gamma(N+\alpha+\beta+p)}\\
&\quad\times\frac{\Gamma(\Delta+N+\alpha+\beta+p+1)}{\Gamma(\Delta+J+\alpha+\beta+p+1)}\\
&\quad\times\frac{\Gamma(M+L+\alpha+\beta+p)}{\Gamma(M+\alpha+\beta+p)}\\
&\quad\times\frac{(\Delta-L)!}{\Delta!}\frac{(\Delta-N)!}{\Delta!}(-1)^{J+L}\\
&\quad\times\sum_{\{x_k\}}\left[\prod_{k=1}^{p}\frac{\Gamma(j_k+x_k+\alpha_k+1)}{(x_k-l_k)!}\right]\\
&\quad\times\frac{\Gamma(\Delta-X+\beta+1)}{\Gamma(\Delta-X+1)},
\end{aligned}\tag{4.1}$$

where the $\{x_k\}$ sum is over the discrete simplex, as discussed earlier. Equation (4.1) can be evaluated from (2.3) by setting $\eta=-\Delta+L$, $\xi_k=j_k+l_k+\alpha_k+1$, and $\gamma=-\Delta+L-\beta$, in which case (4.1) becomes, apart from a multiplicative constant,

$$\begin{aligned}
&\sum_{\{j_k\}}\sum_{\{l_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\binom{m_k}{l_k}\right.\\
&\quad\left.\times\frac{\Gamma(j_k+l_k+\alpha_k+1)}{\Gamma(j_k+\alpha_k+1)\Gamma(l_k+\alpha_k+1)}\right]\\
&\quad\times\frac{\Gamma(N+J+\alpha+\beta+p)\Gamma(M+L+\alpha+\beta+p)}{\Gamma(J+L+\alpha+\beta+p+1)}\\
&\quad\times(-1)^{J+L}.
\end{aligned}\tag{4.2}$$

Without loss of generality we assume $N>M$, so that also $N>L$; then, we concentrate on the $\{j_k\}$ sum:

$$\begin{aligned}
&\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\frac{\Gamma(l_k+j_k+\alpha_k+1)}{\Gamma(j_k+\alpha_k+1)}\right]\\
&\quad\times\frac{\Gamma(N+J+\alpha+\beta+p)}{\Gamma(L+J+\alpha+\beta+p+1)}(-1)^J,
\end{aligned}\tag{4.3}$$

which apart from a multiplicative constant is expression (2.5).[16] As we discuss in Ref. 16, (4.3) can be written in the form

$$\begin{aligned}
&\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial z_k}\right)_{z_k=1}^{l_k}z_k^{l_k+\alpha_k}\right]\\
&\quad\times\sum_{q=0}^{N-L-1}\xi_q\left(-\sum_{k=1}^{p}z_k\frac{\partial}{\partial z_k}\right)^q\prod_{k=1}^{p}(1-z_k)^{n_k},
\end{aligned}\tag{4.4}$$

where ξ_q are some constants and we recall that $N>L$. To reproduce expression (4.3) one uses the binomial theorem to expand each factor in the product $\Pi_{k=1}^p(1-z_k)^{n_k}$. The highest order derivative acting on $\Pi_{k=1}^p(1-z_k)^{n_k}$ is of order $N-1$, so that at least one factor of $(1-z_k)$, for some k, will survive in every term after the differentiations and then will vanish upon setting $z_k=1$. Thus the sum (4.3) vanishes whenever $N>M$ and hence

$$\sum_{\{x_k\}}H_n(x)H_m(x)w(x)=0,\quad\text{if } N\neq M. \tag{4.5}$$

In a completely analogous manner we find the equivalent result for the biorthogonal family:

$$\sum_{\{x_k\}}G_n(x)G_m(x)w(x)=0,\quad\text{if } N\neq M, \tag{4.6}$$

so that $H_n(x)$ and $G_n(x)$ are orthogonal with respect to the subspaces labeled by N. However, this says nothing of polynomials of the same degree. We demonstrate that these two families form a biorthogonal system.

Using representations (3.3), (3.9), and the weight function (3.1), we obtain

$$\begin{aligned}
&\sum_{\{x_k\}}H_n(x)G_m(x)w(x)\\
&=\frac{(\Delta-N)!}{\Delta!}\frac{(\Delta-M)!}{\Delta!}\frac{\Gamma(\Delta+N+\alpha+\beta+p+1)}{\Gamma(N+\alpha+\beta+p)}\\
&\quad\times\sum_{\{j_k\}}\sum_{\{l_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\binom{m_k}{l_k}\right.\\
&\quad\left.\times\frac{\Gamma(\alpha_k+1)}{\Gamma(j_k+\alpha_k+1)}\frac{\Gamma(\alpha_k+1)}{\Gamma(l_k+\alpha_k+1)}\right]\\
&\quad\times\frac{\Gamma(N+J+\alpha+\beta+p)}{\Gamma(\Delta+J+\alpha+\beta+p+1)}\\
&\quad\times\frac{\Gamma(M+\beta+1)}{\Gamma(M-L+\beta+1)}(-1)^{J+L}\\
&\quad\times\sum_{\{x_k\}}\frac{\Gamma(j_k+x_k+\alpha_k+1)}{\Gamma(x_k-l_k+1)}\\
&\quad\times\frac{\Gamma(\Delta-X+\beta+1)}{\Gamma(\Delta-X-M+L+1)},
\end{aligned}\tag{4.7}$$

and if we set $\eta=-\Delta+M$, $\xi_k=j_k+l_k+\alpha_k+1$, and $\gamma=-\Delta+L-\beta$ in (2.3), the above $\{x_k\}$ sum is obtained, leading to the expression

$$\sum_{\{x_k\}} H_n(x)G_m(x)w(x)$$
$$= \left[\prod_{k=1}^{p} \Gamma(\alpha_k+1)\right]^2 \frac{(\Delta-N)!}{\Delta!\Delta!}$$
$$\times \frac{\Gamma(M+\beta+1)\Gamma(\Delta+N+\alpha+\beta+p+1)}{\Gamma(N+\alpha+\beta+p)}$$
$$\times \sum_{\{j_k\}}\sum_{\{l_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k}\binom{m_k}{l_k}\right.$$
$$\left.\times \frac{(-1)^{j_k+l_k}\Gamma(j_k+l_k+\alpha_k+1)}{\Gamma(j_k+\alpha_k+1)\Gamma(l_k+\alpha_k+1)}\right]$$
$$\times \frac{\Gamma(N+J+\alpha+\beta+p)}{\Gamma(M+J+\alpha+\beta+p+1)}. \tag{4.8}$$

Equation (4.8) is essentially equivalent to the analogous expression for the continuous family, so the biorthogonality proof in this discrete case follows in a similar manner.

First we use identity (2.1) to write

$$(-1)^{l_k}\Gamma(j_k+l_k+\alpha_k+1)$$
$$= \Gamma(j_k+\alpha_k+1)\Gamma(-j_k-\alpha_k)$$
$$\times [\Gamma(-j_k-l_k-\alpha_k)]^{-1}. \tag{4.9}$$

Then, the $\{l_k\}$ summations are performed using the Chu–Vandermonde theorem (2.2), which gives

$$\sum_{\{x_k\}} H_n(x)G_m(x)w(x)$$
$$= \left[\prod_{k=1}^{p} \frac{\Gamma(\alpha_k+1)\Gamma(\alpha_k+1)}{\Gamma(m_k+\alpha_k+1)}\right] \frac{(\Delta-N)!}{\Delta!\Delta!}$$
$$\times \frac{\Gamma(M+\beta+1)\Gamma(\Delta+N+\alpha+\beta+p+1)}{\Gamma(N+\alpha+\beta+p)}$$
$$\times (-1)^M \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \frac{n_k!}{(n_k-j_k)!(j_k-m_k)!}\right]$$
$$\times \frac{\Gamma(N+J+\alpha+\beta+p)}{\Gamma(M+J+\alpha+\beta+p+1)}(-1)^J. \tag{4.10}$$

If now we redefine the indices $j_k \to j_k + m_k$ and then use identity (2.1) to write

$$\frac{n_k!}{(n_k-m_k-j_k)!}$$
$$= \frac{n_k!}{(n_k-m_k)!}\frac{\Gamma(-n_k+m_k+j_k)}{\Gamma(-n_k+m_k)}(-1)^{j_k}, \tag{4.11}$$

the inner product becomes

$$\sum_{\{x_k\}} H_n(x)G_m(x)w(x)$$
$$= \left[\prod_{k=1}^{p} \frac{\Gamma(\alpha_k+1)\Gamma(\alpha_k+1)n_k!}{\Gamma(m_k+\alpha_k+1)(n_k-m_k)!}\right] \frac{(\Delta-N)!}{\Delta!\Delta!}$$
$$\times \frac{\Gamma(M+\beta+1)\Gamma(\Delta+N+\alpha+\beta+p+1)}{\Gamma(N+\alpha+\beta+p)}$$
$$\times \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \frac{\Gamma(-n_k+m_k+j_k)}{\Gamma(-n_k+m_k)j_k!}\right]$$
$$\times \frac{\Gamma(N+M+J+\alpha+\beta+p)}{\Gamma(2M+J+\alpha+\beta+p+1)}. \tag{4.12}$$

The $\{j_k\}$ summations can be performed by again using identity (2.3), now with the parameter values $\eta = N+M+\alpha+\beta+p$, $\xi_k = -n_k+m_k$, and $\gamma = 2M+\alpha+\beta+p+1$, which yields

$$\sum_{\{x_k\}} H_n(x)G_m(x)w(x)$$
$$= \left[\prod_{k=1}^{p} \frac{\Gamma(\alpha_k+1)\Gamma(\alpha_k+1)n_k!}{\Gamma(m_k+\alpha_k+1)(n_k-m_k)!}\right]$$
$$\times \frac{1}{(M-N)!}\frac{(\Delta-N)!}{\Delta!\Delta!}$$
$$\times \frac{\Gamma(M+\beta+1)\Gamma(\Delta+N+\alpha+\beta+p+1)}{(N+M+\alpha+\beta+p)\Gamma(N+\alpha+\beta+p)}, \tag{4.13}$$

which is clearly zero unless $n_k = m_k$ for every k; that is,

$$\sum_{\{x_k\}} H_n(x)G_m(x)w(x) = h_n \prod_{k=1}^{p} \delta_{n_k m_k}, \tag{4.14}$$

where the normalization constant is given by

$$h_n = \left[\prod_{k=1}^{p} \frac{\Gamma(\alpha_k+1)\Gamma(\alpha_k+1)n_k!}{\Gamma(n_k+\alpha_k+1)}\right] \frac{(\Delta-N)!}{\Delta!\Delta!}$$
$$\times \frac{\Gamma(N+\beta+1)\Gamma(\Delta+N+\alpha+\beta+p+1)}{(2N+\alpha+\beta+p)\Gamma(N+\alpha+\beta+p)}. \tag{4.15}$$

V. DISCUSSION

In the special case $p = 1$ expressions (3.2), (3.3), and (3.11) reduce to known representations of the single-variable Hahn polynomials. For a single variable Eqs. (4.5) and (4.6) imply that both $H_n(x)$ and $G_n(x)$ form an orthogonal family with the same weight function. In fact, it follows from (3.12) and the equality of $P_n(x)$ and $Q_n(x)$ for $p = 1$ that for a single variable $H_n(x) = G_n(x)$. Equations (3.7)–(3.9) for $p = 1$ are then further representations of the single-variable Hahn polynomials.

An analogous multivariable extension of the Meixner, Krawtchouk, and Meixner–Pollaczek polynomials will appear in a future publication. Currently under investigation are similar generalizations of the Wilson and Racah polynomials and their q analogs.

ACKNOWLEDGMENTS

I would like to thank the referee for a careful reading of the manuscript and for many useful comments.

This work was supported by the Natural Sciences and Engineering Research Council of Canada and the United States Department of Energy.

[1]W. Hahn, Math. Nach. 2, 4 (1949).
[2]P. L. Tchebychef, in *Oeuvres de P.L. Tchebychef, Tome II*, edited by A. Markoff and N. Sonin (St. Petersburg, 1899), (reprinted Chelsea, New York, 1961).
[3]*Gabor Szegö, Collected Papers*, edited by R. Askey (Birkhäuser, Stuttgart, 1982), Vol. 3, p. 866.
[4]M. Weber and A. Erdelyi, Am. Math. Mon. 59, 163 (1952).

[5]S. Karlin and J. McGregor, Scr. Math. **26**, 33 (1961).
[6]J. J. Bartko, Am. Math. Mon. **70**, 978 (1963).
[7]R. J. Levit, SIAM Rev. **9**, 191 (1967).
[8]P. A. Lee, SIAM J. Appl. Math. **19**, 266 (1970).
[9]M. W. Wilson, SIAM J. Math. Anal. **1**, 131 (1970).
[10]G. Gasper, J. Math. Anal. Appl. **42**, 438 (1973).
[11]T. H. Koornwinder, Nieuw Arch. Wisk. **29**, 140 (1981).
[12]A. F. Nikiforov, S. K. Suslov, and V. B. Uvarov, *Classical Orthogonal Polynomials of a Discrete Variable* (Nauka, Moscow, 1985) (in Russian).
[13]S. Karlin and J. McGregor, in *Theory and Application of Special Functions*, edited by R. A. Askey (Academic, New York, 1975).
[14]N. M. Atakishiyev and S. K. Suslov, J. Phys. A **18**, 1583 (1985).
[15]R. Askey, J. Phys. A **18**, L1017 (1985).
[16]M. V. Tratnik, J. Math. Phys. **29**, 1529 (1988).
[17]H. M. Srivastava and P. W. Karlsson, *Multiple Gaussian Hypergeometric Series* (Ellis Horwood, New York, 1985), p. 38.
[18]W. N. Bailey, *Generalized Hypergeometric Series* (Cambridge U.P., London, 1950).
[19]P. Appell and J. Kampé de Fériet, *Fonctions Hypergéométriques et Hypersphériques* (Gauthier–Villars, Paris, 1926).
[20]J. Thomae, J. Math. **87**, 26 (1897).
[21]C. S. Lam and M. V. Tratnik, Can. J. Phys. **63**, 1427 (1985).

Affine structure and isotropy imply Minkowski space-time and the orthochronous Poincaré group

John W. Schutz
Bendigo College of Advanced Education, Bendigo, Victoria 3550, Australia

(Received 14 January 1988; accepted for publication 26 October 1988)

Minkowski space-time is characterized in terms of affine structure and a property of isotropy with respect to the path of a single freely moving observer. It can then be shown that the property of isotropy applies for all freely moving observers, and that the associated isotropy mappings generate the orthochronous Poincaré group.

I. INTRODUCTION

Minkowski space-time $\mathscr{M}$ is assumed to be a four-dimensional affine space. The derivation of the remaining structure is based on properties related to the path of a single freely moving observer. The assumed properties may be stated in the following physical terms.

(i) Freely moving observers move on paths according to Newton's first law.

(ii) Space-time is isotropic with respect to one freely moving observer.

(iii) Action at a distance is not instantaneous.

The Euclidean geometry of position space and the constancy of the velocity of light are not assumed, but are obtained as corollaries to the main results. In Theorem 1, it is shown that the set of all paths through a given event is an ellipsoidal cone. The proof is based on a characterization of ellipsoids in terms of isotropy about a single point as given by Busemann.[1] The ellipsoidal cone structure corresponds to the usual Minkowski space-time, which is coordinatized in the usual way. It can then be shown that isotropy mappings generate the orthochronous Poincaré group. This group of motions can then be used to generate a set of coordinate frames for which position space is Euclidean and the speed of light is constant and equal to unity.

There is a long tradition of characterizing geometries and space-times by properties of symmetry and free mobility. In the case of geometry, this is the celebrated Riemann–Helmholtz–Lie space-form problem, whose solutions are given and reviewed by Busemann[1] and Freudenthal.[2] The analogous space-time problem has been investigated by Busemann,[3,4] Freudenthal,[5] and Mayr,[6] who make extensive use of Lie group theory, and also by Aleksandrov,[7] who assumes from the outset that space-time is affine. Aleksandrov[7] and Busemann[3] consider a symmetry property which Aleksandrov[8] describes as "isotropy of the light cone" (see also Refs. 9–11), together with additional conditions, to show that the light cones are ellipsoidal cones in an affine space. Freudenthal[5] and Mayr[6] use a property of "free mobility." Pimenov[12] discusses "affine space-times" with a property of "mobility of frames."

Characterizations of Minkowski space-time, using properties of symmetry together with properties of the paths of freely moving observers, have been proposed by several authors. Alexandrov[13] uses a property of "reflection in the path of every observer." Schutz[14,15] uses a property of isotropy similar to that of the present treatment, but characterizes Minkowski space-time indirectly by first demonstrating that its velocity space is hyperbolic. The axiomatic system of Szekeres[16] is stated in terms of freely moving observers and lightlike paths, and has an axiom of isotropy and two other symmetry axioms. Walker[17] discusses cosmologies of "fundamental particles" which satisfy two axioms of symmetry, one of which is a property of isotropy similar to that of the present treatment. A useful survey of the literature on characterizations of Minkowski space-time and the Lorentz group is given by Guts.[18]

In the present exposition, Minkowski space-time is also characterized by properties of the paths of freely moving observers. The concept of isotropy (which resembles Axiom X of Walker[17]) may be considered from the point of view of a single "fixed" observer, to whom the rest of space-time appears to be isotropic.

II. CHARACTERIZATION OF MINKOWSKI SPACE-TIME

A. Definitions

A space-time $\mathscr{M} = \langle \mathscr{E}, \mathscr{P} \rangle$ is a set of *events* $\mathscr{E}$ together with a set $\mathscr{P}$ of subsets of $\mathscr{E}$ which are called *(inertial) paths* or *timelike lines*. We will say that a space-time is *four-dimensional* and *affine* if it can be represented with a four-dimensional affine space A^4 where the events of $\mathscr{E}$ correspond to the points of the space, and the set of paths $\mathscr{P}$ corresponds to a subset of the set of straight lines. Individual paths are denoted by the symbols **Q**,**R**,**S**,... . Events which belong to a path are denoted by the path symbol together with a subscript; for example, the events Q_1, Q_a, Q_x belong to the path **Q**. A path **Q** and an event $e \notin \mathbf{Q}$ specify a *plane* $pl[\mathbf{Q},e]$ which contains a subset of paths "in one-dimensional rectilinear motion."

Given a timelike line **Q** and two half-planes with edge **Q**, we say that an automorphism θ of $\mathscr{M}$ which leaves **Q** and all its events invariant, and which maps the first half-plane onto the second, is an *isotropy mapping with invariant timelike line* **Q**. If, for a given timelike line **Q** and any two half-planes with edge **Q**, there is an isotropy mapping with invariant timelike line **Q**, we say that the space-time $\mathscr{M}$ is *isotropic with respect to* **Q**. If $\mathscr{M}$ is isotropic with respect to every timelike line, we say that the *space-time is isotropic*.

Given a path **S** and an event $e \notin \mathbf{S}$, the subset of events of **S** which can be joined to e by single paths is called the *reachable subset of* **S** *from* e, or simply the *reachable set*, and is denoted as $\mathbf{S}(e)$. The subset of events of **S** which can not be

 0022-2488/89/030635-04$02.50

joined to e by single paths is called the *unreachable set* and is denoted as $\mathbf{S}\setminus\mathbf{S}(e)$. If the unreachable set $\mathbf{S}\setminus\mathbf{S}(e)$ contains at least two distinct events and is connected (with respect to the usual order topology on $\mathbf{S}$), this may be interpreted in physical terms by saying that "action at a distance is not instantaneous," or simply by referring to "noninstantaneous interaction." If the reachable set $\mathbf{S}(e)$ is open (with respect to the usual order topology on $\mathbf{S}$), this corresponds to "there being no fastest path between $\mathbf{S}$ and e." Both of these conditions together are equivalent to the statement that the "unreachable set is a closed interval," which is stated in Theorem 1 as condition (iii). An example of a space-time in which these properties are not satisfied is Galilean space-time, in which each unreachable set consists of a single event, and, in a limiting sense, "interaction is instantaneous."

As well as timelike lines, two other types of lines will be considered: *lightlike lines* are straight lines which pass through e and the events which bound the reachable set $\mathbf{S}(e)$, while *spacelike lines* are straight lines which pass through an event e and the nonboundary events of the unreachable set $\mathbf{S}\setminus\mathbf{S}(e)$. Lines of all types will be denoted by upper case letters; when referred to an affine coordinate system, each line has a parametric equation of the form

$$x_i = x_{i(\text{initial})} + \lambda w_i \quad (i = 0,1,2,3),$$

where the four-component direction vector w_i is called a *four-velocity vector.* We will use the index conventions that italic letters $i,j,k,\ldots$ range over the integers 0,1,2,3, while greek letters $\alpha,\beta,\gamma,\ldots$ range over the integers 1,2,3, and repeated subscripts imply a sum in accordance with the Einstein summation convention.

B. Characterization theorem

The main result of the present paper is contained in the next theorem, which gives sufficient conditions to characterize Minkowski space-time.

Theorem 1 (Minkowski space-time): If, (i) $\mathscr{M}$ is a four-dimensional affine space-time and the set of paths $\mathscr{P}$ consists of equivalence classes of parallels, (ii) $\mathscr{M}$ is isotropic with respect to one given path $\mathbf{Q}$, and (iii) each unreachable set is a closed interval, then $\mathscr{M}$ has a set of affine coordinates such that the set of paths $\mathscr{P}$ is characterized by four-velocity vectors which satisfy the inequality

$$w_0^2 - w_1^2 - w_2^2 - w_3^2 > 0,$$

and such that the four-velocity of $\mathbf{Q}$ is (1,0,0,0). That is, the four-velocity vectors of paths lie within an ellipsoidal cone.

Remark: Lightlike lines and spacelike lines have four-velocity vectors which satisfy the corresponding equality and reversed inequality.

Proof: (a) Each isotropy mapping with invariant path $\mathbf{Q}$ *induces an affinity of* A^4. Since each isotropy mapping maps $\mathscr{P}$ onto itself, it follows by a result of Hegerfeldt[19] that each isotropy mapping is an affinity.

(b) Definition of the set of parallels $\mathscr{K}$. The set of straight lines through Q_0 is a three-dimensional projective space with each line being a "point" of the space. Any plane through Q_0 contains a set of lines through Q_0 which are the points of a "straight line" of the projective space. Condition (iii) implies that the set of timelike lines $\mathscr{V}$ is a set of points with the property that any "line" which meets $\mathscr{V}$ must meet it in an "open segment" whose complement is a nontrivial closed interval. Thus by a result of Kneser,[20] $\mathscr{V}$ is an open convex subset of the projective space of lines through Q_0. Now, considering A^4, in each plane which contains $\mathbf{Q}$, there are two lightlike lines through Q_1 and two lightlike lines through Q_3 (see Fig. 1). A pair of nonparallel lightlike lines through these events meets in some event $a \in \mathscr{E}$ which can be joined to Q_0 by a timelike line (since $\mathscr{V}$ is convex); moreover there is a line $\mathbf{A}$ parallel to $\mathbf{Q}$ such that $a \in \mathbf{A}$. Thus in each plane which contains $\mathbf{Q}$ there are two such parallels, one on either side of $\mathbf{Q}$; each of these two parallels corresponds to the intersection of two lightlike lines—one through Q_1 and the other through Q_3. Any two timelike lines which pass, respectively, through Q_1 and Q_3 and which are parallel, respectively, to lines of $\mathscr{V}$ on opposite sides of $\mathbf{Q}$, meet in an event through which there is a parallel to $\mathbf{Q}$; since $\mathscr{V}$ is convex, this parallel is between the previously mentioned pair of parallels generated by the intersection of lightlike lines. In each plane, the set of all such parallels (including the pair defined by the lightlike lines) is connected, contains $\mathbf{Q}$, and is bounded by the pair of parallels which were defined by the lightlike lines. This procedure applies to each plane through $\mathbf{Q}$; the set of all such parallels is denoted by $\mathscr{K}$.

(c) The set of parallels $\mathscr{K}$ *is an ellipsoid.* Since $\mathscr{V}$ is convex, the set of half-lines (corresponding to the timelike and lightlike lines of $\overline{\mathscr{V}}$) is separated into two components by any of the supporting "planes" (hyperplanes). Hence the convexity of both components implies the convexity of $\mathscr{K}$. Furthermore, $\overline{\mathscr{V}}$ is convex and each line (plane) containing $\mathbf{Q}$ has two distinct boundary points (lightlike lines), so the set $\mathscr{K}$ of parallels is closed and therefore bounded. By part (a) above, any isotropy mapping θ with invariant path $\mathbf{Q}$ induces an affinity ψ of the set of parallels (to $\mathbf{Q}$), and so ψ maps $\mathscr{K}$ bijectively onto itself. Let us denote the boundary of $\mathscr{K}$ by $\mathscr{C}$. Then the preconditions of the characterization of ellipsoids given by Busemann[1] are satisfied. It follows that the set of parallels of $\mathscr{C}$, interpreted as points, form an ellipsoidal surface and hence the set of parallels $\mathscr{K}$, interpreted as points, form a solid three-dimensional ellipsoid.

(d) Coordinatization of the set of parallels. Since $\mathscr{K}$ is an ellipsoid, the three-dimensional affine space of parallels (to

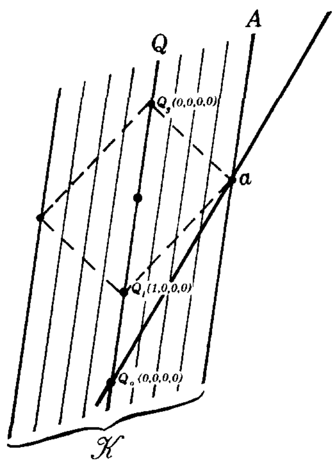


FIG. 1. The set of parallels $\mathscr{K}$ of the proof of Theorem 1.

Q) can be coordinatized with a set of Y_α coordinates ($\alpha = 1,2,3$) such that the set of all parallels of $\mathscr{K}$ satisfy the inequality

$$y_1^2 + y_2^2 + y_3^2 \leqslant 1, \tag{1}$$

where the equality corresponds to the parallels of the boundary $\mathscr{C}$.

(e) Coordinatization of the events of $\mathscr{M}$. The events of $\mathscr{M}$ can now be coordinatized with a set of X_i coordinates ($i = 0,1,2,3$) in the following way: each event lies on some parallel with coordinates (y_1,y_2,y_3), so we define the coordinates X_α ($\alpha = 1,2,3$) for the event to be the same as for the parallel. The line **Q** has coordinates $(y_1,y_2,y_3) = (0,0,0)$ and its events have already been indexed with an affine parameter, so for each event $Q_t \in \mathbf{Q}$ we define $x_0 = t$, and then the coordinates of Q_t are $(t,0,0,0)$.

On each plane through **Q**, there are two bounding parallels in $\mathscr{C}$ such that

$$y_1^2 + y_2^2 + y_3^2 = 1. \tag{2}$$

We consider one such parallel **A** and [as discussed in part (b) of this proof] there are lightlike lines which pass through the events $(1,0,0,0)$ and $(3,0,0,0)$ and which meet at an event $a \in \mathbf{A}$; we specify that the x_i coordinates of a are $(2,y_1,y_2,y_3)$, and for the lightlike line which passes through $(1,0,0,0)$ and through $a(2,y_1,y_2,y_3)$ we specify the event coordinates x_i with the (four) equations

$$(x_0,x_1,x_2,x_3) = (1,0,0,0) + \mu(1,y_1,y_2,y_3), \quad \mu \in \mathbb{R}. \tag{3a}$$

These equations specify the x_0 coordinate at one event on each parallel to **Q** in the plane which contains **Q** and **A**. Now take a lightlike line through the event $(3,0,0,0)$ parallel to the previous lightlike line. Let the event at which this line meets **A** have the coordinates $(4,y_1,y_2,y_3)$. Then this line has the parametric equations

$$(x_0,x_1,x_2,x_3) = (3,0,0,0) + \mu(1,y_1,y_2,y_3), \quad \mu \in \mathbb{R}. \tag{3b}$$

These equations specify an x_0 coordinate on a second event on each parallel to **Q** in the plane spanned by **Q** and **A**. Thus each parallel has been equipped with an affine parameter, and hence the plane has been given an affine coordinatization.

Since each plane through **Q** passes through some bounding parallel of $\mathscr{C}$ [satisfying Eq. (2)], it follows that the entire affine space A^4 is equipped with an affine coordinatization.

(f) Equations of timelike, lightlike, and spacelike lines. Any line in A^4 has an equation of the form

$$x_i = x_{i(\text{initial})} + \mu w_i, \quad \mu \in \mathbb{R} \tag{4}$$

(for $i = 0,1,2,3$), where μ is a line parameter and w_i is a constant four-velocity vector. Equations (2) and (3) imply that for lightlike lines

$$w_0^2 - w_1^2 - w_2^2 - w_3^2 = 0, \tag{5a}$$

hence for timelike lines

$$w_0^2 - w_1^2 - w_2^2 - w_3^2 > 0. \tag{5b}$$

The remaining lines are spacelike lines which must therefore satisfy

$$w_0^2 - w_1^2 - w_2^2 - w_3^2 < 0. \tag{5c}$$

Q.E.D.

III. THE ORTHOCHRONOUS POINCARÉ GROUP

It can be shown[11] that the set of automorphisms of the light cone [Eq. (5a)] is the Lorentz group augmented by dilatations.[21] Thus each Lorentz transformation which has an invariant path is an isotropy mapping. Two special types of isotropy mapping are: (i) Lorentz transformations which leave the path **Q** invariant—these correspond to orthogonal transformations of "position space," and (ii) "reflections in timelike lines." Lorentz boosts may be generated from the composition of mappings of type (ii), and then translations may be generated by the composition of boosts.

It is straightforward to show that these motions generate the orthochronous Poincaré group. Thus the property of isotropy applies for all freely moving observers and the associated isotropy mappings generate the orthochronous Poincaré group.[22]

IV. CONCLUSION

Minkowski space-time has been characterized by assuming affine structure, isotropy with respect to a single observer, and noninstantaneous interaction. It then follows that the property of isotropy applies with respect to every observer, and that isotropy mappings generate the orthochronous Poincaré group.

[1]H. Busemann, *The Geometry of Geodesics* (Academic, New York, 1955). Busemann states and proves the proposition: "(16.11) Let $\mathscr{C}$ be a closed convex surface in A^n and z an interior point of $\mathscr{C}$. If for any two points p and q of $\mathscr{C}$ an affinity exists that leaves z fixed, maps $\mathscr{C}$ on itself and p on q, then $\mathscr{C}$ is an ellipsoid with center z." The convex surface $\mathscr{C}$ of Busemann's proposition corresponds to the boundary of the convex body $\mathscr{K}$ of our Theorem 1.

[2]H. Freudenthal, "Lie groups in the foundations of geometry," Adv. Math. **1**, 145 (1965).

[3]H. Busemann, "Timelike spaces," Dissertationes Math. (Rozprawy Mat.) **53**, 1 (1967).

[4]Busemann considers "timelike Minkowski spaces," which have the additional structure of a pseudometric "gauge function." Accordingly, the usual Minkowski space-time (Lorentz space in Busemann's terminology) is characterized by conditions on timelike and lightlike lines, with an additional condition of isometry, or "triplewise transitivity." In the present paper, Minkowski space-time is considered to be described by the structure of its timelike (and lightlike) lines; the squared interval and the pseudometric are derived concepts.

[5]H. Freudenthal, "Das Helmholtz-Liesche raumproblem bei indefiniter metrik," Math Ann. **156**, 263 (1964).

[6]D. Mayr, "A constructive-axiomatic approach to physical space and space-time geometries of constant curvature by the principle of reproducibility," in *Space-Time and Mechanics*, edited by D. Mayr and G. Sussmann (Reidel, Dordrecht, 1983).

[7]A. D. Aleksandrov, "Cones with a transitive group," Dokl. Akad. Nauk SSSR **189**, 695 (1969) [Sov. Math. Dokl. **10**, 1460 (1969)].

[8]Aleksandrov (Ref. 7) considers a property of "transitivity" on the rays of the future light cone; the corresponding symmetry group is the orthochronous Lorentz group with dilatations as shown by Aleksandrov (Ref. 9) and Zeeman (Ref. 10). The same result is implied by a result of Busemann (Ref. 3) (§ 7, p. 43, Theorem 9). An alternative characterization in terms of transitivity on the lines of the light cone can be based on Busemann's characterization (Ref. 3) (§ 6, p. 34, Proposition 2) of ellipsoids in projective space; then the corresponding symmetry group is the Lorentz group with dilatations, as obtained by Alexandrov (Ref. 9) and Borchers and Hegerfeldt (Ref. 11).

[9]A. D. Alexandrov [sic, same author as Aleksandrov (Ref. 7)], "On Lorentz transformations," Usp. Mat. Nauk **5**, 187 (1950); "Mappings of spaces with families of cones and space-time transformations," Ann. Mat. Pura Appl. **103**, 229 (1975).
[10]E. C. Zeeman, "Causality implies the Lorentz group," J. Math. Phys. **5**, 490 (1964).
[11]H. J. Borchers and G. C. Hegerfeldt, "The structure of space-time transformations," Commun. Math. Phys. **28**, 259 (1972).
[12]R. I. Pimenov, "Kinematic spaces," Zap. Nauč. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **6**, 3 (1968). [Seminars in Mathematics (V. A. Steklova Mathematical Institute, Leningrad) **6**, 1 (1970)]. Pimenov defines an "affine space-time" to have a partial order relation $<$ in addition to the affine properties used in the present paper.
[13]A. D. Alexandrov [sic, same author as Aleksandrov (Ref. 7)], "A contribution to chronogeometry," Can. J. Math. **19**, 1119 (1967).
[14]J. W. Schutz, "Foundations of special relativity: Kinematic axioms for Minkowski space-time," *Lecture Notes in Mathematics*, Vol. 361 (Springer, Berlin, 1973).
[15]J. W. Schutz, "An axiomatic system for Minkowski space-time," J. Math. Phys. **22**, 293 (1981).
[16]G. Szekeres, "Kinematic geometry: An axiomatic system for Minkowski space–time," J. Aust. Math. Soc. **8**, 134 (1968).
[17]A. G. Walker, "Axioms for cosmology," in *The Axiomatic Method*, edited by L. Henkin, P. Suppes, and A. Tarski (North-Holland, Amsterdam, 1959).
[18]A. K. Guts, "Axiomatic relativity theory," Usp. Mat. Nauk **37**, 39 (1982) [Russian Math. Surveys **37**, 41 (1982)].
[19]G. C. Hegerfeldt, "The Lorentz transformations: Derivation of linearity and scale factor," Nuovo Cimento A **10**, 257 (1972). The result of linearity is stated for mappings of lines parallel to the lines within a light cone. However, the proof applies even more generally to the mappings of straight lines which satisfy our conditions (i) and (iii). In the present situation an even simpler proof may be used, since events on one path are invariant and conditions (ii) and (iii) imply that each plane which contains the invariant path contains a set of timelike lines bounded by two lightlike lines.
[20]H. Kneser, "Eine Erweiterung des Begriffes 'konvexer Körper,'" Math. Ann. **82**, 287 (1921). Satz 5 (Theorem 5) on p. 296 includes the stated result as a special case. Alternatively the result may be obtained directly, using elementary methods of projective geometry.
[21]Since we are considering affinities, we may also use the well-known and less general result which is given, for example, by W. Rindler, *Special Relativity* (Oliver and Boyd, Edinburgh, 1960), p. 21.
[22]The results of Sec. III are obtained directly using the same property of isotropy by J. W. Schutz, "The isotropy mappings of Minkowski space-time generate the orthochronous Poincaré group," to be published.

Solutions of certain integrable nonlinear PDE's describing nonresonant N-wave interactions

F. Calogero
Dipartimento di Fisica, Università di Roma "La Sapienza," 00185 Roma, Italy, and Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Roma, Italy

(Received 16 September 1988; accepted for publication 16 November 1988)

The evolution equations that describe, in appropriately "coarse-grained" and "slow" variables, the evolution of the envelopes of N nonresonant dispersive waves, can be reduced to the "universal" form $(\partial/\partial t + v_n \partial/\partial x)\, u_n(x,t) = u_n(x,t) \sum_{m=1}^{N} \beta_{nm} u_m(x,t)$. In this paper the special case with $\beta_{nm} = (v_m - v_n)\beta_m$, which is integrable by quadratures, is investigated. The subclass of localized solutions (i.e., vanishing as $x \to \pm\infty$) gives rise to a novel solitonic phenomenology. The class of solutions that are asymptotically finite contains a richer solitonic phenomenology, including those of novel type (which move with the speeds v_n, and can have any shape) and more standard kinks (which can move with any speed, and have standard shapes). The class of rational solutions, and the integrable dynamical systems naturally associated with these solutions, are also investigated; these dynamical systems include, and extend, known integrable systems. In this paper the treatment is confined to $1+1$ dimensions; at the end, together with some other generalizations, a partially solvable multidimensional extension is reported.

I. INTRODUCTION

It has been recently pointed out that the nonlinear evolution PDE's that describe the interaction of N dispersive waves in the regime of weak nonlinearity have a universal character, and are therefore likely to be both widely applicable and integrable.[1,2] In particular, the standard equations that describe the nonresonant interaction of N dispersive waves have the form

$$\left(\frac{\partial}{\partial t} + v_n \frac{\partial}{\partial x}\right)\Psi_n = i\Psi_n(x,t) \sum_{m=1}^{N} \alpha_{nm} |\Psi_m|^2,$$
$$\Psi_n \equiv \Psi_n(x,t). \tag{1.1}$$

Here x and t are appropriately "coarse-grained" and "slow" space and time variables [our consideration in this paper is limited to the $(1+1)$-dimensional case]; the N real constants v_n are the (different) group velocities associated with the N dispersive waves under consideration, and without loss of generality we assume hereafter that

$$v_n < v_{n+1}, \quad n = 1,2,\ldots,N-1; \tag{1.2}$$

the N dependent variables Ψ_n are the complex amplitudes accounting for the (amplitude) modulation of the N dispersive waves due to the (weak) nonlinearity; and the N^2 constants α_{nm} are generally complex.[1] Of course in (1.1), and throughout this paper [unless otherwise indicated; see, e.g., (1.2)], the index n takes the values $1,2,\ldots,N$.

An interesting case is that in which the imaginary part of α_{nm} is proportional to the difference of the velocities v_n and v_m,

$$\operatorname{Im} \alpha_{nm} = (v_n - v_m)\beta_m, \quad \beta_m \text{ real}; \tag{1.3}$$

because the nonlinear evolution PDE's (1.1) are then C-integrable,[1] namely, their solution can be obtained by quadratures. Purpose and scope of this paper is to analyze the solutions of (1.1) in this case (1.3). As shown below, these solutions display a rich and remarkable phenomenology, that can be exhibited in explicit detail thanks to the C-integrable nature of (1.1) with (1.3).

The paper is organized as follows. Section II summarizes tersely the solution technique.[1] Section III deals with localized solutions, namely, solutions of (1.1) with (1.3) such that

$$\int_{-\infty}^{+\infty} dx |\Psi_n(x,t)| < \infty. \tag{1.4}$$

As we show below, this case gives rise to a new type of solitonic phenomenology, with the following features. *Only solitons are present* (no background or radiation); every solution, in the remote past and future, factors into N separate solitons (see below). Each separated soliton has only one wave present, and travels with the corresponding velocity, say

$$u_n(x,t) = \delta_{n\nu} u_\nu(x - v_\nu t) \tag{1.5}$$

[through the rest of this section we denote by $u_n(x,t)$ the renormalized squared modulus of $\Psi_n(x,t)$, $u_n(x,t) = 2\beta_n |\Psi_n(x,t)|^2$]; the function $u_\nu(y)$ is arbitrary (localized). Hence *these N solitons have fixed speeds but arbitrary shapes.* The collision of two solitons produces a change in the shape of each of them; the new profile of each soliton after the collision, say $\tilde{u}_\nu(y)$, is determined by its profile before the collision, say $\hat{u}_\nu(y)$, according to the *universal* formula

$$\tilde{u}_\nu(y) = \hat{u}_\nu(y)\left(1 + a \exp\left[\int_{-\infty}^{y} dx\, \hat{u}_\nu(x)\right]\right)^{-1}, \tag{1.6}$$

with the parameter a given, in terms of the integrals over the profiles of the two colliding solitons, say

$$\hat{b}_\mu = \int_{-\infty}^{+\infty} dx\, \hat{u}_\mu(x), \quad \hat{b}_\nu = \int_{-\infty}^{+\infty} dx\, \hat{u}_\nu(x), \tag{1.7}$$

by the formula

 0022-2488/89/030639-16$02.50

$$a = \exp(-\hat{b}_\mu) - 1, \quad \text{if } v_\nu < v_\mu, \tag{1.8a}$$

$$a = \exp(-\hat{b}_\nu)[\exp(\hat{b}_\mu) - 1], \quad \text{if } v_\nu > v_\mu. \tag{1.8b}$$

In the remote past and future, every solution factors into N separated solitons,

$$u_n(x,t) \approx u_n^\pm(x - v_n t), \quad t \approx \pm\infty; \tag{1.9}$$

of course the mth soliton is to the right of the $(m+1)$th soliton in the remote past, and to its left in the remote future [see (1.2)]. Hence the final outcome obtains from the configuration in the remote past after all the solitons have overtaken each other. The profile of each soliton in the remote future, $u_n^{(+)}(y)$, is related to its profile in the remote past, $u_n^{(-)}(y)$, by the *universal* formula

$$u_n^{(+)}(y) = u_n^{(-)}(y)\Big(1 + a_n \times \exp\Big[\int_{-\infty}^{y} dx\, u_n^{(-)}(x)\Big]\Big)^{-1}, \tag{1.10}$$

with the constant a_n determined by the integrals over the profiles of the solitons in the remote past,

$$b_n^{(-)} = \int_{-\infty}^{+\infty} dx\, u_n^{(-)}(x), \tag{1.11}$$

according to the formula

$$a_n = \exp\Big(\sum_{m=1}^{n} b_m^{(-)}\Big) + \exp\Big(-\sum_{m=n+1}^{N} b_m^{(-)}\Big) - \exp(b_n^{(-)}) - 1. \tag{1.12}$$

This final outcome corresponds of course to what would result from a sequence of pair soliton collisions, see (1.6)–(1.8); note, however, that *it is independent of the localization of the solitons in the remote past* [other than their order, which is determined by (1.2)], even though this determines the sequence of pair collisions, or the eventual occurrence of multiple collisions (whose relevance depends of course also on the widths of the soliton profiles, both initially and as they get modified by the interactions through the time evolution). It is this last feature that, in our opinion, justifies the use of the term "soliton" to describe this phenomenology.

Section IV deals with solutions that tend asymptotically $(x \to \pm\infty)$ to finite values (rather than vanishing). This class of solutions features, in addition to a soliton phenomenology analogous to that described above, another kind of soliton, of more standard type. These new solitons, which of course also move with constant speed and shape when they are isolated, involve more than one wave; they are genuinely nonlinear. Moreover, *their speed V can take any value* other than the v_n's, $V \neq v_n$, while *their shape is fixed*. The prototype of this kind of soliton is the *kink* involving *two* waves,

$$u_n(x,t) = \delta_{n\mu} p_\mu / \{1 + \exp[(p_\nu - p_\mu)(x - Vt - x_0)]\} + \delta_{n\nu} p_\nu / \{1 + \exp[(p_\mu - p_\nu)(x - Vt - x_0)]\}, \tag{1.13}$$

where the (real) parameters x_0, p_μ, and p_ν are arbitrary $(\mu \neq \nu, p_\mu \neq p_\nu)$, and

$$V = (p_\mu v_\mu - p_\nu v_\nu)/(p_\mu - p_\nu). \tag{1.14}$$

It should be noted that also in this case the solution generally contains only solitons (no radiation), as demonstrated by its behavior in the remote past and future, when it separates into the sum of distinct solitons.

In Sec. V we consider the special class of solutions of (1.1) with (1.3) such that the square moduli $|\Psi_n(x,t)|^2$ are *rational functions* (of both x and t). We also discuss the finite-dimensional integrable dynamical systems (which turn out to generalize known integrable dynamical systems[3]) that are naturally associated with such solutions.

Finally in Sec. VI we outline some future developments of these researches.

II. METHOD OF SOLUTION

We set

$$\Psi_n(x,t) = [\tilde{u}_n(x,t)]^{1/2} \exp[i\theta_n(x,t)]. \tag{2.1}$$

Here we assume that the quantities $\tilde{u}_n(x,t)$ are real and nonnegative, and the quantities $\theta_n(x,t)$ are also real [and of course defined $\operatorname{mod}(2\pi)$].

Then (1.1) with (1.3) yield

$$\Big(\frac{\partial}{\partial t} + v_n \frac{\partial}{\partial x}\Big)\theta_n(x,t) = \sum_{m=1}^{N} (\operatorname{Re}\alpha_{nm})\tilde{u}_m(x,t), \tag{2.2a}$$

$$\Big(\frac{\partial}{\partial t} + v_n \frac{\partial}{\partial x}\Big)\tilde{u}_n(x,t) = 2\tilde{u}_n(x,t) \sum_{m=1}^{N} (v_m - v_n)\beta_m \tilde{u}_m(x,t). \tag{2.2b}$$

The explicit solution of (2.2a) reads

$$\theta_n(x,t) = \hat{\theta}_n(x - v_n t) + \int_0^t dt' \sum_{m=1}^{N} \times (\operatorname{Re}\alpha_{nm})\tilde{u}_m[x - v_n(t - t'), t']. \tag{2.3}$$

Here the n functions $\hat{\theta}_n(x)$ are arbitrary; in the context of the Cauchy problem, they are determined by the initial conditions

$$\hat{\theta}_n(x) = \theta_n(x,0), \tag{2.4}$$

that are clearly implied by (2.3).

The functions $\tilde{u}_m(x,t)$ that appear in the rhs of (2.3) are the solutions of the nonlinear system of evolution equations (2.2a). It is clearly convenient to set

$$\tilde{u}(x,t) = u(x,t)/(2\beta_m), \tag{2.5}$$

so that (2.2b) read

$$\Big(\frac{\partial}{\partial t} + v_n \frac{\partial}{\partial x}\Big)u_n(x,t) = u_n(x,t) \sum_{m=1}^{N} (v_m - v_n)u_m(x,t). \tag{2.6}$$

Our attention hereafter will be focused on this system of nonlinear evolution PDE's. Note that, to the extent that this system is related to (1.1) with (1.3) via (2.1), one should limit consideration to solutions of (2.6) such that

$$\beta_n u_n(x,t) \geqslant 0. \tag{2.7}$$

To solve (2.6) it is expedient to set

$$u_n(x,t) = w_n(x - v_n t)/F(x,t). \tag{2.8}$$

Then (2.6) yields

$$F_t(x,t) = -\sum_{m=1}^{N} v_m w_m(x-v_m t), \tag{2.9a}$$

$$F_x(x,t) = \sum_{m=1}^{N} w_m(x - v_m t). \tag{2.9b}$$

The compatibility of these two equations, $F_{tx} = F_{xt}$, is evident.

Hence the general solution of (2.6) reads

$$u_n(x,t) = w_n(x - v_n t)\Big(\Big[C + \sum_{m=1}^{N} \int_{x_0}^{x - v_m t} dx' w_m(x')\Big]\Big), \tag{2.10}$$

where the two real constants (x and t independent) C and x_0, and the N real functions $w_m(x)$, can be chosen arbitrarily.

It is easily seen that (2.10) implies

$$w_n(x) = C\, u_n(x,0)\exp\Big[\int_{x_0}^{x} dx' \sum_{m=1}^{N} u_m(x',0)\Big]; \tag{2.11}$$

this formula provides an explicit expression of the functions $w_n(x)$ in terms of the initial data $u_m(x,0)$.

The formulas (2.11), (2.10), and (2.3) with (2.4) and (2.5) provide, via (2.1), the explicit solution of the Cauchy problem for (1.1) with (1.3); the two constants C and x_0 may be chosen at one's convenience (see below).

In the following we discuss the behavior of certain classes of solutions. Our treatment will focus on the solutions of (2.6), as given by (2.10) and (2.11).

Let us end this section with two remarks.

The system of nonlinear evolution PDE's (2.6) is clearly invariant under the transformations $x \to \alpha x$, $t \to \alpha t$, $u_n \to u_n/\alpha$ (scale transformation), with x_0, t_0, and α arbitrary constants ($\alpha \neq 0$). It is moreover invariant under the (Galileo) transformation $v_n \to v_n + v_0$, $x \to x - v_0 t$; hence we may hereafter assume, without loss of generality, that all the velocities v_n are positive,

$$0 < v_1 < v_2 < \cdots < v_N. \tag{2.12}$$

The system of nonlinear evolution PDE's (2.6) implies that the quantities

$$V_r(x,t) = \sum_{m=1}^{N} (v_m)^r u_m(x,t), \quad r = 0,1,2,\ldots,N, \tag{2.13a}$$

satisfy the equations

$$\frac{\partial V_r}{\partial t} + \frac{\partial V_{r+1}}{\partial x} = V_1 V_r - V_0 V_{r+1}, \quad r = 0,1,2,\ldots,N-1. \tag{2.13b}$$

Note that for $r = 0$ this yields the local conservation law

$$\frac{\partial V_0}{\partial t} + \frac{\partial V_1}{\partial x} = 0, \tag{2.13c}$$

while $V_N(x,t)$ is, as implied by (2.13a), a linear combination of the N V_r's with $r = 0,1,2,\ldots,N-1$,

$$V_N(x,t) = \sum_{r=0}^{N-1} \gamma_r^{(N)} V_r(x,t). \tag{2.13d}$$

The quantities $\gamma_r^{(N)}$ are evaluated in Appendix E.

III. LOCALIZED SOLUTIONS: A NOVEL SOLITON PHENOMENOLOGY

In this section we discuss the localized solutions of (2.6), characterized by the condition to vanish asymptotically at both ends ($x \to \pm\infty$), or more precisely by the condition to be integrable,

$$\int_{-\infty}^{+\infty} dx\, u_n(x,t) = b_n(t), \tag{3.1a}$$

$$|b_n(t)| < \infty. \tag{3.1b}$$

It is sufficient (see below) that this condition, (3.1b), hold at any one time, to hold for all time.

It is then expedient to set $C = 1$ and $x_0 = -\infty$, so that (2.10) and (2.11) read

$$u_n(x,t) = w_n(x - v_n t)\Big(1 + \sum_{m=1}^{N} \int_{-\infty}^{x - v_m t} dx' w_m(x')\Big)^{-1}, \tag{3.2}$$

$$w_n(x) = u_n(x,0)\exp\Big[\int_{-\infty}^{x} dx' \sum_{m=1}^{N} u_m(x',0)\Big]. \tag{3.3}$$

In fact, the second of these equations is a special case of the more general formula

$$w_n(x - v_n t) = u_n(x,t)\exp\Big[\int_{-\infty}^{x} dx' \sum_{m=1}^{N} u_m(x',t)\Big], \tag{3.4}$$

that can be easily derived from (3.2) (see Appendix A). Note that this equation implies that the nonlinear evolution equations (2.6) possess the N (nonlocal) conserved quantities

$$c_n = \int_{-\infty}^{+\infty} dx\, u_n(x,t)\exp\Big[\int_{-\infty}^{x} dx' \sum_{m=1}^{N} u_m(x',t)\Big]. \tag{3.5}$$

More generally, conserved quantities are also provided by the more general formula

$$C_n = \int_{-\infty}^{+\infty} dx\, F_n\Big\{u_n(x,t) \times \exp\Big[\int_{-\infty}^{x} dx' \sum_{m=1}^{N} u_m(x',t)\Big]\Big\}, \tag{3.6a}$$

where the functions $F_n(z)$ are arbitrary [except for the condition

$$F_n(0) = 0, \tag{3.6b}$$

as required in order that the integrals on the rhs of (3.6a) converge].

Note moreover that (3.5) implies the relation

$$\sum_{n=1}^{N} c_n = \exp\Big[\sum_{n=1}^{N} b_n(t)\Big] - 1, \tag{3.7}$$

with the quantities $b_n(t)$ defined by (3.1a). Hence the sum

$$B = \sum_{n=1}^{N} b_n(t) = \sum_{n=1}^{N} \int_{-\infty}^{+\infty} dx\, u_n(x,t) \tag{3.8}$$

is a constant of the motion; as can also be evinced directly from the local conservation law (2.13c).

Let us now introduce the functions $u_n^{(\pm)}(x)$ via the formula

$$\lim_{t\to\pm\infty}[u_n(x+v_nt,t)-u_n^{(\pm)}(x)]=0, \tag{3.9a}$$

whose consistency with (3.1)–(3.3) is obvious (see below), and whose significance is perhaps best understood by rewriting this formula, (3.9a), in the following equivalent, if less rigorous, form:

$$u_n(x,t) \underset{t\to\pm\infty}{\to} u_n^{(\pm)}(x-v_nt). \tag{3.9b}$$

It is moreover convenient to define the $2N$ constants

$$b_n^{(\pm)}=\int_{-\infty}^{+\infty}dx\,u_n^{(\pm)}(x); \tag{3.10a}$$

there clearly hold the relations [see (3.1a)]

$$b_n^{(\pm)}=\lim_{t\to\pm\infty}b_n(t). \tag{3.10b}$$

It is now easy (see Appendix B) to obtain the formula

$$u_n^{(+)}(x)=u_n^{(-)}(x)\Big(1+a_n\times\exp\Big[-\int_{-\infty}^{x}dx'\,u_n^{(-)}(x')\Big]\Big)^{-1}, \tag{3.11a}$$

$$u_n^{(+)}(x)=\frac{d}{dx}\ln\Big\{a_n+\exp\Big[\int_{-\infty}^{x}dx'\,u_n^{(-)}(x')\Big]\Big\}, \tag{3.11b}$$

with

$$a_n=\exp\Big(\sum_{m=1}^{n}b_m^{(-)}\Big)+\exp\Big(-\sum_{m=n+1}^{N}b_m^{(-)}\Big)-\exp(b_n^{(-)})-1. \tag{3.12a}$$

In writing this formula, and throughout this paper, we conventionally set to zero the value of a sum if its lower limit exceeds its upper limit; hence, (3.12a) implies in particular

$$a_N=\exp\Big[\sum_{m=1}^{N}b_m^{(-)}\Big]-\exp[b_N^{(-)}], \tag{3.12b}$$

as well of course as

$$a_1=\exp\Big[-\sum_{m=2}^{N}b_m^{(-)}\Big]-1. \tag{3.12c}$$

Let us emphasize that, to obtain this formula, we have used the condition (1.2) (see Appendix B).

It is easily seen that (3.11b) implies, via (3.10a), the formula

$$b_n^{(+)}=\ln\{[a_n+\exp(b_n^{(-)})]/[a_n+1]\}, \tag{3.13a}$$

namely, via (3.12a)

$$b_n^{(+)}=\ln\Big\{\Big[\exp\Big(\sum_{m=1}^{n}b_m^{(-)}\Big)+\exp\Big(-\sum_{m=n+1}^{N}b_m^{(-)}\Big)-1\Big]\Big[\exp\Big(\sum_{m=1}^{n}b_m^{(-)}\Big)+\exp\Big(-\sum_{m=n+1}^{N}b_m^{(-)}\Big)-\exp(b_n^{(-)})\Big]^{-1}\Big\}, \tag{3.13b}$$

$$b_n^{(+)}=\ln\Big\{1+[1-\exp(-b_n^{(-)})]\Big[\exp\Big(\sum_{m=1}^{n-1}b_m^{(-)}\Big)+\exp\Big(-\sum_{m=n}^{N}b_m^{(-)}\Big)-1\Big]^{-1}\Big\}, \tag{3.13c}$$

$$b_n^{(+)}=b_n^{(-)}+\ln\Big\{\Big[\exp\Big(\sum_{m=1}^{n-1}b_m^{(-)}\Big)+\exp\Big(-\sum_{m=n}^{N}b_m^{(-)}\Big)-\exp(-b_n^{(-)})\Big]\times\Big[\exp\Big(\sum_{m=1}^{n}b_m^{(-)}\Big)+\exp\Big(-\sum_{m=n+1}^{N}b_m^{(-)}\Big)-\exp(b_n^{(-)})\Big]^{-1}\Big\}. \tag{3.13d}$$

To understand the significance of these remarkable formulas, it is appropriate to first consider the $N=2$ case. Then

$$a_1=\exp(-b_2^{(-)})-1, \tag{3.14a}$$

$$a_2=\exp(b_2^{(-)})[\exp(b_1^{(-)})-1], \tag{3.14b}$$

$$b_1^{(+)}=b_1^{(-)}+\beta, \tag{3.15a}$$

$$b_2^{(+)}=b_2^{(-)}-\beta, \tag{3.15b}$$

$$\beta=\ln[\exp(b_2^{(-)})+\exp(-b_1^{(-)})-\exp(b_2^{(-)}-b_1^{(-)})]. \tag{3.15c}$$

Note incidentally that the last three formulas imply the relation

$$b_1^{(+)}+b_2^{(+)}=b_1^{(-)}+b_2^{(-)}, \tag{3.16a}$$

which is clearly a special case of the more general formula

$$\sum_{n=1}^{N}b_n^{(+)}=\sum_{n=1}^{N}b_n^{(-)}, \tag{3.16b}$$

that follows from (3.13); and which is itself clearly a consequence of the time independence of B [see (3.8) and (3.10b)].

The second fact that is instrumental to understand this phenomenology is the observation that the recursion relation

$$f_{s+1}(x)=f_s(x)\Big(1+\eta_s\exp\Big[-\int_{-\infty}^{x}dx'\,f_s(x')\Big]\Big)^{-1},\quad s=0,1,2,..., \tag{3.17a}$$

$$f_{s+1}(x)=\frac{d}{dx}\ln\Big\{\eta_s+\exp\Big[\int_{-\infty}^{x}dx'\,f_s(x')\Big]\Big\},\quad s=0,1,2,..., \tag{3.17b}$$

can be solved to yield

$$f_s(x)=f_0(x)\Big(1+\alpha_s\exp\Big[-\int_{-\infty}^{x}dx'\,f_0(x')\Big]\Big)^{-1},\quad s=0,1,2,..., \tag{3.18a}$$

$$f_s(x)=\frac{d}{dx}\ln\Big\{\alpha_s+\exp\Big[\int_{-\infty}^{x}dx'\,f_0(x')\Big]\Big\},\quad s=0,1,2,..., \tag{3.18b}$$

as can be easily verified; namely, the functions $f_s(x)$,

$s=0,1,2,\ldots$, defined by the recursion relations (3.17), all have the same functional form,

$$f_s(x)=\varphi(x,\alpha_s),\quad s=0,1,2,\ldots, \tag{3.19a}$$

$$\varphi(x,\alpha)=f_0(x)\Big(1+\alpha\exp\Big[-\int_{-\infty}^{x}dx'\,f_0(x')\Big]\Big)^{-1}, \tag{3.19b}$$

with the constants α_s determined by the recursion relations

$$\alpha_{s+1}=\eta_s+(1+\eta_s)\alpha_s,\quad \alpha_0=0,\quad s=0,1,2,\ldots, \tag{3.20a}$$

whose explicit solution reads

$$\alpha_s=\sum_{r=0}^{s-1}\eta_r\prod_{q=r+1}^{s-1}(1+\eta_q),\quad s=0,1,2,\ldots . \tag{3.20b}$$

Let us recall that, by convention, a sum vanishes if its lower limit exceeds its upper limit; and by the same token we assume, here and hereafter, that a product takes the value unity if its lower limit exceeds its upper limit; so that (3.20b) yields $\alpha_0=0$ and $\alpha_1=\eta_0$ [consistently with (3.18a) and (3.17a)].

Note that, with the definition

$$\beta_s=\int_{-\infty}^{+\infty}dx\,f_s(x),\quad s=0,1,2,\ldots, \tag{3.21a}$$

we obtain [from (3.18b)]

$$\beta_s=\ln\{[\alpha_s+\exp(\beta_0)]/(\alpha_s+1)\}, \tag{3.21b}$$

and from (3.17b)

$$\beta_{s+1}=\ln\{[\eta_s+\exp(\beta_s)]/(\eta_s+1)\}. \tag{3.21c}$$

We are now in the position to describe the general behavior of the localized solutions of the system of nonlinear evolution PDE's (2.6).

In the remote past ($t\sim-\infty$) each function $u_n(x,t)$ has the form $u_n^{(-)}(x-v_nt)$ [where each $u_n^{(-)}(y)$ is a localized function, which can be assigned arbitrarily]; hence each of them travels with its own constant velocity v_n, without changing its profile; and since the velocities are different (and $|t|$ is large), these profiles are widely spaced, with $u_N(x,t)$ localized farther to the left, then $u_{N-1}(x,t)$ to its right, then $u_{N-2}(x,t)$ and so on [see (1.2)]. Note that this picture is consistent with the evolution equations (2.6), since these nonlinear evolution equations reduce to the linear wave equations $(\partial/\partial t+v_n\,\partial/\partial x)u_n(x,t)=0$ whenever their right-hand sides vanish, as is the case when the functions $u_m(x,t)$, $m=1,2,\ldots,N$, do not overlap.

As time progresses, the waves $u_n(x,t)$ associated with larger values of n [which were in the remote past located more to the left, and which move faster towards the right; see (2.12)], catch up with the waves associated with smaller values of n and overtake them; in the process, the nonlinear character of the evolution comes into play, causing different waves to interact as they cross each other.

The final outcome of the process (namely, the picture in the remote future, $t\sim+\infty$) shows again N widely separated waves, each traveling with its own constant velocity without change of shape, $u_n(x,t)\approx u_n^{(+)}(x-v_nt)$; but now the wave $u_N(x,t)$ is farther to the right, followed by $u_{N-1}(x,t)$ and so on; and as time proceeds further, these waves continue to separate and do not interact any more.

The effect on each wave of its interaction with the other $N-1$ waves (which it has overtaken, or by which it has been overtaken) is to change its profile from $u_n^{(-)}(x-v_nt)$ to $u_n^{(+)}(x-v_nt)$, according to the formulas (3.11) and (3.12). Hence the profile in the remote future, $u_n^{(+)}(y)$, is essentially determined by the profile of the same wave in the remote past, $u_n^{(-)}(y)$; the influence of all the other waves is only felt through the constant a_n, see (3.11), whose value is given by (3.12) in terms of the integrals $b_m^{(-)}$ of the profiles of the N waves in the remote past [see (3.10) and (3.9)]. Note in particular that the final effect does not depend on the localization of the waves in the remote past, nor on their velocities v_n; it depends only on their ordering in the remote past and future [as determined by the inequalities (1.2)], and on the integrals $b_n^{(-)}$ of their profiles in the remote past (the values of these integrals is of course independent of the localization of these profiles and of the velocity with which they move).

The final outcome we have described may of course result from a sequence of $\frac{1}{2}N(N-1)$ essentially distinct pair crossings, the effect of each of them being then accounted for by the formulas (3.14) and (3.15) [or, equivalently, (1.6)–(1.8)], and their combined result being calculable according to the formulas (3.17)–(3.20). The fact that the final outcome is independent of the order in which such pair encounters occur is a nontrivial feature of these formulas; the diligent reader may wish to verify this in the $N=3$ case, where two distinct sequences of pair encounters are possible, namely those characterized, sequentially in time, by the following orderings (from left to right) of the three waves: $(321)\to(312)\to(132)\to(123)$, $(321)\to(231)\to(213)\to(123)$ (of course throughout such a computation the changes in the values of the integrals over the profiles resulting from each encounter must be taken into account). Note that the order in which the different encounters occur is determined by the initial localization of the waves as well as by their velocities; and these parameters determine as well whether the interaction indeed develops only through a sequence of pair encounters, or instead also features multiple encounters (this depends moreover on how wide is the profile of each wave).

Let us emphasize again that the independence of the final outcome from these features of the actual time evolution is remarkable; it justifies (in our opinion) the use of the term "solitons" to denote these localized solutions.

Note however that the effect of the collisions on the solitons is not merely to shift their positions; it induces a change of profile. One might wonder whether, for some special profile, this change might reduce to a shift in position; the results of Appendix C imply that, for localized solutions, this cannot happen.

Our discussion so far has been mainly focused on the time evolution of the solution from the remote past to the remote future; but these findings are of course also relevant to understand the time evolution in the framework of the Cauchy problem, namely in terms of data given, say, at the initial time $t=0$. The relevant formulas are (3.2), (3.3), (3.9), and (B2), (B3) with (B1a) and (3.3). The outcome is sufficiently clear, in terms of the preceding discussion, not to require any additional elaboration.

Let us end with some quantitative comments.

If all the functions $u_n^{(-)}(x)$ [or equivalently, in the context of the Cauchy problem, all the functions $u_n(x,0)$] are non-negative, it is easily seen that the same is true for all the functions $u_n^{(+)}(x)$ or indeed, more generally, $u_n(x,t)$. It is moreover easily seen that $a_n > -1$ [see (3.12a); and note that in this case all the $b_m^{(-)}$'s are positive], and therefore the functions $u_n^{(+)}(x)$ are regular for all real values of x [see (3.11a); we assume of course all the functions $u_n^{(-)}(x)$, or, equivalently, $u_n(x,0)$, to be regular to begin with]. More generally, the fact that the functions $u_n(x,t)$ are regular (and non-negative) for all real values of x and t is an immediate consequence of (3.1)–(3.3), under the hypothesis that all the functions $u_n(x,0)$ [or, equivalently, $u_n^{(-)}(x)$] are themselves regular and non-negative. Note on the other hand that, while a_N is positive, a_1 is negative [see (3.12b) and (3.12c)].

If instead some of the initial data $u_n(x,0)$ are nonpositive [as it would be appropriate if some of the parameters β_n in (1.3) were negative; see (2.7)], then it might happen that, even though all the $u_n(x,0)$ are regular, all the functions $u_n(x,t)$ become singular at some finite time t_c [this would occur through the vanishing of their denominator, which is common to all of them; see (3.2) or (2.8)]. It is, however, clear from (3.2) that this phenomenon may, but need not, happen; for instance if only one of the $u_n(x,0)$ is nonpositive while all the others are non-negative, say $u_\nu(x,0)\leqslant 0$, $u_n(x,0)\geqslant 0$ for $n\neq\nu$, then the inequality

$$\int_{-\infty}^{+\infty} dx|u_\nu(x,0)|\exp\left[\sum_{m=1}^{N}\int_{-\infty}^{x} dx'\, u_m(x',0)\right] < 1 \tag{3.22}$$

is clearly sufficient to exclude the emergence of any singularity at any, future or past, time [see (3.2) and (3.3)].

IV. KINKLIKE SOLUTIONS

In this section we consider solutions of (2.6) that tend to finite values asymptotically,

$$\lim_{x\to\pm\infty}[u_n(x,t)] = u_n(\pm\infty,t). \tag{4.1}$$

The simpler instance of this kind is the uniform (space-independent) solution,

$$u_n(x,t) = u_n(t). \tag{4.2a}$$

It reads

$$u_n(t) = p\, u_n(0)\left(\sum_{m=1}^{N} u_m(0)\exp[p(v_n - v_m)t]\right)^{-1}, \tag{4.2b}$$

$$p = \sum_{n=1}^{N} u_n(0) = \sum_{n=1}^{N} u_n(t). \tag{4.2c}$$

The fact that this solution contains the N arbitrary constants $u_n(0)$ implies that it is the most general of its kind. It is easy to obtain this solution setting $u_n(x,0)=u_n(0)$ in (2.10) and (2.11).

Another simple solution is the *traveling wave*

$$u_n(x,t) = \tilde{u}_n(x - Vt). \tag{4.3a}$$

It reads

$$\tilde{u}_n(y) = \alpha A_n \exp\left[\frac{\alpha y}{(v_n - V)}\right] \times\left(\sum_{m=1}^{N}(v_m - V)A v_m \exp\left[\frac{\alpha y}{(v_m - V)}\right]\right)^{-1}. \tag{4.3b}$$

Here the overall speed V is arbitrary, except for the restriction

$$V\neq v_n; \tag{4.3c}$$

and the $N+1$ constants α and A_m are also arbitrary. Note however that, for any given V, the solution (4.3b) depends effectively on N arbitrary constants, since a change of scale of the constants A_m does not affect it.

This solution may, but need not, become singular at some real value of y, due to the vanishing of the denominator. For instance, the restrictions

$$(v_m - V)A_m \geqslant 0, \quad m = 1,2,\ldots,N, \tag{4.3d}$$

are sufficient to guarantee that this solution be regular for all (real) values of y (we assume of course that all the constants α, V, A_m, and v_m are real).

In Appendix D we indicate how this solution has been obtained, and we show that it is the most general solution of its kind.

Special cases of this solution are obtained if some of the constants A_n vanish. In particular, the simpler nontrivial solution is obtained if only two of these N constants do not vanish, say $A_n = 0$ if $n\neq\mu,\nu$. It reads

$$\tilde{u}_n(y) = 0, \quad \text{if } n\neq\mu,\nu, \tag{4.4a}$$

$$\tilde{u}_\mu(y) = p_\mu/\{1+\exp[(p_\nu - p_\mu)(y - y_0)]\}, \tag{4.4b}$$

$$\tilde{u}_\nu(y) = p_\nu/\{1+\exp[(p_\mu - p_\nu)(y - y_0)]\}, \tag{4.4c}$$

where we have set

$$y_0 = (p_\mu - p_\nu)^{-1}\ln[(p_\mu A_\nu)/(p_\nu A_\mu)], \tag{4.5a}$$

$$p_\mu = \alpha/(v_\mu - V), \quad p_\nu = \alpha/(v_\nu - V), \tag{4.5b}$$

implying

$$V = (p_\mu v_\mu - p_\nu v_\nu)/(p_\mu - p_\nu). \tag{4.5c}$$

As we will see below, this solution is the prototypical one representing a *kink*. Note that it is certainly regular (if y_0 is real), and that it has the following asymptotic values:

$$\tilde{u}_\mu(-\infty) = p_\mu, \quad \tilde{u}_\mu(+\infty) = 0, \tag{4.6a}$$

$$\tilde{u}_\nu(-\infty) = 0, \quad \tilde{u}_\nu(+\infty) = p_\nu, \tag{4.6b}$$

provided

$$p_\nu > p_\mu > 0. \tag{4.6c}$$

Note that these asymptotic relations are special cases of the general formula, applicable to the solution (4.3b),

$$\tilde{u}_n(-\infty) = \delta_{n\mu}p_\mu, \quad \text{if } A_\mu\neq 0 \text{ and } A_m = 0 \quad \text{for } m<\mu, \tag{4.7a}$$

$$\tilde{u}_n(+\infty) = \delta_{n\nu}p_\nu, \quad \text{if } A_\nu\neq 0 \text{ and } A_m = 0 \quad \text{for } m>\nu. \tag{4.7b}$$

Here we are of course using the notation (4.5b), and we have moreover assumed that α is negative, $\alpha<0$, so that the restrictions in (4.7) (which imply $\nu > \mu$), together with (1.2), imply (4.6c).

Note that (4.7) implies that the traveling wave solution (4.3b) has a kinklike shape; only one of the waves does not vanish to the left, and another one to the right; generally these are characterized by the extreme (i.e., largest and smallest) values of the index n, among the waves which are effectively present.

For $V=0$, the formulas (4.3) provide the (most general) static solution of (2.6).

Let us now exhibit a third class of solutions, that encompasses those discussed thus far. It is obtained by setting in (2.10)

$$w_n(x) = \widetilde{w}_n(x) + \sum_r A_{nr} \exp(p_{nr} x), \tag{4.8}$$

with the constants A_{nr} and p_{nr} arbitrary and the functions $\widetilde{w}_n(x)$ vanishing (or, more precisely, integrable) at both ends,

$$c_n = \int_{-\infty}^{+\infty} dx\, \widetilde{w}_n(x), \tag{4.9a}$$

$$|c_n| < \infty. \tag{4.9b}$$

Hence this solution reads as follows:

$$\begin{aligned} u_n(x,t) = &\Big\{\widetilde{w}_n(x - v_n t) + \sum_r A_{nr} \exp[p_{nr}(x - v_n t)]\Big\} \\ &\times\Bigg(A_0 + \sum_{m=1}^{N} \Big\{\widetilde{W}_n(x - v_m t) \\ &+ \sum_r \Big(\frac{A_{mr}}{p_{mr}}\Big)\exp[p_{mr}(x - v_m t)]\Big\}\Bigg)^{-1}, \end{aligned} \tag{4.10a}$$

with

$$A_0 = C - \sum_{m=1}^{N} \sum_r \Big(\frac{A_{mr}}{p_{mr}}\Big)\exp(p_{mr} x_0) \tag{4.10b}$$

and

$$\widetilde{W}_n(x) = \int_{x_0}^{x} dx'\, \widetilde{w}_n(x'). \tag{4.10c}$$

The asymptotic behavior of this solution as $x \to \pm\infty$ is easily ascertained from these formulas, taking into account the exploding or vanishing asymptotics of the exponentials and the vanishing or constant asymptotics of the functions $\widetilde{w}_n(x)$ and $\widetilde{W}_n(x)$; it is clear from such an analysis that the generic solution of this kind is asymptotically limited. Indeed, if all the parameters p_{nr} are different among themselves, then clearly

$$u_n(\pm\infty,t) = \delta_{n\nu^{(\pm)}} p^{(\pm)}, \tag{4.10d}$$

where $p^{(+)}$ is the largest of the p_{nr}'s and $p^{(-)}$ is the smallest, and $\nu^{(+)}$, $\nu^{(-)}$ are the corresponding values of the index n.

By the same token it is easy to analyze the behavior of this solution in the remote past and future. To this end it is expedient to set $x = y + Vt$, and to analyze the behavior of the solution for fixed values of y and V, and for diverging t (graphical techniques, analogous to those employed, in an analogous context, in Ref. 4, may prove expedient to perform such an investigation). It is clear that the generic solution of this type yields, in the remote past and future, a solution whose profile is a sequence of regions where one of the u_n's has a constant nonvanishing value and all the others vanish, say

$$u_n = \delta_{n\nu} p_\nu. \tag{4.11a}$$

Two adjacent regions, one characterized by the solution (4.11a) and the other, say, characterized by

$$u_n = \delta_{n\mu} p_\mu, \tag{4.11b}$$

are separated by a kink that travels with the constant velocity (4.5c) and interpolates (as a "boundary layer") among these two constant solutions, according to the formulas (4.3a), (4.4), and (4.6). A different (but easily computable; see below) shape of the boundary layer may also occur, if its velocity coincides with one of the v_n's.

Note that the solution (4.10) might, but need not, be singular for real values of x and t; assuming all the functions $\widetilde{w}_m(x)$ to be regular [as suggested by (4.9)], this question hinges of course on the possibility that the denominator in the rhs of (4.10a) (which is common to all the u_n's) vanish.

The motivation for focusing on the solution (4.10) is because such a solution is clearly yielded [see (2.10) and (2.11)] by the Cauchy problem with initial data $u_n(x,0)$ having the asymptotic behavior ($x \to s\infty$, $s = \pm$)

$$\begin{aligned} u_n(x,0) = u_n(s\infty,0) &+ \sum_r B_{nr}^{(s)} \exp(-|q_{nr}^{(s)} x|) \\ &+ o[\exp(-p^{(s)} x)], \quad s = \pm, \end{aligned} \tag{4.12a}$$

with

$$p^{(s)} = \sum_{n=1}^{N} u_n(s\infty,0), \quad s = \pm. \tag{4.12b}$$

Of course, if $p^{(+)}$ is negative and $p^{(-)}$ positive, no exponential corrections appear in the rhs of this asymptotic formula [namely, all the coefficients $B_{nr}^{(s)}$ vanish; in general, $B_{nr}^{(+)} = 0$ if $p^{(+)} < 0$, $B_{nr}^{(-)} = 0$ if $p^{(-)} > 0$; this case is treated below].

The fact that (4.12) yield (4.8)–(4.10) [via (2.10)–(2.12)] is easily verified; note however that the quantities p_{nr} and A_{nr} in (4.8), although uniquely defined by the initial data $u_m(x,0)$, depend generally on fine details (of the asymptotic behavior) of these functions; this may cause the solutions to depend *sensitively* on the initial data. Moreover, a *generic* choice of the parameters p_{nr} and A_{nr} in (4.8) is generally far from being yielded by a *generic* choice of the parameters $u_n(s\infty,0)$, $B_{nr}^{(s)}$, and $q_{nr}^{(s)}$ in (4.12a). Since the relevance of these facts has been elaborated in some detail (in a different, but analogous, context) in Ref. 4(b), we forsake any further discussion of these points and we limit our treatment here to an analysis of the solution of (2.6) yielded, in the context of the Cauchy problem, by initial data $u_n(x,0)$ that are positive and yield no exponential corrections in (4.12a),

$$u_n(x,0) \geqslant 0, \tag{4.13a}$$

$$u_n(\pm\infty,0) \geqslant 0, \tag{4.13b}$$

$$p^{(\pm)} \equiv \sum_{n=1}^{N} u_n(\pm\infty,0) > 0, \tag{4.13c}$$

$$\lim_{x\to+\infty}\{[u_n(x,0)-u_n(+\infty,0)]\exp(p^{(+)}x)\}=0. \tag{4.13d}$$

Then the formulas (2.10) and (2.11) [or (3.2) and (3.3)] may be replaced by

$$u_n(x,t)=w_n(x-v_nt)\left(\sum_{m=1}^N\int_{-\infty}^{x-v_mt}dx'\,w_m(x')\right)^{-1}, \tag{4.14}$$

$$w_n(x)=u_n(x,0)\exp\left\{p^{(-)}x+\int_{-\infty}^{x}dx'\sum_{m=1}^N \times[u_m(x',0)-u_m(-\infty,0)]\right\}. \tag{4.15}$$

Note that the consistency [both among themselves and with the asymptotic behavior (4.1)] of these expressions is implied by the exponential vanishing of the functions $w_n(x)$ to the left,

$$\lim_{x\to-\infty}[w_n(x)\exp(-p^{(-)}x)]=u_n(-\infty,0), \tag{4.16a}$$

and their exponential divergence to the right

$$\lim_{x\to+\infty}[w_n(x)\exp(-p^{(+)}x)]=u_n(+\infty,0)\lambda, \tag{4.16b}$$

$$\lambda=\exp\left\{\int_{-\infty}^0 dx'\sum_{m=1}^N[u_m(x',0)-u_m(-\infty,0)] +\int_0^\infty dx'\sum_{m=1}^N[u_m(x',0)-u_m(+\infty,0)]\right\}. \tag{4.16c}$$

It is now clear that these formulas imply the following special (nongeneric!) case of (4.8):

$$w_n(x)=\tilde{w}_n(x)+A_n\exp(p^{(+)}x), \tag{4.17a}$$

$$A_n=u_n(+\infty,0)\lambda, \tag{4.17b}$$

of course with $\tilde{w}_n(x)$ integrable at both ends [see (4.9)]. The corresponding expression of $u_n(x,t)$ is obtained from (4.14),

$$u_n(x,t)=\{\tilde{w}_n(x-v_nt)+A_n\exp[p^{(+)}(x-v_nt)]\} \times\left(\sum_{m=1}^N\left\{\int_{-\infty}^{x-v_mt}dx'\,\tilde{w}_m(x') +\left(\frac{A_m}{p^{(+)}}\right)\exp[p^{(+)}(x-v_mt)]\right\}\right)^{-1}. \tag{4.18}$$

It is easily seen that these formulas imply

$$u_n(s\infty,t) =p^{(s)}u_n(s\infty,0) \times\left(\sum_{m=1}^N u_m(s\infty,0)\exp[p^{(s)}(v_n-v_m)t]\right)^{-1}, \quad s=\pm. \tag{4.19a}$$

For $s=+$, this expression is obtained from (4.18), (4.13c), (4.9), and (4.17b); for $s=-$, it is obtained directly from (4.14), (4.16), and (4.13c). The consistency of this formula at $t=0$ is easily verified [via (4.12b)]. Note that this expression, (4.19a), satisfies, as it should [see (2.6) and (4.1)], the system of ODE's

$$\dot{u}_n(s\infty,t)=u_n(s\infty,t)\sum_{m=1}^N(v_m-v_n)u_m(s\infty,t),\quad s=\pm \tag{4.19b}$$

[see (4.2a) and (4.2b)], as well as the conservation law

$$p^{(s)}\equiv\sum_{n=1}^N u_n(s\infty,0)=\sum_{n=1}^N u_n(s\infty,t),\quad s=\pm. \tag{4.19c}$$

Let us now investigate the behavior of this solution, (4.18), in the remote past and future. To this end we set

$$\lim_{t\to\pm\infty}[u(y+Vt,t)]=u^{(\pm)}(y), \tag{4.20}$$

and we find

$$u_n^{(-)}(y)=\delta_{nN}p^{(+)},\quad\text{for } V<v_N, \tag{4.21a}$$

$$u_n^{(-)}(y)=\delta_{nN}w_N(y) \times\left(\int_{-\infty}^y dx\,w_N(x)\right)^{-1},\quad\text{for } V=v_N, \tag{4.21b}$$

$$u_n^{(-)}(y)=\delta_{nN}p^{(-)},\quad\text{for } V>v_N; \tag{4.21c}$$

$$u_n^{(+)}(y)=\delta_{n1}p^{(+)},\quad\text{for } V>v_1, \tag{4.22a}$$

$$u_n^{(+)}(y)=\delta_{n1}w_1(y)\left(\int_{-\infty}^y dx\,w_1(x)\right)^{-1},\quad\text{for } V=v_1, \tag{4.22b}$$

$$u_n^{(+)}(y)=\delta_{n1}p^{(-)},\quad\text{for } V<v_1. \tag{4.22c}$$

To obtain these results we have of course used the inequalities (1.2); (4.21a) and (4.22a) follow from (4.18), (4.13c), and (4.9); (4.21b), (4.21c) and (4.22b), (4.22c) follow from (4.14), (4.16), and (4.13c). The functions $w_1(y)$ and $w_N(y)$ are of course given, in terms of the initial data $u_n(x,0)$, by (4.15).

Hence, in the remote future, only the wave with $n=1$ does not vanish, and it has to the left the constant value $p^{(-)}$, to the right the constant value $p^{(+)}$; the two regions where $u_1(x,t)$ (for $t\approx+\infty$) is constant are separated by a boundary layer that travels with the constant velocity $V=v_1$ and whose shape is given by the formula (4.22b) [it is easy to verify that this formula has the limiting values $p^{(\pm)}$ as y tends to $\pm\infty$; this follows from (4.17a), (4.13c), (4.9), and (4.17b) for $y\to+\infty$, and from (4.16) and (4.13c) for $y\to-\infty$]. Note the consistency of these results with those implied by (4.19), namely,

$$u_n(+\infty,+\infty)=\delta_{n1}p^{(+)}, \tag{4.23a}$$

$$u_n(-\infty,+\infty)=\delta_{n1}p^{(-)}. \tag{4.23b}$$

In the remote past the results are analogous, with the index $n=1$ replaced by $n=N$.

Note that knowledge of the behavior of the solutions in the remote past (or, for that matter, in the remote future), as given by these equations, is insufficient to determine the solution for all time; this is in contrast to the case of localized solutions (i.e., vanishing as $x\to\pm\infty$), treated in Sec. III. This fact provides another indication of the sensitive dependence of the time evolution (especially the long time behavior) on the initial data.

Let us end this section with a treatment of the case when $p^{(+)}$ is negative and $p^{(-)}$ positive,

$$p^{(+)}<0, \tag{4.24a}$$

$$p^{(-)}>0. \tag{4.24b}$$

The formulas (4.14) and (4.15) are then again applicable, and they moreover imply that the function $w_n(x)$ vanish exponentially at both ends,

$$\lim_{x\to-\infty}[w_n(x)\exp(-p^{(-)}x)]=u_n(-\infty,0), \tag{4.24c}$$

$$\lim_{x\to+\infty}[w_n(x)\exp(-p^{(+)}x)]=u_n(+\infty,0)\lambda. \tag{4.24d}$$

Here $p^{(+)}$ and $p^{(-)}$ are of course always related to the initial data by (4.12b), and λ is defined by (4.16c).

In this case therefore the analysis of the behavior of the solution (4.14) becomes analogous to that given in Sec. III, except for the modifications implied by the differences that distinguish (4.14) from (3.2) and (4.15) from (3.3). We feel therefore justified in reporting the relevant formulas (in self-explanatory notation), without elaborating on their derivation or on their significance,

$$\lim_{t\to\pm\infty}[u(y+Vt,t)]=u^{(\pm)}(y); \tag{4.25}$$

for $V=v_n$,

$$u_n^{(-)}(y)=w_n(y)\left(\sum_{m=n+1}^{N}c_m+\int_{-\infty}^{y}dx\,w_n(x)\right)^{-1}, \tag{4.26a}$$

$$u_n^{(-)}(-\infty)=\delta_{nN}p^{(-)}, \tag{4.26b}$$

$$u_n^{(-)}(+\infty)=0; \tag{4.26c}$$

for $V>v_N$

$$u_n^{(-)}(y)=\delta_{nN}p^{(-)}; \tag{4.26d}$$

for $V<v_N$, $V\neq v_n$,

$$u_n^{(-)}(y)=0; \tag{4.26e}$$

for $V=v_n$,

$$u_n^{(+)}(y)=w_n(y)\left(\sum_{m=1}^{n-1}c_m+\int_{-\infty}^{y}dx\,w_n(x)\right)^{-1}, \tag{4.27a}$$

$$u_n^{(+)}(-\infty)=\delta_{n1}p^{(-)}, \tag{4.27b}$$

$$u_n^{(+)}(+\infty)=0; \tag{4.27c}$$

for $V<v_1$,

$$u_n^{(+)}(y)=\delta_{n1}p^{(-)}; \tag{4.27d}$$

for $V>v_1$, $V\neq v_n$,

$$u_n^{(+)}(y)=0. \tag{4.27e}$$

It is easily seen that the expression (4.27a) may be related to (4.26a) as follows: for $n<N$,

$$u_n^{(+)}(y)=u_n^{(-)}(y)\left(1+a_n\exp\left[\int_{-\infty}^{y}dx\,u_n^{(-)}(x)\right]\right)^{-1}, \tag{4.28a}$$

$$a_n=\left(\sum_{m=1}^{n-1}c_m-\sum_{m=n+1}^{N}c_m\right)\left(\sum_{m=n+1}^{N}c_m\right)^{-1}; \tag{4.28b}$$

for $n=N$,

$$u_N^{(+)}(y)=u_N^{(-)}(y)\times\left(1+a_N\exp\left[\int_{-\infty}^{y}dx\,u_N^{(-)}(x)\right]\right)^{-1}, \tag{4.29a}$$

$$a_N=\left(\sum_{m=1}^{n-1}c_m\right)(c_N)^{-1}. \tag{4.29b}$$

In this formula the quantities c_n are of course defined as follows:

$$c_n=\int_{-\infty}^{+\infty}dx\,w_n(x). \tag{4.30}$$

For $n<N$, they can be related to the integrals over the expressions (4.26a) of $u_n^{(-)}$,

$$b_n^{(-)}=\int_{-\infty}^{+\infty}dx\,u_n^{(-)}(x),\quad n<N, \tag{4.31}$$

as follows:

$$c_n=c_N\left\{\exp\left[\sum_{m=n}^{N-1}b_m^{(-)}\right]-\exp\left[\sum_{m=n+1}^{N-1}b_m^{(-)}\right]\right\},\quad n<N \tag{4.32}$$

[note that (4.26b), (4.26c), and (4.24b) imply that $u_n^{(-)}$ is integrable if $n<N$]. Hence the quantities a_n [see (4.28) and (4.29)] can also be expressed through these quantities:

$$a_n=\exp\left[\sum_{m=1}^{n}b_m^{(-)}\right]-\exp[b_m^{(-)}]-1,\quad n<N, \tag{4.33a}$$

$$a_N=\exp\left[\sum_{m=1}^{N-1}b_m^{(-)}\right]-1. \tag{4.33b}$$

Note that these formulas do not contain the (undefined) quantity $b_N^{(-)}$; and that, for $n<N$, they are obtained from (3.12a) by setting $b_N^{(-)}=+\infty$.

A main finding of this section is the crucial role played by the values (and in particular the signs) of the quantities $p^{(-)}$ and (especially) $p^{(+)}$, defined, in the context of the Cauchy problem, by (4.12b), or, more generally, by the formula (4.19c) applicable at any time. Note that these quantities may have any value, even if each of the initial data $u_n(x,0)$ has a constant sign, consistently with the condition (2.7); provided not all these signs are the same, and the functions u_n are not all localized, namely, do not all vanish asymptotically [see (4.1) and (4.12b) or (4.19c)]. And it is also possible (although, of course, not generically) that, when these conditions prevail, $p^{(-)}$ or $p^{(+)}$ (or both of them) vanish.

V. RATIONAL SOLUTIONS AND RELATED DYNAMICAL SYSTEMS

In this section we discuss rational solutions of (2.6), and related integrable dynamical systems.

The existence of such solutions is immediately implied by (2.10); indeed this formula yields, for any polynomial choice of the functions $w_n(x)$,

$$w_n(x)=\sum_{k=0}^{J}A_{nk}x^k, \tag{5.1}$$

the following rational solution of (2.6):

$$u_n(x,t)=\sum_k A_{nk}(x-v_n t)^k[F(x,t)]^{-1}, \qquad (5.2a)$$

$$F(x,t)=A_0+\sum_{m=1}^{N}\sum_k\left[\frac{A_{mk}}{k+1}\right](x-v_m t)^{k+1}, \quad (5.2b)$$

$$A_0=C-\sum_{m=1}^{N}\sum_k\left[\frac{A_{mk}}{k+1}\right]x_0^{k+1}. \qquad (5.2c)$$

And of course, in the context of the Cauchy problem, the initial (rational) datum $u_n(x,0)$ that is obtained by setting $t=0$ in (5.2a) reproduces the solution (5.2), that is rational (both in x and t) for all time.

With these rational solutions of (2.6) one can associate an (integrable) dynamical system via the positions

$$u_n(x,t)=u_n(\infty,t)+\sum_{j=1}^{J}\rho_{nj}(t)[x-\xi_j(t)]^{-1}. \quad (5.3)$$

Indeed, it is easily seen that the insertion of this *ansatz* in (2.6) yields the following evolution equations for the quantities $u_n(\infty,t)$, $\rho_{nj}(t)$, and $\xi_j(t)$:

$$\dot{\xi}_j(t)=\sum_{n=1}^{N}v_n\rho_{nj}(t), \qquad (5.4a)$$

$$\sum_{n=1}^{N}\rho_{nj}(t)=1, \qquad (5.4b)$$

$$\begin{aligned}\dot{\rho}_j(t)&=\sum_{m=1}^{N}(v_m-v_n)[u_n(\infty,t)\rho_{mj}(t)+u_m(\infty,t)\rho_{nj}(t)]\\&\quad+\sum_{m=1}^{N}(v_m-v_n)\sum_{k=1}^{J}[\xi_j(t)-\xi_k(t)]^{-1}\\&\quad\times[\rho_{nj}(t)\rho_{mk}(t)+\rho_{nk}(t)\rho_{mj}(t)],\end{aligned} \qquad (5.4c)$$

$$\dot{u}_n(\infty,t)=u_n(\infty,t)\sum_{m=1}^{N}(v_m-v_n)u_n(\infty,t). \qquad (5.4d)$$

Here, and always below, the superimposed dot denotes time differentiation, and the prime appended to a summation symbol indicates omission of the (singular) term with $k=j$ (likewise for products; see below).

The ansatz (5.3), with (5.4b), is of course consistent with the solution (5.2); in particular, the number of poles J coincides with the degree of the polynomial (5.2b),

$$F(x,t)=\sum_{j=0}^{J}f_j(t)x^j. \qquad (5.5)$$

To express the dynamical system (5.4) in a more interesting form, it is convenient to restrict attention to the subcase with

$$u_n(\infty,t)=0 \qquad (5.6)$$

[note the consistency of this restriction with (5.4d)], and to introduce the quantities

$$R_{rj}(t)=\sum_{n=1}^{N}(v_n)^r\rho_{nj}(t), \quad r=0,1,\ldots,N, \qquad (5.7a)$$

so that [see (5.4a) and (5.4b)]

$$R_{0j}(t)=1, \qquad (5.7b)$$

$$R_{1j}(t)=\dot{\xi}_j(t), \qquad (5.7c)$$

and

$$R_{Nj}(t)=\sum_{r=0}^{N-1}\gamma_r^{(N)}R_{rj}(t). \qquad (5.7d)$$

The last formula is obtained from (5.7a); the constants $\gamma_r^{(N)}$ are computed in Appendix E. In particular, for $N=2$,

$$\gamma_0^{(2)}=-v_1v_2, \quad \gamma_1^{(2)}=v_1+v_2, \qquad (5.8a)$$

and for $N=3$,

$$\begin{aligned}&\gamma_0^{(3)}=v_1v_2v_3, \quad \gamma_1^{(3)}=-(v_1v_2+v_2v_3+v_3v_1),\\&\gamma_2^{(3)}=v_1+v_2+v_3.\end{aligned} \qquad (5.8b)$$

Using these formulas, the dynamical system (5.4) can be reformulated in the following (more appealing) form:

$$\ddot{\xi}_j=\sum_{k=1}^{J}{}'(\xi_j-\xi_k)^{-1}(2\dot{\xi}_j\dot{\xi}_k-R_{2j}-R_{2k}), \quad j=1,2,\ldots,J, \qquad (5.9a)$$

$$\dot{R}_{rj}=\sum_{k=1}^{J}{}'(\xi_j-\xi_k)^{-1}(\dot{\xi}_jR_{rk}+\dot{\xi}_kR_{rj}-R_{r+1,j}-R_{r+1,k}),$$
$$j=1,2,\ldots,J, \quad r=2,3,\ldots,N-1, \qquad (5.9b)$$

with

$$R_{Nj}=\gamma_0^{(N)}+\gamma_1^{(N)}\dot{\xi}_j+\sum_{r=2}^{N-1}\gamma_r^{(N)}R_{rj}, \quad j=1,2,\ldots,J. \qquad (5.9c)$$

Here the dynamical variables are the J ξ_j's, $j=1,2,\ldots,J$ and the $J(N-2)$ R_{rj}'s, $r=2,3,\ldots,N-1$; $j=1,2,\ldots,J$.

In particular, for $N=2$ the system reads

$$\begin{aligned}\ddot{\xi}_j&=\sum_{k=1}^{J}{}'(\xi_j-\xi_k)^{-1}[2\dot{\xi}_j\dot{\xi}_k-(v_1+v_2)(\dot{\xi}_j+\dot{\xi}_k)\\&\quad+2v_1v_2],\end{aligned} \qquad (5.10a)$$

$$\begin{aligned}\ddot{\xi}_j&=\sum_{k=1}^{J}{}'(\xi_j-\xi_k)^{-1}[(\dot{\xi}_j-v_1)(\dot{\xi}_k-v_2)\\&\quad+(\dot{\xi}_j-v_2)(\dot{\xi}_k-v_1)],\end{aligned} \qquad (5.10b)$$

and for $N=3$,

$$\ddot{\xi}_j=\sum_{k=1}^{J}{}'(\xi_j-\xi_k)^{-1}(2\dot{\xi}_j\dot{\xi}_k-\eta_j-\eta_k), \qquad (5.11a)$$

$$\begin{aligned}\dot{\eta}_j&=\sum_{k=1}^{J}{}'(\xi_j-\xi_k)^{-1}[\dot{\xi}_j\eta_k+\dot{\xi}_k\eta_j-2\gamma_0^{(3)}\\&\quad-\gamma_1^{(3)}(\dot{\xi}_j+\dot{\xi}_k)-\gamma_2^{(3)}(\eta_j+\eta_k)],\end{aligned} \qquad (5.11b)$$

where we have set

$$R_{2j}(t)\equiv\eta_j(t). \qquad (5.11c)$$

It is therefore seen that, for $N=2$, the dynamical system (5.9) reproduces a known integrable system[3]; while for $N=3$ (and, *a fortiori*, for $N>3$), it yields integrable systems which are (to the best of our knowledge) new. It may also be easily verified that the system (5.10) is contained in (5.11); indeed the special solution of (5.11) characterized by the relation

$$\eta_j=(v_1+v_2)\dot{\xi}_j-v_1v_2 \qquad (5.12a)$$

[whose compatibility with (5.11) is easily verified by equating (5.11a) to (5.11b)], clearly sends (5.11) into (5.10). Incidentally, the position

$$\eta_j=a\dot{\xi}_j+b \qquad (5.12b)$$

is compatible with (5.11) *only* if

$$a=v_1+v_2,\quad b=-v_1v_2 \tag{5.12c}$$

[in which case (5.12b) yields (5.12a)], or in the other two cases that are obtained from (5.12c) by permuting cyclically the three indices 1, 2, and 3.

Note that in the system (5.9), the N constants $\gamma_r^{(N)}$ may be chosen arbitrarily; the corresponding quantities v_n are then identified (see Appendix E) as the N roots of the polynomial of degree N in v

$$p_N(v)=v^N-\sum_{r=0}^{N-1}\gamma_r^{(N)}v^r. \tag{5.13}$$

Clearly the system (5.9) features the following conservation laws:

$$\sum_{j=1}^{J}\ddot{\xi}_j=0, \tag{5.14a}$$

$$\sum_{j=1}^{J}\dot{R}_{rj}=0,\quad r=2,3,\ldots,N-1. \tag{5.14b}$$

It is easily seen that the system (5.9) has no real equilibrium configuration [if $\gamma_0^{(N)}\neq 0$, namely, if none of the v_n's vanishes; see (E6a)]. Indeed insertion of the position

$$\xi_j(t)=\bar{\xi}_j,\quad \dot{\xi}_j=0,\quad j=1,2,\ldots,J, \tag{5.15a}$$

$$R_{rj}(t)=\bar{R}_{rj},\quad \dot{R}_{rj}=0,\quad j=1,2,\ldots,J;\quad r=2,3,\ldots,N-1, \tag{5.15b}$$

in (5.9) yields

$$\sum_{k=1}^{J}{}'(\bar{\xi}_j-\bar{\xi}_k)^{-1}(\bar{R}_{rj}+\bar{R}_{rk})=0,$$
$$j=1,2,\ldots,J;\quad r=2,3,\ldots,N, \tag{5.15c}$$

namely [multiplying by $\gamma_r^{(N)}$, summing over r from 2 to N and using (5.9c), (5.15a), and (5.15c) with $r=N$]

$$\gamma_0^{(N)}\sum_{k=1}^{J}{}'(\bar{\xi}_j-\bar{\xi}_k)^{-1}=0,\quad j=1,2,\ldots,J. \tag{5.15d}$$

Clearly these relations cannot be satisfied if all the $\bar{\xi}_j$'s are real [indeed, if all the $\bar{\xi}_k$'s are real and $\bar{\xi}_j$ is the largest, or smallest, of them, all the terms in the sum (5.15d) have the same sign].

The system (5.9) may instead possess (similarity) solutions in which each ξ_j moves with its own constant speed, and the quantities R_{rj} are constant,

$$\xi_j(t)=\xi_0+\nu_j(t-t_0), \tag{5.16a}$$

$$R_{rj}(t)=\bar{R}_{rj},\quad \dot{R}_{rj}(t)=0. \tag{5.16b}$$

The quantities ν_j and $\dot{R}_{rj}$ must then satisfy the $(J-1)(N-1)$ equations

$$\sum_{k=1}^{J}{}'(\nu_j-\nu_k)^{-1}(2\nu_j\nu_k-\bar{R}_{2j}-\bar{R}_{2k})=0,$$
$$j=1,2,\ldots,J-1, \tag{5.16c}$$

$$\sum_{k=1}^{J}{}'(\nu_j-\nu_k)^{-1}(\nu_j\bar{R}_{rk}+\nu_k\bar{R}_{rj}-\bar{R}_{r+1\,j}-\bar{R}_{r+1,k})=0,$$
$$j=1,2,\ldots,J-1,\quad r=2,3,\ldots,N-1, \tag{5.16d}$$

with

$$\bar{R}_{Nj}=\gamma_0^{(N)}+\gamma_1^{(N)}\nu_j+\sum_{r=2}^{N-1}\gamma_r^{(N)}\bar{R}_{rj},\quad j=1,2,\ldots J. \tag{5.16e}$$

[the equations (5.16c) and (5.16d) with $j=J$ need not be enforced, since they are automatically implied by the others].

Note that, for $t=t_0$, all these ξ_j's coincide [see (5.16a)].

For instance, for $J=2$, these restrictions become simply

$$2\nu_1\nu_2-\bar{R}_{21}-\bar{R}_{22}=0, \tag{5.17a}$$

$$\nu_1\bar{R}_{r2}+\nu_2\bar{R}_{r1}-\bar{R}_{r+1,1}-\bar{R}_{r+1,2}=0,$$
$$r=2,3,\ldots,N-1, \tag{5.17b}$$

with

$$\bar{R}_{Nj}=\gamma_0^{(N)}+\gamma_1^{(N)}\nu_j+\sum_{r=2}^{N-1}\gamma_r^{(N)}\bar{R}_{rj},\quad j=1,2 \tag{5.17c}$$

and in particular, for $N=2$ [see (5.10)],

$$(\nu_1-v_1)(\nu_2-v_2)+(\nu_1-v_2)(\nu_2-v_1)=0, \tag{5.18}$$

and for $N=3$ [see (5.11)],

$$2\nu_1\nu_2-\bar{\eta}_1-\bar{\eta}_2=0, \tag{5.19a}$$

$$\nu_1\bar{\eta}_2+\nu_2\bar{\eta}_1-2\gamma_0-\gamma_1(\nu_1+\nu_2)-\gamma_2(\bar{\eta}_1+\bar{\eta}_2)=0. \tag{5.19b}$$

Clearly the relation (5.18) admits a one-parameter family of solutions for the two unknowns ν_1 and ν_2; while the two relations (5.19) admit a two-parameter family of solutions for the four unknowns ν_1, ν_2, $\bar{\eta}_1$, and $\bar{\eta}_2$.

The solution of the Cauchy problem for the system (5.9) is given by the following result (derived in Appendix F): *the quantities $\xi_j(t)$, $j=1,2,\ldots,J$, are the J roots of the following algebraic equation of degree J in x*:

$$\sum_{n=1}^{N}\mu_n\prod_{k=1}^{J}[x-\xi_k(0)-v_nt]$$
$$=\sum_{n=1}^{N}\sum_{j=1}^{J}Q_{nj}\int_{x_0}^{x-v_nt}dx'\prod_{k=1,k\neq j}^{J}[x'-\xi_k(0)], \tag{5.20a}$$

where x_0 is an arbitrary constant (*independent of n*),

$$\mu_n=-\prod_{m=1}^{N}{}'\left[\frac{v_m}{v_m-v_n}\right], \tag{5.20b}$$

and the quantities Q_{jn} are given, in terms of the initial data, by the formula

$$Q_{nj}=\lambda_{n1}^{(N)}\dot{\xi}_j(0)+\sum_{r=2}^{N-1}\lambda_{nr}^{(N)}R_{rj}(0) \tag{5.20c}$$

with the constants $\lambda_{nr}^{(N)}$ given, in terms of the v_n's, in Appendix E. In particular, for $N=2$,

$$\lambda_{11}^{(2)}=-\lambda_{21}^{(2)}=1/(v_1-v_2) \tag{5.21a}$$

and

$$Q_{nj}=\lambda_{n1}^{(2)}\dot{\xi}_j(0),\quad n=1,2,\quad j=1,2,\ldots,J, \tag{5.21b}$$

and for $N=3$,

$$\lambda_{n1}^{(3)}=-\left(\sum_{m=1,m\neq n}^{3}v_m\right)\prod_{m=1}^{3}{}'(v_n-v_m)^{-1},\quad n=1,2,3, \tag{5.22a}$$

$$\lambda_{n2}^{(3)} = \prod_{m=1}^{3}{}' (v_n - v_m)^{-1}, \quad n = 1,2,3, \tag{5.22b}$$

and [see (5.11c)]

$$Q_{nj} = \lambda_{n1}^{(3)} \dot{\xi}_j(0) + \lambda_{n2}^{(3)} \eta_j(0), \quad n = 1,2,3, \quad j = 1,2,...,J. \tag{5.22c}$$

For $J=2$ and $J=3$ the trajectories $\xi_j(t)$ can be obtained in explicit form, by solving (5.20a); this is left as an exercise for the diligent reader. Here we limit ourselves to indicate how to obtain, for any J, the asymptotic ($t \to \pm\infty$) values V_j of the velocities $\dot{\xi}_j(t)$. Clearly these values are the roots of the algebraic equation of degree J in V,

$$J \sum_{n=1}^{N} \mu_n (V - v_n)^J = \sum_{n=1}^{N} \sum_{j=1}^{J} Q_{nj} (V - v_n)^J, \tag{5.23a}$$

that is obtained from (5.20a) by replacing x with Vt and letting t diverge. Using (F10c), (F10b), and (F9b), this equation can be conveniently rewritten as follows:

$$\sum_{j=1}^{J} \sum_{n=1}^{N} \rho_{nj}(0)(V - v_n)^J = 0, \tag{5.23b}$$

$$\sum_{j=1}^{J} \sum_{k=0}^{J} \binom{J}{k} (-)^k V^{J-k} \sum_{n=1}^{N} (v_n)^k \rho_{nj}(0) = 0. \tag{5.23c}$$

If $J<N$, this equation reads simply [see (5.7a)–(5.7c)]

$$\sum_{k=0}^{J} \binom{J}{k} (-)^k V^{J-k} \sum_{j=1}^{J} R_{kj}(0) = 0. \tag{5.24}$$

For $J \geqslant N$, this equation remains valid provided the $R_{kj}(t)$'s with $k>N$ are defined by the natural extension of (5.7a); but then the quantities $R_{kj}(0)$ with $k \geqslant N$ that enter in (5.24) must be expressed in terms of the quantities $R_{nj}(0)$ with $n<N$. The relevant formula to do this is of course [see (E7), and (5.7a)–(5.7c)]

$$R_{kj}(0) = \sum_{s=0}^{N-1} \sum_{n=1}^{N} (v_n)^k \lambda_{ns}^{(N)} R_{sj}(0),$$
$$j = 1,2,...,J, \quad k = 0,1,2,... \tag{5.25}$$

[for $k<N$ this formula becomes an identity; see (E11)].

VI. OUTLOOK

In this final section we outline some generalizations and extensions of the results reported above.

A trivial extension of the system (2.6) is obtained via the change of dependent variables

$$u_n(x,t) = f_n[\hat{u}_n(x,t)], \tag{6.1a}$$

that yields

$$\left(\frac{\partial}{\partial t} + v_n \frac{\partial}{\partial x}\right) \hat{u}_n(x,t) = g_n[\hat{u}_n(x,t)] \sum_{m=1}^{N} (v_m - v_n) f_m[\hat{u}_m(x,t)], \tag{6.1b}$$

with

$$g_n(y) = f_n(y)/f_n'(y) = 1/[d \ln f_n(y)/dy]. \tag{6.1c}$$

A case of special interest is obtained for $f_n(y) = \exp(\alpha_n y)$, $g_n(y) = 1/\alpha_n$.

A less trivial extension of (2.6) [or, equivalently, of (2.2b)] is obtained by replacing the discrete variable n with the continuous variable k,

$$\left(\frac{\partial}{\partial t} + v(k)\frac{\partial}{\partial x}\right) u(k,x,t) = u(k,x,t) \int dk' [v(k') - v(k)] f(k') u(k',x,t). \tag{6.2}$$

Here $v(k)$ and $f(k)$ are given functions; $f(k)$ might have compact support, or in any case it should guarantee the convergence of the integral in the rhs. The solution of this nonlinear integrodifferential equation is given by the formula

$$u(k,x,t) = w[k,x - v(k)t]/F(x,t), \tag{6.3}$$

that yields for $F(x,t)$ the two compatible equations

$$F_t(x,t) = -\int dk\, v(k) f(k) w[k,x - v(k)t], \tag{6.4a}$$

$$F_x(x,t) = \int dk\, f(k) w[k,x - v(k)t]. \tag{6.4b}$$

Hence

$$F(x,t) = 1 + \int dk\, f(k) \int_{x_0}^{x - v(k)t} dx'\, w(k,x'). \tag{6.5}$$

From (6.3) with $t=0$ and (6.5) there also obtains

$$F(x,0) = \exp\left[\int_{x_0}^{x} dx' \int dk\, f(k) u(k,x',0)\right], \tag{6.6}$$

implying, via (6.3) with $t=0$,

$$w(k,x) = u(k,x,0) \exp\left[\int_{x_0}^{x} dx' \int dk' f(k') u(k',x',0)\right]. \tag{6.7}$$

The two formulas (6.7) and (6.3) with (6.5) provide the solution of the Cauchy problem for (6.2), with $u(k,x,0)$ the imput datum.

A variation of the system (2.6) with (2.12) that can be easily solved is the Cauchy/boundary value problem, which seeks to determine the solution $u_n(x,t)$ for $x \geqslant 0$ and $t \geqslant 0$, in terms of $u_n(x,0)$ given for $x \geqslant 0$ and $u(0,t)$ given for $t \geqslant 0$.

Finally we mention a multidimensional extension of (2.6) [or, equivalently, (2.2b)], that can be partially solved. It reads

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_n \cdot \nabla\right) u_n(\mathbf{x},t) = u_n(\mathbf{x},t) \sum_{m=1}^{N} [(\mathbf{v}_n - \mathbf{v}_m) \cdot \boldsymbol{\alpha}_m] u_n(\mathbf{x},t), \tag{6.8}$$

with the $\mathbf{v}_n$'s and $\boldsymbol{\alpha}_m$'s given constant vectors (in d-dimensional space). It is then easily seen that a class of solutions (that is, however, far from including all possible solutions) is provided by the *ansatz*

$$u_n(\mathbf{x},t) = w_n[\boldsymbol{\alpha}_n \cdot (\mathbf{x} - \mathbf{v}_n t)]/F(\mathbf{x},t), \tag{6.9}$$

where the functions $w_n(y)$ are arbitrary (but depend only on a *scalar* variable), and the denominator $F(\mathbf{x},t)$ satisfies the *compatible* equations

$$\nabla F(\mathbf{x},t) = \sum_{m=1}^{N} \boldsymbol{\alpha}_m w_m[\boldsymbol{\alpha}_m \cdot (\mathbf{x} - \mathbf{v}_m t)], \tag{6.10a}$$

$$F_t(\mathbf{x},t) = -\sum_{m=1}^{N} (\boldsymbol{\alpha}_m \cdot \mathbf{v}_m) w_m[\boldsymbol{\alpha}_m \cdot (\mathbf{x} - \mathbf{v}_m t)], \tag{6.10b}$$

and can be therefore easily computed in terms of the functions w_m [for instance by integrating (6.10a) along some convenient path].

APPENDIX A: DERIVATION OF (3.4)

The derivation of (3.4) from (3.2) is easy; we report it here for completeness.

The starting point is (3.2), which can be rewritten as follows:

$$u_n(x,t) = w_n(x - v_n t) \times\left(1 + \sum_{m=1}^{N} \int_{-\infty}^{x} dx'\, w_m(x' - v_m t)\right)^{-1}. \quad \text{(A1)}$$

Hence

$$\sum_{m=1}^{N} u_n(x,t) = \left(\frac{d}{dx}\right)\ln\left[1 + \sum_{m=1}^{N} \int_{-\infty}^{x} dx'\, w_m(x' - v_m t)\right]. \quad \text{(A2)}$$

This yields

$$\int_{-\infty}^{x} dx' \sum_{n=1}^{N} u_n(x',t) = \ln\left[1 + \sum_{m=1}^{N} \int_{-\infty}^{x} dx'\, w_m(x' - v_m t)\right], \quad \text{(A3a)}$$

namely

$$1 + \sum_{m=1}^{N} \int_{-\infty}^{x - v_m t} dx'\, w_m(x') = \exp\left[\int_{-\infty}^{x} dx' \sum_{m=1}^{N} u_m(x',t)\right]. \quad \text{(A3b)}$$

The insertion of this formula in (3.2) yields (3.4). Q.E.D.

APPENDIX B: DERIVATION OF (3.11)

In this Appendix we derive (3.11).

It is easily seen that (3.3) and (3.1) imply that the functions $w_n(x)$ are integrable,

$$\int_{-\infty}^{+\infty} dx\, w_n(x) = c_n, \quad \text{(B1a)}$$

$$|c_n| < \infty. \quad \text{(B1b)}$$

Note, incidentally, the consistency of the notation used here with that employed above: see (3.5).

These formulas, together with (3.2), (3.9), and (1.2), imply the formulas

$$u_n^{(+)}(x) = w_n(x)\left(1 + \sum_{m=1}^{n-1} c_m + \int_{-\infty}^{x} dx'\, w_n(x')\right)^{-1}, \quad \text{(B2)}$$

$$u_n^{(-)}(x) = w_n(x)\left(1 + \sum_{m=n+1}^{N} c_m + \int_{-\infty}^{x} dx'\, w_n(x')\right)^{-1}. \quad \text{(B3)}$$

From (B3) we obtain

$$\int_{-\infty}^{x} dx'\, u_n^{(-)}(x') = \ln\left\{\left[1 + \sum_{m=n+1}^{N} c_m + \int_{-\infty}^{x} dx'\, w_n(x')\right] \times\left(1 + \sum_{m=n+1}^{N} c_m\right)^{-1}\right\}, \quad \text{(B4)}$$

yielding

$$\int_{-\infty}^{x} dx'\, w_n(x') = \left[1 + \sum_{m=n+1}^{N} c_m\right]\left\{\exp\left[\int_{-\infty}^{x} dx'\, u_n^{(-)}(x')\right] - 1\right\}, \quad \text{(B5a)}$$

$$w_n(x) = u_n^{(-)}(x)\left[1 + \sum_{m=n+1}^{N} c_m\right] \times\exp\left[\int_{-\infty}^{x} dx'\, u_n^{(-)}(x')\right]. \quad \text{(B5b)}$$

The insertion of these expressions in (B2) yields the formula

$$u_n^{(+)}(x) = u_n^{(-)}(x) \times\left(1 + a_n \exp\left[-\int_{-\infty}^{x} dx'\, u_n^{(-)}(x')\right]\right)^{-1}, \quad \text{(B6)}$$

with

$$a_n = \left[\sum_{m=1}^{n-1} c_m - \sum_{m=n+1}^{N} c_m\right]\left(1 + \sum_{m=n+1}^{N} c_m\right)^{-1}. \quad \text{(B7)}$$

The formula (B6) coincides with the formula, (3.11), we set out to prove; there remains to show that the expression (B7) of the constants a_n coincides with (3.12).

To this end we set $x = \infty$ in (B5a) and, using the definitions (B1a) and (3.10a), we get

$$c_n = \left[1 + \sum_{m=n+1}^{N} c_m\right]\left[\exp(b_n^{(-)}) - 1\right]. \quad \text{(B8)}$$

This formula yields, for $n = N$,

$$c_N = \exp(b_N^{(-)}) - 1; \quad \text{(B9a)}$$

and all the preceding c_n's can be computed recursively. The result can be easily expressed in compact form,

$$c_n = \left[\exp(b_n^{(-)}) - 1\right]\exp\left[\sum_{m=n+1}^{N} b_m^{(-)}\right], \quad \text{(B9b)}$$

$$c_n = \exp\left[\sum_{m=n}^{N} b_m^{(-)}\right] - \exp\left[\sum_{m=n+1}^{N} b_m^{(-)}\right]. \quad \text{(B9c)}$$

This last formula yields the relations

$$\sum_{m=1}^{n-1} c_m = \exp\left[\sum_{m=1}^{N} b_m^{(-)}\right] - \exp\left[\sum_{m=n}^{N} b_m^{(-)}\right], \quad \text{(B10a)}$$

$$\sum_{m=n+1}^{N} c_m = \exp\left[\sum_{m=n+1}^{N} b_m^{(-)}\right] - 1. \quad \text{(B10b)}$$

The insertion of these two expressions in (B7) yields (3.12). Q.E.D.

Note the consistency of the formulas (B9) with (3.7).

For completeness let us also report the expressions of c_n

in terms of the $b_m^{(+)}$'s [whose derivation is analogous to that of (B9)],

$$c_n = [\exp(b_n^{(+)}) - 1]\exp\left[\sum_{m=1}^{n-1} b_m^{(+)}\right], \tag{B11a}$$

$$c_n = \exp\left[\sum_{m=1}^{n} b_m^{(+)}\right] - \exp\left[\sum_{m=1}^{n-1} b_m^{(+)}\right], \tag{B11b}$$

implying of course

$$\sum_{m=\mu}^{\nu} c_m = \exp\left[\sum_{m=1}^{\nu} b_m^{(+)}\right] - \exp\left[\sum_{m=1}^{\mu-1} b_m^{(+)}\right], \quad \nu \geqslant \mu - 1. \tag{B12}$$

Note finally that the formulas (B9a) and (B11a) can also be obtained, more directly if less rigorously, by integrating (3.4) over x from $-\infty$ to $+\infty$ [see (B1a)] in the limit $t \to \mp\infty$ [see (1.2) and (3.1a), (3.10b); and recall that, in these limits, the function $u_n(x,t)$ is localized around $x = v_n t$, so that the different functions $u_m(x,t)$ are widely separated, i.e., they do not overlap].

APPENDIX C: SOLUTION OF A FUNCTIONAL EQUATION

In this Appendix we solve the functional equation

$$f[x - y(a)] = f(x)\left(1 + a\exp\left[-\int_{-\infty}^{x} dx' f(x')\right]\right)^{-1}. \tag{C1}$$

We assume of course that the function $f(x)$ vanishes as $x \to -\infty$, so that the integral in the rhs of (C1) converges.

It is convenient to set

$$F(x) = \int_{-\infty}^{x} dx' f(x'), \tag{C2}$$

and to integrate (C1) to yield

$$F[x - y(a)] = \ln[\{\exp[F(x)] + a\}/(1+a)]. \tag{C3}$$

We now set

$$a = \varepsilon, \tag{C4a}$$

$$y(a) = \varepsilon/p + O(\varepsilon^2), \quad y'(0) = 1/p, \tag{C4b}$$

and obtain from (C3), in the $\varepsilon \to 0$ limit,

$$p^{-1}F'(x) = 1 - \exp[-F(x)]. \tag{C5}$$

This ODE is easily integrated,

$$F(x) = \ln[1 + c^{-1}\exp(px)]. \tag{C6}$$

Here c is an arbitrary constant, and the condition

$$p > 0 \tag{C7}$$

is required in order that $F(-\infty)$ vanish [see (C2)].

It is now easy to check that the expression (C6) does indeed satisfy (C3), with

$$y(a) = p^{-1}\ln(1+a) \tag{C8}$$

[whose consistency with (C4) is obvious].

In conclusion, the general solution of the functional equation (C1) reads

$$f(x) = p/[1 + c\exp(-px)] \tag{C9}$$

with c and p arbitrary constants and $y(a)$ given by (C8). Here p is constrained by the condition (C7), and c is also required to be positive in order that $f(x)$ be regular for all real values of x.

Note that (C9) with (C7) implies

$$f(+\infty) = p > 0. \tag{C10}$$

APPENDIX D: TRAVELING WAVE SOLUTION

In this Appendix we indicate how to obtain the traveling wave solution,

$$u_n(x,t) = \tilde{u}_n(y), \tag{D1a}$$

$$y = x - Vt, \tag{D1b}$$

of (2.6). Here V is any given constant ($V \neq v_n$; see below).

The system (2.6) yields

$$(v_n - V)\tilde{u}_n' = \tilde{u}_n \sum_{m=1}^{N} (v_n - v_m)\tilde{u}_m. \tag{D2}$$

Note that the *general* solution of this nonlinear system of N ODE's must contain N arbitrary constants.

We now set

$$\tilde{u}_n(y) = A_n \exp(\alpha_n y)/F(y), \tag{D3}$$

with the $2N$ constants A_n, α_n, and the function $F(y)$, as yet undetermined. Then (D2) yields

$$(v_n - V)\alpha_n F(y) - (v_n - V)F'(y) = \sum_{m=1}^{N} (v_m - v_n)A_m \exp(\alpha_m y). \tag{D4}$$

This suggests the assignment

$$\alpha_n = \alpha/(v_n - V), \tag{D5}$$

with α an undetermined constant. Then the N equations (D4) yield two equations:

$$F'(y) = \sum_{m=1}^{N} A_m \exp\left[\frac{\alpha y}{v_m - V}\right], \tag{D6a}$$

$$\alpha F(y) + VF'(y) = \sum_{m=1}^{N} v_m A_m \exp\left[\frac{\alpha y}{v_m - V}\right]. \tag{D6b}$$

It is easily seen that these two equations are compatible, being both solved by the same expression

$$F(y) = \alpha^{-1}\sum_{m=1}^{N} (v_m - V)A_m \exp\left[\frac{\alpha y}{v_m - V}\right]. \tag{D7}$$

Insertion of this formula in (D3) yields

$$\tilde{u}_n(y) = \alpha A_n \exp\left[\frac{\alpha y}{v_m - V}\right] \times\left(\sum_{m=1}^{N} (v_m - V) A_m \exp\left[\frac{\alpha y}{v_m - V}\right]\right)^{-1}. \tag{D8}$$

This solution contains the $N+1$ arbitrary constants A_m and α, of which only N are effectively relevant since the overall scale of the constants A_m does not affect the solution (D8) (without loss of generality one could set to unity one of the constants A_m, say $A_\nu = 1$). Hence (D8) provides the *general* solution of (D2).

Throughout this discussion we have implicitly assumed that V differs from all the velocities v_n,

$$V \neq v_n, \quad n = 1,2,\ldots,N. \tag{D9}$$

If instead V coincides with one of the v_n's, say,

$$V = v_\nu, \tag{D10a}$$

then clearly a (trivial) traveling wave solution of (2.6) reads

$$u_n(x,t) = \delta_{n\nu}\tilde{u}_\nu(x - v_\nu t), \tag{D10b}$$

with $\tilde{u}_\nu(y)$ any arbitrary function.

APPENDIX E: EVALUATION OF $\gamma_r^{(N)}$ AND $\gamma_{nr}^{(N)}$

The constants γ_r (we omit, for notational simplicity, the superscript N) are defined by the relations

$$y_r = \sum_{n=1}^{N} (v_n)^r x_n, \quad r = 0,1,\ldots,N, \tag{E1a}$$

$$y_N = \sum_{r=0}^{N-1} \gamma_r y_r. \tag{E1b}$$

Here the N v_n's are assumedly given (all different), and the N x_n's (or, equivalently, the N y_r's, $r = 0,1,\ldots,N-1$) are arbitrary.

From (E1a) with $r = N$, and (E1b), we obtain

$$\sum_{n=1}^{N} (v_n)^N x_n = \sum_{r=0}^{N-1} \gamma_r y_r. \tag{E2}$$

We now replace y_r in the rhs [using (E1a)] and exchange the order of the two summations, getting

$$\sum_{n=1}^{N} x_n\left[(v_n)^N - \sum_{r=0}^{N-1} \gamma_r (v_n)^r\right] = 0, \tag{E3a}$$

namely,

$$(v_n)^N - \sum_{r=0}^{N-1} \gamma_r (v_n)^r = 0, \quad n = 1,2,\ldots,N. \tag{E3b}$$

This implies that the polynomial in v of degree N

$$p_N(v) = v^N - \sum_{r=0}^{N-1} \gamma_r v^r \tag{E4}$$

has the N roots v_n. Hence

$$v^N - \sum_{r=0}^{N-1} \gamma_r v^r = \prod_{n=1}^{N} (v - v_n). \tag{E5}$$

It is thus seen that *the quantities γ_r coincide with the symmetric invariants of the set $\{v_n\}$*. In particular,

$$\gamma_0 = (-)^{N-1} \prod_{n=1}^{N} v_n, \tag{E6a}$$

$$\gamma_{N-1} = \sum_{n=1}^{N} v_n. \tag{E6b}$$

The constants λ_{nr} (we omit again, for notational simplicity, the superscript N) are defined by the equations

$$y_r = \sum_{n=1}^{N} (v_n)^r x_n, \quad r = 0,1,\ldots,N-1, \tag{E7a}$$

$$x_n = \sum_{r=0}^{N-1} \lambda_{nr} y_r, \quad n = 1,2,\ldots,N. \tag{E7b}$$

Here, as above, the N v_n's are given, and the N x_n's (or, equivalently, the N y_r's, $r = 0,1,\ldots,N-1$) are arbitrary (actually in our case $y_0 = 1$; but this has no relevance to the calculation of the λ_{nr}'s).

Insertion of (E7a) in (E7b) yields the formula

$$\sum_{r=0}^{N-1} \lambda_{nr}(v_m)^r = \delta_{nm}, \quad n,m = 1,2,\ldots,N. \tag{E8}$$

Hence the (Lagrangian interpolational) polynomial of degree $N-1$ in v,

$$q_{N-1}^{(n)}(v) = \sum_{r=0}^{N-1} \lambda_{nr} v^r, \tag{E9a}$$

has the following explicit representation:

$$q_{N-1}^{(n)}(v) = \prod_{m=1, m\neq n}^{N} \left[\frac{v - v_m}{v_n - v_m}\right]. \tag{E9b}$$

We have therefore shown that *λ_{nr} is the coefficient of v^r in the polynomial (E 9b)*.

In particular,

$$\lambda_{n0} = \prod_{m=1}^{N}{}' \left[\frac{v_m}{v_m - v_n}\right], \tag{E10a}$$

$$\lambda_{n,N-1} = \prod_{m=1}^{N}{}' (v_n - v_m)^{-1}, \tag{E10b}$$

$$\lambda_{n,N-2} = -\left(\sum_{m=1, m\neq n}^{3} v_m\right)\left(\prod_{m=1}^{N}{}' (v_n - v_m)\right)^{-1}. \tag{E10c}$$

Finally note that, summing (E7b) multiplied by $(v_n)^p$ over n and comparing with (E7a), we obtain the formula

$$\sum_{n=1}^{N} (v_n)^p \lambda_{nr} = \delta_{pr}, \quad p,r = 0,1,2,\ldots,N-1. \tag{E11}$$

APPENDIX F: SOLUTION OF CAUCHY PROBLEM FOR (5.9)

In this Appendix we indicate how to solve the Cauchy problem for the system (5.9). We limit our treatment to providing an explicit formula that reduces the computation of $\xi_j(t), j = 1,2,\ldots,J$, to the solution of an algebraic equation of degree J; the derivation of the corresponding expressions for $R_{rj}(t)$, $r = 2,3,\ldots,N-1$; $j = 1,2,\ldots,J$, is left as an easy exercise for the diligent reader.

The solution is of course achieved by exploiting the connection [via (5.3) with (5.4b), (5.6), and (5.7)] of (5.9) with (2.6), and using the solution of the Cauchy problem for (2.6) [see (2.10) and (2.11)].

At $t = 0$ (5.3) with (5.6) yields

$$u_n(x,0) = \sum_{j=1}^{N} \frac{\rho_{nj}(0)}{x - \xi_j(0)}. \tag{F1}$$

Via (2.11) and (5.4b) this yields

$$w_n(x) = \sum_{j=1}^{J} \left\{\frac{\rho_{nj}(0)}{x_0 - \xi_j(0)}\right\} \prod_{k=1, k\neq j}^{J} \left\{\frac{x - \xi_k(0)}{x_0 - \xi_k(0)}\right\}. \tag{F2}$$

The positions $\xi_j(t)$ of the poles of $u_n(x,t)$ coincide [see (5.3)] with the zeros of the denominator in the rhs of (2.10),

$$F(x,t) = 1 + \sum_{n=1}^{N}\sum_{j=1}^{N} \left\{\frac{\rho_{nj}(0)}{x_0 - \xi_j(0)}\right\} \int_{x_0}^{x - v_m t} dx' \times \prod_{k=1, k\neq j}^{J} \left\{\frac{x' - \xi_k(0)}{x_0 - \xi_k(0)}\right\}, \tag{F3}$$

$$F[\xi_j(t), t] = 0, \quad j = 1,2,\ldots,J. \tag{F4}$$

These formulas provide the solution; but some further elaboration is in order. First, we note that, for $t=0$ [using (5.4b)], (F3) becomes

$$F(x,0)=1+\left\{\prod_{k=1}^{J}[x_0-\xi_k(0)]^{-1}\right\}\times\sum_{j=1}^{J}\int_{x_0}^{x}dx'\prod_{k=1,k\neq j}^{J}[x'-\xi_k(0)]; \qquad \text{(F5)}$$

and (F4) implies that this quantity must vanish for $x=\xi_p(0)$,

$$1+\left\{\prod_{k=1}^{J}[x_0-\xi_k(0)]^{-1}\right\}\times\sum_{j=1}^{J}\int_{x_0}^{\xi_p(0)}dx'\prod_{k=1,k\neq j}^{J}[x'-\xi_k(0)]=0. \qquad \text{(F6)}$$

Note incidentally that this identity is easily verified, since

$$\sum_{j=1}^{J}\prod_{k=1,k\neq j}^{J}[x-\xi_k(0)]=\left(\frac{d}{dx}\right)\prod_{k=1}^{J}[x-\xi_k(0)]. \qquad \text{(F7)}$$

Now from (F6), (F3), and (F4), and using (5.4b), we get

$$\sum_{n=1}^{N}\sum_{j=1}^{N}\rho_{nj}(0)\int_{\xi_p(0)}^{\xi_s(t)-v_nt}dx\prod_{k=1,k\neq j}^{J}[x-\xi_k(0)]=0, \quad s=1,2,\ldots,J. \qquad \text{(F8)}$$

To render this formula more suitable to solve (5.9), it is moreover convenient to express the quantities $\rho_{nj}(0)$ in terms of the quantities $R_{rj}(0)$. The relevant formula,

$$\rho_{nj}(0)=\sum_{r=0}^{N-1}\lambda_{nr}^{(N)}R_{rj}(0), \qquad \text{(F9a)}$$

is obtained by inverting (5.7a); the coefficients $\lambda_{nr}^{(N)}$ are computed in Appendix E. Note that, via (5.7a)–(5.7c) this equation may be rewritten as follows:

$$\rho_{nj}(0)=\lambda_{n0}^{(N)}+\lambda_{n1}^{(N)}\dot{\xi}_j(0)+\sum_{r=2}^{N-1}\lambda_{nr}^{(N)}R_{rj}(0). \qquad \text{(F9b)}$$

The insertion of (F9b) in (F8) now yields [after using (F6)]

$$\sum_{n=1}^{N}\sum_{j=1}^{N}Q_{nj}(0)\int_{\xi_p(0)}^{\xi_s(t)-v_nt}dx\prod_{k=1,k\neq j}^{J}[x-\xi_k(0)]=\sum_{n=1}^{N}\mu_n\prod_{k=1}^{J}[\xi_s(t)-v_nt-\xi_k(0)], \quad s=1,2,\ldots,J, \qquad \text{(F10a)}$$

with

$$Q_{nj}=\lambda_{nl}^{(N)}\dot{\xi}_j(0)+\sum_{r=2}^{N-1}\lambda_{n2}^{(N)}R_{rj}(0), \qquad \text{(F10b)}$$

$$\mu_n=-\lambda_{n0}^{(N)}. \qquad \text{(F10c)}$$

Finally note that, since for $r>0$,

$$\sum_{n=1}^{N}\lambda_{nr}^{(N)}=0 \qquad \text{(F11a)}$$

[see (E11)], (F10b) implies the identity

$$\sum_{n=1}^{N}Q_{nj}=0. \qquad \text{(F11b)}$$

Hence the lower limit $\xi_p(0)$ in the lhs of (F10a) can be replaced by any arbitrary quantity, say x_0, that does not depend on the index n.

This completes the proof of (5.20).

[1]F. Calogero, "Universality and integrability of the nonlinear evolution PDE's describing N-wave interactions," J. Math. Phys. 30, 28 (1989).

[2]F. Calogero, "Why are certain nonlinear PDEs both widely applicable and integrable?," in *What is integrability (for nonlinear PDEs)?*, edited by V. E. Zakharov (Springer, Berlin, to be published).

[3]F. Calogero, "Motion of poles and zeros of special solutions of nonlinear and linear partial differential equations, and related 'solvable' many-body problems," Nuovo Cimento B 43, 177 (1978).

[4](a) F. Calogero, "The evolution PDE $u_t=u_{xxx}+3(u_{xx}u^2+3u_x{}^2u)+3u_xu^4$," J. Math. Phys. 28, 538 (1987); (b) F. Calogero and S. De Lillo "The Eckhaus PDE $i\Psi_t+\Psi_{xx}+2(|\Psi|^2)_x\Psi+|\Psi|^4\Psi=0$," Inverse Problems 3, 633 (1987).

Exact path integral solution of the Coulomb plus Aharonov–Bohm potential

L. Chetouani and L. Guechi
Département de Physique Théorique, Institut de Physique, Université de Constantine, Constantine, Algeria
T. F. Hammann
Laboratoire de Mathématiques, Physique Mathématique et Informatique, Faculté des Sciences et Techniques, Université de Haute Alsace, 4, rue des Frères Lumière, F 68093 Mulhouse, France

(Received 18 February 1988; accepted for publication 19 October 1988)

The Green's function for the sum of the Coulomb and Aharonov–Bohm potentials is calculated exactly in the path integral formalism. The energy spectrum is deduced.

I. INTRODUCTION

We calculate the Green's function relative to a particle of charge e', submitted to a potential that is the sum of the Coulomb and Aharonov–Bohm potentials.

The vector potential **A** of Aharonov–Bohm is defined in Cartesian coordinates as $\mathbf{A} = F(x\mathbf{j} - y\mathbf{i})/2\pi(r^2 - z^2)$, and in spherical coordinates (r,θ,ϕ) as

$$A_r = A_\theta = 0 \quad \text{and} \quad A_\phi = F/2\pi r \sin\theta,$$

where F is the flux created within a very thin and infinitely long solenoid, directed along the $0z$ axis. A punctual charge $(-e)$ is located at the origin of the axis system. The action S of this compound system that has a reduced mass μ, is written as follows:

$$S = \int_0^T \mathscr{L}_{cl}(\mathbf{r},\dot{\mathbf{r}})\, dt = \int_0^T \left(\frac{1}{2}\mu\dot{\mathbf{r}}^2 + \frac{e'}{c}\mathbf{A}\cdot\dot{\mathbf{r}} + \frac{e'e}{r}\right) dt. \tag{1}$$

The potential under consideration belongs to the class of noncentral potentials that has been studied very recently in the algebraical group approach,[1–3] in the functional integral formalism with parabolic coordinates,[4–6] as well as through the Schrödinger equation.[7] In this paper we show that this system with a velocity depending potential may also be solved in spherical coordinates through the path integral formalism.

II. PROPAGATOR

In the Feynman functional integral approach, the propagator may be written formally in Cartesian coordinates and in standard notation[8] as follows:

$$\begin{aligned} K(\mathbf{r}_f,\mathbf{r}_i;T) &= \int \mathscr{D}x(t)\mathscr{D}y(t)\mathscr{D}z(t) \\ &\quad\times \exp\left[\frac{i}{\hbar}\int_0^T \mathscr{L}_{cl}(\mathbf{r},\dot{\mathbf{r}})dt\right] \\ &= \lim_{N\to\infty} \int \left[\frac{\mu}{2i\pi\hbar\epsilon}\right]^{3N/2} \prod_{j=1}^{N-1} dx_j\, dy_j\, dz_j \\ &\quad\times \exp\left[\frac{i}{\hbar}\sum_{j=1}^{N} S(j,j-1)\right], \end{aligned} \tag{2}$$

where

$$\begin{aligned} S(j,j-1) &= \frac{\mu}{2\epsilon}[\Delta x_j^2 + \Delta y_j^2 + \Delta z_j^2] + \frac{e'}{c}[A_x(\tilde{\mathbf{r}}_j)\Delta x_j \\ &\quad + A_y(\tilde{\mathbf{r}}_j)\Delta y_j + A_z(\tilde{\mathbf{r}}_j)\Delta z_j] \\ &\quad + \frac{e'e\epsilon}{[\tilde{x}_j^2 + \tilde{y}_j^2 + \tilde{z}_j^2]^{1/2}}, \end{aligned}$$

represents the action corresponding to the elementary time interval $[t_{j-1},t_j]$. We used the following usual notations:

$$\epsilon = t_j - t_{j-1}, \quad T = N\epsilon = t_f - t_i,$$

$$\mathbf{r}_f = \mathbf{r}(t_N), \quad \mathbf{r}_i = \mathbf{r}(t_0), \quad u_j = u(t_j),$$

$$\Delta u_j = u_j - u_{j-1}, \quad \tilde{u}_j = (u_j + u_{j-1})/2.$$

Taking into account the equality,[9]

$$(e'/c)\mathbf{A}\cdot\dot{\mathbf{r}} = e'F\dot{\phi}/2\pi c,$$

the propagator $K(\mathbf{r}_f,\mathbf{r}_i;T)$ of (2) may be written in the usual spherical coordinates, $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$, as follows:

$$\begin{aligned} &K(\mathbf{r}_f,\mathbf{r}_i;T) \\ &\quad= \lim_{N\to\infty} \int \left[\frac{\mu}{2i\pi\hbar\epsilon}\right]^{3N/2} \prod_{j=1}^{N-1} r_j^2 \sin\theta_j\, dr_j\, d\theta_j\, d\phi_j \\ &\quad\times \exp\Bigg[\frac{i}{\hbar}\sum_{j=1}^{N}\Bigg\{\frac{\mu}{2\epsilon}[r_j^2 + r_{j-1}^2 - 2r_j r_{j-1} \\ &\quad\times(\cos\theta_j\cos\theta_{j-1} + \sin\theta_j\sin\theta_{j-1}\cos(\Delta\phi_j))] \\ &\quad+ \frac{ee'\epsilon}{\tilde{r}_j} + \frac{e'F\Delta\phi_j}{2\pi c}\Bigg\}\Bigg]. \end{aligned} \tag{3}$$

Let us use the Fourier expansion,[10]

$$\exp[z\cos(\Delta\phi)] = \sum_{m=-\infty}^{+\infty} I_m(z)e^{im(\Delta\phi)},$$

and, for small ϵ, the two following equations:

$$\cos(\Delta\phi) \approx \cos(\Delta\phi + a\epsilon) + a\epsilon\Delta\phi + \tfrac{1}{2}a^2\epsilon^2$$

and the asymptotic behavior[11] of the modified Bessel functions,

$$I_m\left(\frac{v}{\epsilon}\right) \approx \left(\frac{\epsilon}{2\pi v}\right)^{1/2} \exp\left[\frac{v}{\epsilon} - \frac{\epsilon}{2v}\left(m^2 - \frac{1}{4}\right)\right].$$

By integration on the angular variable ϕ, the propagator (3) may then be decomposed into partial kernels

$$K(\mathbf{r}_f,\mathbf{r}_i;T) = \sum_{m=-\infty}^{+\infty} \frac{e^{im(\phi_f-\phi_i)}}{2\pi} K_m(r_f,\theta_f,r_i,\theta_i;T), \tag{4}$$

with

$$K_m(r_f,\theta_f,r_i\theta_i;T) = \frac{1}{(r_i^2 r_f^2 \sin\theta_i \sin\theta_f)^{1/2}} \lim_{N\to\infty} \int \left[\frac{\mu}{2i\pi\hbar\epsilon}\right]^N \times \prod_{j=1}^{N} (r_j r_{j-1})^{1/2} \prod_{j=1}^{N-1} dr_j\, d\theta_j \times \exp\left[\frac{i}{\hbar}\sum_{j=1}^{N}\left\{\frac{\mu}{2\epsilon}\left[\Delta r_j^2 + 4r_j r_{j-1}\sin^2\frac{\Delta\theta_j}{2}\right] - \frac{\hbar^2(M^2-1/4)\epsilon}{2\mu\tilde{r}_j^2\sin^2\tilde{\theta}_j} + \frac{e'e\epsilon}{\tilde{r}_j}\right\}\right], \tag{5}$$

where

$$M^2 = (m - e'F/2\pi\hbar c)^2 .$$

Let us symmetrize this expression (5) with respect to the middle of the interval $[j, j-1]$; let us expand the measure up to order 2 in Δr,

$$\prod_{j=1}^{N} (r_j r_{j-1})^{1/2} \prod_{j=1}^{N-1} dr_j\, d\theta_j \approx \prod_{j=1}^{N} \tilde{r}_j\left(1 - \frac{\Delta r_j^2}{8\tilde{r}_j^2}\right) \prod_{j=1}^{N-1} dr_j\, d\theta_j,$$

and the action $A(j,j-1)$ up to order 4 in Δu_j,

$$A(j,j-1) \approx \frac{\mu}{2\epsilon}\{\Delta r_j^2 + \tilde{r}_j^2\Delta\theta_j^2\} - \frac{\hbar^2(M^2-1/4)\epsilon}{2\mu\tilde{r}_j^2\sin^2\tilde{\theta}_j} + \frac{e'e\epsilon}{\tilde{r}_j} - \frac{\mu}{8\epsilon}\left[\Delta r_j^2\Delta\theta_j^2 + \frac{1}{3}\tilde{r}_j^2\Delta\theta_j^4\right].$$

By means of the McLaughlin–Schulman procedure,[12] let us introduce a pure quantum effective centrifugal potential, by making the following substitutions:

$$\Delta r_j^2 \to \frac{i\hbar\epsilon}{\mu}, \quad \Delta r_j^2\Delta\theta_j^2 \to \frac{1}{\tilde{r}_j^2}\left(\frac{i\hbar\epsilon}{\mu}\right)^2, \quad \Delta\theta_j^4 \to \frac{3}{\tilde{r}_j^4}\left(\frac{i\hbar\epsilon}{\mu}\right)^2.$$

The partial kernel (5) then becomes

$$K_m(r_f,\theta_f;r_i,\theta_i;T) = \frac{1}{(r_i^2 r_f^2 \sin\theta_i \sin\theta_f)^{1/2}} \lim_{N\to\infty} \int \left[\frac{\mu}{2i\pi\hbar\epsilon}\right]^N \times \prod_{j=1}^{N} \tilde{r}_j \prod_{j=1}^{N-1} dr_j\, d\theta_j \times \exp\left\{\frac{i}{\hbar}\sum_{j=1}^{N}\left\{\frac{\mu}{2\epsilon}(\Delta r_j^2 + \tilde{r}_j^2\Delta\theta_j^2) + \left(\frac{\hbar^2}{8\mu\tilde{r}_j^2} - \frac{\hbar^2(M^2-1/4)}{2\mu\tilde{r}_j^2\sin^2\tilde{\theta}_j} + \frac{e'e}{\tilde{r}_j}\right)\epsilon\right\}\right\}. \tag{6}$$

At this point we notice that the separation of the radial and angular variables r and θ is not possible.

III. GREEN'S FUNCTIONS AND ENERGY SPECTRUM

In order to make the integration on the angular variable θ, we introduce the energy E by means of the Green's function [Fourier transform of the propagator (4)],

$$G(\mathbf{r}_f,\mathbf{r}_i;E) = \int_0^\infty dT \exp\left(\frac{iET}{\hbar}\right) K(\mathbf{r}_f,\mathbf{r}_i;T). \tag{7}$$

Let us utilize the Duru–Kleinert procedure[13] of reparametrization of the paths, by making use of the time transformation $t \to s$ defined by

$$\frac{dt}{ds} = r^2(s), \tag{8}$$

or, in discrete form,

$$\epsilon = \tau_j r_j r_{j-1} = \tau_j \tilde{r}_j^2(1 - \Delta r_j^2/4\tilde{r}_j^2), \quad \tau_j = s_j - s_{j-1}.$$

Taking into account the constraint $T = \int_0^S ds\, r^2(s)$, the Green's function is rewritten as

$$G(\mathbf{r}_f,\mathbf{r}_i;E) = \sum_{m=-\infty}^{+\infty} \frac{e^{im(\phi_f-\phi_i)}}{2\pi} \int_0^\infty dS\, P_E^m(\mathbf{r}_f,\theta_f,\mathbf{r}_i,\theta_i;S), \tag{9}$$

where

$$P_E^m(\mathbf{r}_f,\theta_f,\mathbf{r}_i,\theta_i;S) = \frac{1}{(\sin\theta_i \sin\theta_f)^{1/2}} \lim_{N\to\infty} \int \prod_{j=1}^{N} \left(\frac{\mu}{2i\pi\hbar\tau_j}\right)\frac{1}{\tilde{r}_j} \times \left(1 + \frac{\Delta r_j^2}{4\tilde{r}_j^2}\right) \prod_{j=1}^{N-1} dr_j\, d\theta_j \times \exp\left\{\frac{i}{\hbar}\sum_{j=1}^{N}\left\{\frac{\mu}{2\tau_j}\left(1 + \frac{\Delta r_j^2}{4\tilde{r}_j^2}\right)\left(\frac{\Delta r_j^2}{\tilde{r}_j^2} + \Delta\theta_j^2\right) + \left[\frac{\hbar^2}{8\mu} - \frac{\hbar^2(M^2-1/4)}{2\mu\sin^2\tilde{\theta}_j} + e'e\tilde{r}_j + E\tilde{r}_j^2\right]\tau_j\right\}\right\}, \tag{10}$$

represents the new kernel.

Let us replace the terms that appear in the measure and in the action, by reutilizing the McLaughlin–Schulman procedure, with the following substitutions:

$$\Delta r_j^2 \to \tilde{r}_j^2(i\hbar\tau_j/\mu), \quad \Delta r_j^2\Delta\theta_j^2 \to \tilde{r}_j^2(i\hbar\tau_j/\mu)^2,$$

$$\Delta r_j^4 \to 3\tilde{r}_j^4(i\hbar\tau_j/\mu)^2 .$$

This leads to a pure quantum correction $(-(\hbar^2/4\mu)\tau_j)$. Thus the separation of the variables r and θ is possible and the expression (10) is decomposed into a product of radial and angular kernels,

$$P_E^m(\mathbf{r}_f,\theta_f,\mathbf{r}_i,\theta_i;S) = [1/(\sin\theta_i \sin\theta_f)^{1/2}]\, P_r^m(r_f,r_i;S) P_\theta^m(\theta_f,\theta_i;S), \tag{11}$$

where

$$P_r^m(r_f,r_i;S) = \lim_{N\to\infty} \int \prod_{j=1}^{N} \left(\frac{\mu}{2i\pi\hbar\tau_j}\right)^{1/2} \frac{1}{\tilde{r}_j} \prod_{j=1}^{N-1} dr_j \times \exp\left[\frac{i}{\hbar}\sum_{j=1}^{N}\left\{\frac{\mu}{2\tau_j}\frac{\Delta r_j^2}{\tilde{r}_j^2} + \left(E\tilde{r}_j^2 + e'e\tilde{r}_j - \frac{\hbar^2}{8\mu}\right)\tau_j\right\}\right], \tag{12}$$

and

$$P_\theta^m(\theta_f,\theta_i;S) = \int \mathscr{D}\theta(s) \exp\left[\frac{i}{\hbar}\int_0^S \left\{\frac{\mu}{2}\dot{\theta}^2 - \frac{\hbar^2(M^2-1/4)}{2\mu\sin^2\theta}\right\} ds\right]. \tag{13}$$

This latter kernel can be calculated. It can be deduced from the calculus of the rigid rotator propagator,[14,15]

$$P_\theta^m(\theta_f,\theta_i;S) = (\sin\theta_i \sin\theta_f)^{1/2} \sum_{n=0}^{\infty}\left(n+|M|+\frac{1}{2}\right)\frac{(n+2|M|)!}{n!} \times \exp\left[-\frac{i}{\hbar}\frac{(n+|M|+1/2)^2\hbar^2 S}{2\mu}\right] \cdot P_{n+|M|}^{|M|}(\cos\theta_i)P_{n+|M|}^{|M|}(\cos\theta_f). \quad (14)$$

Let us insert (14), (12), and (11) into (9) and let us go back to the former temporal variable t, utilizing Eq. (8). In this time transformation[15] $s\to t$,

$$\frac{ds}{dt}=\frac{1}{r^2(t)},$$

we take into account the pure quantum correction that results from the McLaughlin–Schulman procedure. The Green's function (7) then becomes

$$G(\mathbf{r}_f,\mathbf{r}_i;E) = \frac{1}{r_i r_f}\sum_{m=-\infty}^{+\infty}\sum_{n=0}^{+\infty}\left(n+|M|+\frac{1}{2}\right) \times\frac{(n+2|M|)!}{n!}\frac{e^{im(\phi_f-\phi_i)}}{2\pi}P_{n+|M|}^{|M|}(\cos\theta_i) \times P_{n+|M|}^{|M|}(\cos\theta_f)\int_0^\infty dT\exp\left(\frac{iET}{\hbar}\right) \times\left\{\int \mathscr{D}r(t)\exp\left[\frac{i}{\hbar}\int_0^T \mathscr{L}(r,\dot r)dt\right]\right\}, \quad (15)$$

where

$$\mathscr{L}(r,\dot r)=\frac{\mu}{2}\dot r^2-\frac{\hbar^2}{8\mu r^2}[(2n+2|M|+1)^2-1]+\frac{e'e}{r},$$

is the radial Lagrangian of the hydrogen atom submitted to a centrifugal potential.

Let us make a last space-time transformation $(r,t)\to(u,s')$ defined by[16]

$$r=u^2,\quad \frac{dt}{ds'}=4u^2(s'), \quad (16)$$

or, in discrete form,

$$r_j=u_j^2,\quad r_{j-1}=u_{j-1}^2,$$
$$\epsilon=4(s'_j-s'_{j-1})u_ju_{j-1}=4\sigma_j u_j u_{j-1}.$$

Let us symmetrize the measure,

$$\prod_{j=1}^{N}\left(\frac{\mu}{2i\pi\hbar\epsilon}\right)^{1/2}\prod_{j=1}^{N-1}dr_j = \frac{1}{(4u_iu_f)^{1/2}}\prod_{j=1}^{N}\left(\frac{\mu}{2i\pi\hbar\sigma_j}\right)^{1/2}\prod_{j=1}^{N-1}du_j,$$

and the action S' relevant to the propagator shown in (15),

$$S'(j,j-1)=\frac{\mu}{2\epsilon}(r_j-r_{j-1})^2-\epsilon\left[\frac{\hbar^2}{8\mu r_j r_{j-1}}\{(2n+2|M|+1)^2-1\}-\frac{e'e}{\tilde r_j}\right] \approx\frac{\mu}{2\sigma_j}\Delta u_j^2-\sigma_j\left\{\frac{\hbar^2}{2\mu\tilde u_j^2}[(2n+2|M|+1)^2-1]-4e'e\right\}+\frac{\mu\Delta u_j^4}{8\sigma_j\tilde u_j^2},$$

and let us apply one last time the McLaughlin–Schulman procedure that consists of replacing $(\mu/8\sigma_j)(\Delta u_j^4/\tilde\mu_j^2)$ by $(-3\hbar^2\sigma_j/8\mu)$. The Green's function (15) can then be put into the following form:

$$G(\mathbf{r}_f,\mathbf{r}_i;E)=\frac{2}{(u_iu_f)^{3/2}}\sum_{m,n}\left(n+|M|+\frac{1}{2}\right)\frac{(n+2|M|)!}{n!}\frac{e^{im(\phi_f-\phi_i)}}{2\pi}P_{n+|M|}^{|M|}(\cos\theta_i)P_{n+|M|}^{|M|}(\cos\theta_f) \times\int_0^\infty dS' e^{i(4e'e/\hbar)S'}\int\mathscr{D}u(s')\exp\left\{\frac{i}{\hbar}\left[\int_0^{S'}ds'\left[\frac{\mu}{2}\dot u^2-\frac{\hbar^2}{2\mu u^2}\left((2n+2|M|+1)^2-\frac{1}{4}\right)+4Eu^2\right]\right]\right\}. \quad (17)$$

The Green's function of the potential under consideration can finally be calculated,[17] and gets,

$$G(\mathbf{r}_f,\mathbf{r}_i;E)=\frac{2}{u_iu_f}\frac{\mu\omega}{i\hbar}\sum_{m,n}\left(n+|M|+\frac{1}{2}\right)\frac{(n+2|M|)!}{n!}\frac{e^{im(\phi_f-\phi_i)}}{2\pi}P_{n+|M|}^{|M|}(\cos\theta_i) \times P_{n+|M|}^{|M|}(\cos\theta_f)\int_0^\infty dS'\left\{\frac{e^{i(4e'e/\hbar)S'}}{\sin(\omega S')}I_{2n+2|M|+1}\left(\frac{\mu\omega u_iu_f}{i\hbar\sin(\omega S')}\right)\exp\left[\frac{i\mu\omega}{2\hbar}(u_i^2+u_f^2)\cot(\omega S')\right]\right\},$$

where $\frac{1}{2}\mu\omega^2=-4E$.

This Green's function can also be expressed by means of the Whittaker functions[18]

$$G(\mathbf{r}_f,\mathbf{r}_i;E)=\sum_{n,m}\frac{(-1)^{n+|M|+1/2}}{\omega r_i r_f}\frac{(n+2|M|)!}{n!}\frac{\Gamma(n+|M|+p+1)}{\Gamma(2n+2|M|+1)}\frac{e^{im(\phi_f-\phi_i)}}{2\pi} \times P_{n+|M|}^{|M|}(\cos\theta_i)P_{n+|M|}^{|M|}(\cos\theta_f)\,M_{-p,n+|M|+1/2}\left(-\frac{\mu\omega}{\hbar}r_i\right)W_{p,n+|M|+1/2}\left(\frac{\mu\omega}{\hbar}r_f\right),\quad r_f>r_i,$$

where $p = -2e'e/\hbar$.

The energy spectrum can then be deduced from the poles of the Euler's function $\Gamma(n+|M|+p+1)$,

$$n+|M|+p+1 = -n', \quad \text{with} \quad n' = 0,1,2,...,\infty,$$

or

$$E = \frac{-\mu e'^2 e^2}{2\hbar^2(n+n'+|M|+1)^2}.$$

Let us notice that if the flux F is quantized, that is to say, if $F = (2\pi\hbar c/e')\times$ integer, then $|M|$ is integer and the spectrum thus obtained, is exactly that of the hydrogen atom. In this case, there is no Aharonov–Bohm effect.

IV. CONCLUSION

We have obtained in spherical coordinates, the Green's function associated with the sum of the Coulomb and Aharonov–Bohm potentials, utilizing simple space-time transformations. The solution is much more simple in parabolic coordinates,[6] but since the spherical coordinates are more commonly used, it is necessary to know how to treat this type of noncentral potentials, also in spherical coordinates.

Let us finally note that the Green's function of the Coulomb problem, with magnetic charges, obtained by Kleinert[19] via the Kustaanheimo–Stiefel transformation, the vector potential created by the magnetic monopole being

$$\mathbf{A} = g[z/r(r^2 - z^2)](x\mathbf{j} - y\mathbf{i}),$$

can be obtained in spherical coordinates, according to the same method.

ACKNOWLEDGMENT

The authors (L. C.) and (L. G.) would like to express their appreciation to Professor M. Kibler of the I. P. N. in Lyon, for having directed their attention to this compound noncentral potential.

[1] M. Kibler and P. Winternitz, J. Phys. A: Math. Gen. **20**, 4097 (1987).
[2] M. Kibler and T. Negadi, Phys. Lett. A **124**, 42 (1987).
[3] C. C. Gerry, Phys. Lett. A **118**, 445 (1986).
[4] M. V. Carpio and A. Inomata, in *Path Integrals from meV to MeV, Bielefeld Encounters in Physics and Mathematics VII*, edited by M. C. Gutzwiller, A. Inomata, J. R. Klauder, and L. Streit (World Scientific, Singapore, 1985), pp. 261–270.
[5] I. Sokmen, Phys. Lett. A **115**, 249 (1986).
[6] L. Chetouani, L. Guechi, and T. F. Hammann, Phys. Lett. A **125**, 277 (1987); L. Chetouani and T. F. Hammann, Phys. Rev. A **34**, 4737 (1986).
[7] A. Guha and S. Mukherjee, J. Math. Phys. **28**, 840 (1987).
[8] R. P. Feynman and H. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[9] S. F. Edwards, Proc. Phys. Soc. London **91**, 513 (1967).
[10] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 973, Eq. (8.511.4).
[11] W. Langguth and A. Inomata, J. Math. Phys. **20**, 499 (1979).
[12] D. McLaughlin and L. S. Schulman, J. Math. Phys. **12**, 2520 (1971).
[13] I. H. Duru and H. Kleinert, Phys. Lett. B **84**, 185 (1979).
[14] K. Pak and I. Sokmen, Phys. Lett. A **103**, 298 (1984); D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969).
[15] L. Chetouani, L. Guechi, and T. F. Hammann, Nuovo Cimento B **101**, 547 (1988).
[16] L. Chetouani and T. F. Hammann, J. Math. Phys. **27**, 2944 (1986).
[17] D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969).
[18] Reference 10, p. 729, Eq. (6-699-4).
[19] H. Kleinert, Phys. Lett. A **116**, 201 (1986).

Generalized coherent states and Berry's phase

G. Giavarini
Dipartimento di Fisica, Università di Parma, 43100 Parma, Italy and Istituto Nazionale di Fisica Nucleare, Gruppo Collegato di Parma, Parma, Italy

E. Onofri
Dipartimento di Fisica, Università di Trento, 38050 Povo, Italy and Istituto Nazionale di Fisica Nucleare, Gruppo Collegato di Parma, Parma, Italy

(Received 9 August 1988; accepted for publication 19 October 1988)

The quantum dynamics of time-dependent Hamiltonians of the kind $H(t) = U(t)H(0)U(t)^\dagger$ is studied in the framework of generalized coherent states. It is shown that the quantum and the classical dynamics are isomorphic and that the phase of $|\psi(t)\rangle$ is simply the classical action. In adiabatic approximation it is easy to extract Berry's phase from this. A discussion of generalized coherent states (CS's) necessary to deal with excited states is given; in particular for SU(2), SU(1,1), and E(2) their structure is completely characterized, showing that the deep connection with geometric quantization extends to these generalized CS's. This technique may be employed to estimate higher-order correction to the adiabatic approximation.

I. INTRODUCTION

There has recently been an increasing interest in Berry's phase, because of the remarkable variety of applications, ranging from quantum mechanics of molecules[1] to the more fundamental connection with Wess–Zumino terms and, in general, with anomalies.[2] In a number of examples it has been realized[3] that the Hamiltonian can be described in terms of some Lie algebra and Berry's phase can then be calculated either group theoretically or by coherent-state techniques. In the present paper we expand this idea by presenting a systematic treatment of Hamiltonians of the form

$$H(t) = U(t)H(0)U(t)^\dagger, \tag{1}$$

where $U(t)$ takes values in any unitary irreducible representation $\mathscr{D}$ of some Lie group G and $H(0)$ is some generator in its Lie algebra $\mathbf{g}$. [A slightly more general form, namely $H(t) = \lambda(t)U(t)H(0)U(t)^\dagger$, $\lambda(t)>0$, can always be reduced to Eq. (1) by redefining the time scale.] We will write $U(t) = \mathscr{D}(g(t))$, where $t \mapsto g(t)$ is some path in G. The relevant case for discussing Berry's phase consists of restricting $H(0)$ to belong to some Cartan subalgebra $\mathbf{h} \subset \mathbf{g}$ which implies that the spectrum of $H(t)$ is discrete. The parameter space that defines $H(t)$ is then identified in the coset space G/H, H being the Cartan subgroup with algebra $\mathbf{h}$.

Our main issue will be the following: the quantum dynamics generated by $H(t)$ is given exactly by some path in G, i.e.,

$$T\exp\left\{-i\int_0^t H(\tau)d\tau\right\} = \mathscr{D}(w(t)), \quad w(t)\in G, \tag{2}$$

where $w(t)$ satisfies a system of ordinary differential equations (essentially the Hamilton's equations of a corresponding classical system). This fact has the consequence that starting from any eigenstate $|n\rangle$ of $H(0)$ at $t=0$, the state at time t is a generalized coherent state[4,5] $\mathscr{D}(w(t))\cdot|n\rangle$ whose phase relative to a specific choice of coherent state vectors [i.e., Eq. (12)] is just the classical action along the path $\{w(t)\,|\,t\in\mathfrak{R}\}$, that is,

$$\Phi(t) = \int_{w(0)}^{w(t)} \boldsymbol{\theta} - \int_0^t h(w(\tau))d\tau. \tag{3}$$

Here $\boldsymbol{\theta}$ is the connection form whose curvature $\boldsymbol{\omega}$ is a two-form that can be considered as a classical symplectic structure on the parameter space G/H; $h(w)$ denotes the classical Hamiltonian with respect to ω.

Notice moreover that $\boldsymbol{\theta}$ is in general dependent on the initial state $|n\rangle$. The connection θ clearly gives an invariant phase $\oint\boldsymbol{\theta}$ when it happens that $w(T) = w(0)$ for some T. All this holds *exactly*, but we can easily get the adiabatic approximation too; this consists simply of substituting $w(t)$ for $g(t)$ in Φ and replacing the exact evolution operator $\mathscr{D}(w(t))$ by $\mathscr{D}(g(t))e^{i\Phi}$. In this case we will see that the same dependence on the initial state of $\boldsymbol{\theta}$ will be obviously shared by Berry's phase. Corrections to the adiabatic approximation can be easily evaluated [see the Appendix for the detailed calculation of the *displaced harmonic oscillator*, where $c(t)$ takes the role of $g(t)$ and $z(t)$ that of $w(t)$].

The clue to these results will be the idea that a unitary irreducible representation of a (compact semisimple) Lie group can always be described in a basis of coherent states (CS's) that establish a map CS: $\Omega \to \mathscr{K}$ from a classical phase space Ω into the space $\mathscr{K}$ carrier of the representation $\mathscr{D}$. The restriction to compact groups is dictated by simplicity; our results extend to the noncompact case with no substantial modification, at least for the case of discrete series representations (which correspond to Hamiltonians bounded from below). Our results can then be applied to simple models described by Hamiltonians of the form:

$$H(t) = \tfrac{1}{2}(\alpha p^2 + \beta(pq+qp) + \gamma q^2),$$
$$H(t) = \tfrac{1}{2}(\alpha p^2 + \beta(pq+qp) + \gamma q^2 + \chi/q^2),$$

both of which imply the group $G = \mathrm{SU}(1,1)$. Except for the particularly simple case of Hermitian symmetric spaces, however, it is in general necessary to introduce spaces of holomorphic differential forms, the so-called $\bar\partial$*-cohomology* spaces.

Our results have their roots in *geometric quantization*; it is well-known that geometric quantization leads naturally to coherent states.[6,7] What is less known is that the higher representations that are cut out by imposing a polarization are related to the generalized coherent states built over nonex-

 0022-2488/89/030659-05$02.50

tremal weights. This result will be proven for SU(2) and SU(1,1) and it would be nice to find its extension to groups of rank greater than one.

The paper is organized as follows: in Sec. II we discuss the dynamics generated by $H(t)$ and show how the phase of the state is given by a classical action. In Sec. III the problem of spin coherent states built over *any* eigenstate $|jm\rangle$ of J_3 is solved and the corresponding Berry's phase is shown to be given by the classical symplectic structure ω on the sphere S^2 with $\int\omega = 4\pi m$; at the end of this section the straightforward extension of the formalism to the case of discrete series representations of SU(1,1) is sketched. Section IV contains our conclusions and in the Appendix we give the detailed calculations for the Hamiltonian $(a^\dagger - c(t))(a - \bar{c}(t))$.

II. DYNAMICS OVER LIE GROUPS

Let G be a compact semisimple Lie group, $\mathbf{g}$ its Lie algebra, H a maximal torus in G (a Cartan subgroup), $\mathbf{h}$ its Lie algebra; also let $\mathscr{D}(G)$ be a unitary irreducible representation of G in a Hilbert space $\mathscr{H}$ with maximal weight λ (corresponding to a vector $|0\rangle \in \mathscr{H}$). Let the Hamiltonian be given by

$$H(t) = \mathscr{D}(g(t))H(0)\mathscr{D}(g(t))^\dagger \tag{4}$$

for some path $t \mapsto g(t) \in G$ with $H(0) \in \mathbf{h}$. Hereafter we simply drop $\mathscr{D}$ and use the group element symbol to denote its own representation.

The quantum dynamics generated by $H(t)$ is given by the time-ordered exponential

$$w(t) = T\exp\left\{-i\int_0^t H(\tau)d\tau\right\}. \tag{5}$$

By expanding $w(t)$ by Trotter's formula or by applying Magnus' formula,[8] namely

$$w(t) = \exp\left\{-i\int_0^T H(t)dt - \frac{1}{2}\int_0^T\int_0^t [H(t),H(t')]dt\,dt' + \cdots\right\}, \tag{6}$$

it follows that $w(t) \in G$, which shows that the quantum dynamics is given by a path on the group that can always be calculated by solving a system of ordinary differential equations. To prove this last assertion, let us introduce the basis of coherent states $\{|z\rangle\}$ labelled by complex coordinates $(z_1,z_2,...,z_n)$ which parametrize the coset space G/H. Here $\{|z\rangle\}$ represents a set of minimal uncertainty states labelled by the points of G/H which is a classical phase space when endowed with its Kirillov–Kostant two-form.[7,9] All vectors in $\mathscr{H}$ are represented by polynomials in $(z_1,z_2,...,z_n)$ and all generators of G act as *first-order holomorphic* differential operators. This means that the Schrödinger equation is of the form

$$i\frac{\partial\psi}{\partial t} = \sum_k \beta_k(z,t)\frac{\partial\psi}{\partial z_k} + \beta_0(z,t)\psi, \tag{7}$$

where the coefficients $\beta_k(z,t)$ are uniquely determined by $H(t)$. Any linear first-order differential equation like Eq. (7) can be solved by the method of characteristics, namely[10] the general solution is given by an arbitrary relation

$$\Psi(I_1,I_2,...,I_n) = 0 \tag{8}$$

among the first integrals of the associated (ordinary) system

$$\frac{d\psi}{\beta_0\psi} = \frac{dz_1}{\beta_1} = \cdots = \frac{dz_n}{\beta_n} = i\,dt. \tag{9}$$

The first integrals can be chosen as $I_k(z,t) = z_k(0)$ and the solution is formally given by

$$\psi(z,T) = \exp\left\{i\int_0^T \beta_0(t,z(z(0),t))dt\right\}\psi(z(0),0), \tag{10}$$

where $z(0)$ must be reexpressed as a function of (z,t) after having computed the integral. The flow $z(t)$ represents the projection on G/H of the path $w(t)^\dagger \in G$ given by Eq. (6). Now let $|\psi,t=0\rangle = |0\rangle$, i.e., the highest weight in the representation. Then

$$|\psi,t\rangle = w(t)|0\rangle = \exp(i\alpha(t))|z(t)\rangle, \tag{11}$$

having introduced the coherent state basis

$$|z\rangle = |g\cdot 0\rangle \equiv \mathscr{N}\frac{U(g)|0\rangle}{\langle 0|U(g)|0\rangle}, \tag{12}$$

with $\mathscr{N} = |\langle 0|U(g)|0\rangle|$. Here $g \in G$ and z is the complex coordinate of the coset space $\{gH\}$. The normalization adopted here is a very general recipe to fix the phase convention of the CS basis for any semisimple Lie group.[6] By the Schrödinger equation we get

$$\dot{\alpha}(t)|z(t)\rangle = i\frac{d}{dt}\Big|z(t)\Big\rangle - H(t)|z(t)\rangle \tag{13}$$

or

$$\dot{\alpha}(t) = -\langle z(t)|H(t)|z(t)\rangle + \left\langle z(t)\left|\frac{d}{dt}\right|z(t)\right\rangle. \tag{14}$$

From the general theory it is known that the diagonal matrix elements of $H(t)$ in the CS basis coincide with the *classical Hamiltonian* evaluated at $z(t)$; to evaluate $\langle z|d|z\rangle$, we recall that[6] $\langle z|\zeta\rangle = \exp(f(z,\bar{\zeta}) - \frac{1}{2}f(z,\bar{z}) - \frac{1}{2}f(\zeta,\bar{\zeta}))$, f being the *Kähler potential* that defines the symplectic structure on G/H. The phase of $|\psi(t)\rangle$ is thus given by

$$\alpha(t) = -\int_0^t h(z(\tau),\bar{z}(\tau))d\tau + \mathrm{Im}\left\{\int_0^{z(t)} \frac{\partial f}{\partial z}dz\right\}, \tag{15}$$

which is just the classical action along the path $z(t)$. Clearly we identify the geometrical term of the phase $\mathrm{Im}\{\int(\partial f/\partial z)dz\}$ with Berry's phase. The same calculation also works if we start from an arbitrary coherent state $|z_0\rangle$ at $t=0$. In the adiabatic approximation one can avoid solving the differential system [Eq. (9)] and just set $w(t) = g(t)$.

What happens if we start from an excited state, i.e., from a nonextremal weight? The dynamics is of course the same, but the identification with the classical action is slightly different. This has been worked out in detail for the case of rank one, namely SU(2) and SU(1,1) will be discussed in the following section; the nonsemisimple case E(2) is presented in the Appendix.

III. SU(2) and SU(1,1) COHERENT STATES FOR A GENERAL WEIGHT

We consider the spin j representation of SU(2), namely

$$(U(g)\psi)(z) = (\bar{\beta}z + \alpha)^{2j}\psi((\bar{\alpha}z - \beta)/(\bar{\beta}z + \alpha)), \tag{16}$$

where

$$g=\begin{pmatrix}\alpha & \beta\\ -\bar\beta & \bar\alpha\end{pmatrix},\quad \psi(z)=\sum_0^{2j}\psi_k z^k, \tag{17}$$

the vector $|jm\rangle$ being represented by z^{j-m}. Now consider the family of CS built over $|jm\rangle$:

$$|z_{(m)}\rangle \equiv \mathcal{N}\frac{U(g)|jm\rangle}{\langle jm|U(g)|jm\rangle}, \tag{18}$$

the normalization factor being in this case $\mathcal{N}=\Lambda(d(z,0))(1+z\bar z)^{-j}$, where $d(z,z')=|(z-z')/(1+\bar z'z)|$ is the invariant distance on the two-sphere and

$$\Lambda(x)=\sum_{k=0}^{j-|m|}\binom{j+m}{k}\binom{j-m}{k}(-x^2)^k. \tag{19}$$

We have the following theorem.

Theorem 1: The reproducing kernel is given by

$$\langle\zeta_{(m)}|\zeta'_{(m)}\rangle=\Lambda(d(\zeta,\zeta'))\cdot\frac{(1+\zeta\bar\zeta')^{j+m}(1+\bar\zeta\zeta')^{j-m}}{(1+\zeta\bar\zeta)^j(1+\zeta'\bar\zeta')^j}. \tag{20}$$

Proof: Let $\zeta=g\cdot 0$ and $\zeta'=g'\cdot 0$; we calculate $\langle jm|U(g)^\dagger U(g')|jm\rangle$ as the coefficient of z^{j-m} in the polynomial $\langle z|U(g)^\dagger U(g')|jm\rangle$, according to Eq. (16). After dividing by the normalization factors coming from definition (18) we get the result by simple algebra. ◁

The time evolution under $w(t)$ is the same as in Sec. II, but now the phase $\alpha(t)$ must be evaluated using the new reproducing kernel. The factors Λ do not contribute to α and we get

$$\alpha(t)=m\cdot\mathrm{Im}\left\{\int_{z(0)}^{z(t)}\frac{z\,d\bar z}{(1+\bar z z)}\right\}-\int_0^t\langle H\rangle d\tau. \tag{21}$$

We thus obtain the connection form of classical spin m. The $\langle H\rangle$ term is given by diagonal matrix elements of $H(t)$ in the new CS basis. It turns out that it is still given by the classical Hamiltonian with classical spin m. To understand this *migration* of classical spin from j to m we prove the following theorem.

Theorem 2: The representatives $\langle z_{(m)}|\psi\rangle$ in the CS basis $|z_{(m)}\rangle$ are given by

$$\psi_{(m)}(z)=(1+\bar z z)^{-m}\nabla_{m+1}\nabla_{m+2}\cdots\nabla_j\phi(z), \tag{22}$$

where

$$\nabla_k\equiv\frac{\partial}{\partial z}-\frac{2k\bar z}{1+\bar z z}, \tag{23}$$

$\phi(z)$ is a polynomial of degree $2j$, and the normalization is given by

$$\|\psi_{(m)}\|^2=\int|\psi_{(m)}(z)|^2[dz] \tag{24}$$

($[dz]$ denotes the invariant volume on $G/H=S^2$).

Theorem 3: The vectors $\tilde\psi_{(m)}(z)\equiv(1+\bar z z)^m\psi_{(m)}(z)$, with $\psi_{(m)}$ given by Theorem 2, span the invariant subspace with Casimir $j(j+1)$ of the prequantized orbit[7,11] $\mathcal{O}_m$ through the point $(0,0,m)$ in the coadjoint representation of SU(2).

Proof of Theorem 2: Let us compute $\langle z_{(m)}|\psi\rangle$, with $|z_{(m)}\rangle$ given by Eq. (18). The raising operator J_+ acts as $\partial/\partial z$ when acting on polynomials $\psi(z)$ of Eq. (17). Apart from a normalization factor we get

$$\begin{aligned}\langle z_{(m)}|\psi\rangle&=\mathcal{N}'\langle jj|J_+^{j-m}U(g)^\dagger|\psi\rangle\\&=\mathcal{N}\left(\frac{\partial}{\partial z}\right)^{j-m}(U(g)^\dagger\psi)(z)|_{z=0}\\&=\mathcal{N}\left(\frac{\partial}{\partial z}\right)^{j-m}\left[(\bar\alpha-\bar\beta z)^{2j}\psi\left(\frac{\alpha z+\beta}{-\bar\beta z+\bar\alpha}\right)\right]_{z=0}.\end{aligned}$$

Setting $\zeta=\beta/\bar\alpha$ we get

$$\langle z_{(m)}|\psi\rangle=\mathcal{N}\alpha^{-2m}\left[(1+\bar\zeta\zeta)^2\frac{\partial}{\partial\zeta}\right]^{j-m}(1+\bar\zeta\zeta)^{-2j}\psi(\zeta).$$

Now the point is that one can roll the factor $(1+\bar\zeta\zeta)^{-2j}$ to the left of $\partial/\partial\zeta$ converting it to the operators ∇_k,

$$(1+\bar\zeta\zeta)^2\frac{\partial}{\partial\zeta}(1+\bar\zeta\zeta)^{-2k}=(1+\bar\zeta\zeta)^{2-2k}\nabla_k.$$

Taking into account the contribution of $\mathcal{N}$ one gets the final form of Eq. (22). ◁

Proof of Theorem 3: First, we note that to compare the functions of Theorem 2 with functions defined over the orbit $\mathcal{O}_{(m)}$ we absorbed the factor $(1+\bar\zeta\zeta)^{-m}$ in order to have the correct normalization pertaining to the prequantized $\mathcal{O}_{(m)}$. The Casimir operator in the Hilbert space built over $\mathcal{O}_{(m)}$ by the prequantization procedure is given by

$$\mathscr{C}_m=-(1+\bar z z)^2\left(\frac{\partial}{\partial z}-\frac{2m\bar z}{1+\bar z z}\right)\frac{\partial}{\partial\bar z}$$

[having subtracted the ground state $m(m+1)$]. We have to prove that $\mathscr{C}_m\tilde\psi_{(m)}=[j(j+1)-m(m+1)]\tilde\psi_{(m)}$. Since $\mathscr{C}_m=-(1+\bar z z)^2\nabla_m(\partial/\partial z)$, the result follows by iterating the commutation relation

$$\begin{aligned}\mathscr{C}_m\nabla_{m+1}&=-(1+\bar z z)^2\nabla_m\frac{\partial}{\partial\bar z}\nabla_{m+1}\\&=\nabla_{m+1}(\mathscr{C}_{m+1}+2(m+1))\end{aligned}$$

obtained using

$$(1+\bar z z)^2\nabla_k=\nabla_{k+1}(1+\bar z z)^2,$$

$$(1+\bar z z)^2\left[\frac{\partial}{\partial\bar z},\nabla_k\right]=-2k.$$

Eventually we reach $\mathscr{C}_j$ on the right which annihilates $\psi(z)$ and the additive terms cumulate to $j(j+1)-m(m+1)$. In particular for $m=0$ this is just a hard way to calculate spherical harmonics! ◁

In other words, the coherent states built over $|jm\rangle$ "live" in the prequantized orbit with spin m where they span an *excited Landau level*, in the language of Ref. 11, i.e., at the right level with quantum spin j. It is thus clear why the diagonal matrix elements over these coherent states *feel* the classical spin m. This has been noticed in a special case by Klauder and Skagerstam[4] in the case of spin 1. In particular, this implies that in the case $m=0$ the diagonal matrix elements of any $H\in\mathfrak{g}$ vanish and, unlike the standard case $m=j$, they do not identify H as an operator, which is bad for the CS basis, but it is just right for the calculation of Berry's phase (which indeed vanishes).

All the results obtained in this section can be readily converted to the case of discrete series representations of the noncompact group SU(1,1). In this case the coset space G/H is the upper sheet of the two sheets hyperboloid. The index j representation now is given by

$(U(g)\psi)(z) = (-\bar{\beta}z+\alpha)^{-2j}\psi((\bar{\alpha}z-\beta)/(-\bar{\beta}z+\alpha)),$

where

$$g=\begin{pmatrix}\alpha & \beta\\ \bar{\beta} & \bar{\alpha}\end{pmatrix}$$

and $\psi(z)$ is a function analytic in the disk $D=\{z\colon |z|<1\}$ properly normalizable in the orbit $\mathscr{O}_{(j)}$.[5] The reproducing kernel in this case is $\langle z|\zeta\rangle = (1-\bar{z}z)^j(1-\bar{\zeta}\zeta)^j/(1-z\bar{\zeta})^{2j}$.

Constructing the CS built over $\{|jm\rangle\,|m=j,j+1,...\}$ as previously done in the SU(2) case we have

$$\langle\zeta_{(m)}|\zeta'_{(m)}\rangle = \Lambda(d(\zeta,\zeta'))\cdot\frac{(1-\zeta\bar{\zeta})^j(1-\zeta'\bar{\zeta}')^j}{(1-\zeta\bar{\zeta}')^{j+m}(1-\bar{\zeta}\zeta')^{j-m}},$$

where as expected

$$\Lambda(x)=\sum_{k=0}^{m-j}\frac{\Gamma(j+m+k)}{k!\Gamma(j+m)}\binom{m-j}{k}x^{2k}$$

and $d(z,z')=|(z-z')/(1-\bar{z}'z)|$ is the invariant distance on the hyperboloid. The extension to this case of the theorems already proven for SU(2) is immediate.

IV. CONCLUSIONS

It appears that the quantum (time-dependent) Hamiltonians considered in this paper are of a very special kind; they are in a sense "generalized" free systems in a given dynamical group G and time dependence does not constitute a serious problem. Actually, the quantum dynamics is isomorphic to the classical one, so these systems are necessarily very stable also at the classical level; since an invariant distance is defined in phase space (Kähler metric), these systems can be ergodic but not mixing. In our opinion, it is thus dangerous to exercise our intuition only on such examples, because they may turn out to be too special. Nevertheless, a remarkable number of systems *can* be described in this way, and our effort has been to give a general technique to study them. One result to be noticed for these systems is that Berry's phase can be defined even independently of the adiabatic approximation (as pointed out by Aharonov and Anandan),[12] but it unveils itself for what it really is, namely the piece $\int p\,dq$ of the classical action on the coset space G/H. The algebraic approach to these systems also clarifies what should be considered as the effective parameter space of the system in consideration. In fact, since the time dynamics involves an infinite series of successive commutators of the form $[\cdots[H(t),H(t')],...]$, as shown explicitly by Eq. (6), it is the full Lie algebra which is relevant for discussing Berry's phase, even if the Hamiltonian depends explicitly on a reduced number of parameters.[13]

Finally, it will be interesting to explore what happens to our results (in particular Theorem 3) when dealing with groups of higher rank and weights with multiplicity greater than one; it is known that non-Abelian connections should be considered in this case.[14]

ACKNOWLEDGMENTS

We would like to warmly thank Ennio Gozzi for introducing us to the subject of Berry's phase and for stimulating discussions. One of us (E.O.) warmly thanks Professor John R. Klauder for enlightening discussions on quantization theory which provided some key ideas in the development of this paper.

APPENDIX: THE DISPLACED HARMONIC OSCILLATOR

Let $H(t)=(a^\dagger-c(t))(a-\bar{c}(t))$ where $a^\dagger,a$ are ordinary bosonic operators and $c(t)$ is any continuous curve in the complex plane. Obviously, we have $H(t)=U(t)a^\dagger aU(t)^\dagger$ with $U(t)=\exp\{\bar{c}(t)a^\dagger-c(t)a\}$, which is a translation. The group acting in this example is the Euclidean group E(2) generated by $a^\dagger a,\ a,\ a^\dagger$. We calculate the quantum dynamics both in the interaction picture and by the method of characteristics. It holds that

$$w(t)=T\exp\left\{-i\int_0^t(a^\dagger a-c(\tau)a-\bar{c}(\tau)a^\dagger+|c(\tau)|^2)d\tau\right\}. \tag{A1}$$

Applying the interaction picture, we have to compute

$$T\exp\left\{i\int_0^T(c(t)ae^{-it}+\bar{c}(t)a^\dagger e^{it})\right\}dt,$$

which is easily calculated by the continuous Baker–Hausdorff formula.[8] The result is

$$w(t)=e^{i\phi(t)}e^{-ita^\dagger a}e^{\xi(t)a-\bar{\xi}(t)a^\dagger}, \tag{A2}$$

where

$$\xi(t)=i\int_0^t c(\tau)e^{-i\tau}\,d\tau, \tag{A3}$$

$$\Phi(t)=-\int_0^t|c(\tau)|^2\,d\tau+\mathrm{Im}\left\{\int_0^{\xi(t)}\xi\,d\bar{\xi}\right\}. \tag{A4}$$

If the initial state is a coherent state $|z_0\rangle\equiv\exp(\bar{z}_0a^\dagger-z_0a)|0\rangle$, we get

$$w(t)|z_0\rangle=e^{i\Phi(t)+i\,\mathrm{Im}\{\xi(t)\bar{z}_0\}}|(z_0-\xi(t))e^{it}\rangle. \tag{A5}$$

By setting $z(t)\equiv(z_0-\xi(t))e^{it}$ we readily find

$$w(t)|z_0\rangle=|z(t)\rangle\cdot\exp\{i\alpha(t)\}, \tag{A6}$$

$$\alpha(t)=-\int|z(t)-c(t)|^2\,dt+\mathrm{Im}\left\{\int\bar{z}\,dz\right\}, \tag{A7}$$

which is of the general form of Eq. (15).

The same result can be obtained by the method of characteristics. Since $a\to\partial/\partial z, a^\dagger\to z$, the Schrödinger equation is just

$$i\frac{\partial\psi}{\partial t}=(z-c(t))\left(\frac{\partial}{\partial z}-\bar{c}(t)\right)\psi \tag{A8}$$

and Eq. (9) reads in this case

$$i\,dt=\frac{dz}{z-c(t)}=\frac{d\psi}{\bar{c}(t)(z-c(t))\psi}, \tag{A9}$$

with first integrals

$I_1(z,t)=e^{-it}z+\xi,$

$$I_2(z,\psi,t) = \psi\cdot\exp\left\{\bar{\xi} z e^{-it} + \bar{\xi}\xi + i\int \bar{c}c\, dt + i\int \bar{c}\xi\, dt\right\},$$

which yields the same result [with ξ given by Eq. (A3)]. Notice that in the SU(2) case $z(t)$ satisfies a Riccati equation whose general solution cannot be given in closed form. Going back to the displaced harmonic oscillator, if the initial state is some coherent state built over an excited state $|z_{0(n)}\rangle = e^{\bar{z}_0 a^\dagger - z_0 a}|n\rangle$ we still can apply Eq. (A2) to get the *same* result as before, except for an additional phase e^{-int} coming from the expectation value of $H(t)$. In fact, the reproducing kernel $\langle z_{(n)}|\zeta_{(n)}\rangle$ for the CS built over $|n\rangle$ coincides with the standard one of the Glauber coherent state, except for a prefactor depending only on $|z-\zeta|^2$, which does not contribute to the connection form. The explicit calculation yields

$$\langle z_{(n)}|\zeta_{(n)}\rangle = \Xi(|z-\zeta|)e^{\mathrm{Im}\{z\bar{\zeta}\}}, \tag{A10}$$

where

$$\Xi(x) \equiv \sum_{k=0}^{n} \frac{1}{k!}\binom{n}{k}(-x^2)^k \exp\left(-\frac{1}{2}x^2\right).$$

As a consistency check, notice that Eq. (A10) can also be obtained by Eq. (20) according to the contraction $\mathrm{SU}(2)\to\mathrm{E}(2)$.[5,15]

[1] M. V. Berry, Proc. R. Soc. London A **392**, 45 (1984); J. Moody, A. Shapere, and F. Wilczek, Phys. Rev. Lett. **56**, 893 (1986); R. Jackiw, Comments At. Mol. Phys. **21**, 71 (1988).

[2] P. Nelson and L. Alvarez-Gaume, Commun. Math. Phys. **99**, 103 (1985); A. Niemi and G. W. Semenoff, Phys. Rev. Lett. **55**, 927 (1985); M. Stone, Phys. Rev. D **33**, 1191 (1986); H. Sonoda, Phys. Lett. B **156**, 220 (1985); Nucl. Phys. B **266**, 410 (1986); A. Niemi, G. W. Semenoff, and Y. S. Wu, *ibid.* **276**, 173 (1986).

[3] S. Chaturvedi, M. S. Sriram, and V. Srinivasan, J. Phys. A **20**, L1071 (1987); T. F. Jordan, J. Math. Phys. **29**, 2042 (1988); R. Jackiw, Int. J. of Mod. Phys. A **3**, 285 (1988); L. Vinet, Phys. Rev. D **37**, 2369 (1988).

[4] J. R. Klauder and Bo-S. Skagerstam, *Coherent states, Applications in Physics and Mathematical Physics* (World Scientific Singapore, 1985).

[5] A. Perelomov, *Generalized Coherent States and their Applications* (Springer, Berlin, 1986).

[6] E. Onofri, "Quantization theory for homogeneous Kuahler manifolds," Parma University, 1974; J. Math. Phys. **16**, 1087 (1975).

[7] N. P. Whoodhouse, *Geometric Quantization* (Oxford U. P., Oxford, 1980); V. Guillemin and S. Sternberg, *Geometric Asymptotics*, Math. Surveys No. 14 (Amer. Math. Soc., Providence, RI, 1977).

[8] W. Magnus, Commun. Pure Appl. Math. **VII**, 649 (1954).

[9] A. Kirillov, *Eléments de la Théorie des Représentations* (MIR, Moscow, 1974); B. Kostant, in *Lecture Notes in Mathematics*, edited by C. C. Tamm (Springer, New York, 1970).

[10] E. Goursat, *A Course in Mathematical Analysis* (Dover, New York, 1959), Vol. II, part 2, Chap. V.

[11] J. R. Klauder, Ann. Phys. **188**, 120 (1988); J. R. Klauder and E. Onofri, submitted to Int. J. Mod. Phys. A.

[12] Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).

[13] G. Giavarini, E. Gozzi, D. Rohrlich, and W. D. Thacker, to be published in Phys. Rev. D.

[14] F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).

[15] E. Onofri and M. Pauri, Lett. Nuovo Cimento **3**, 35 (1972).

Ring characterization of scalar product spaces and Gel'fand triples

Paweł Kruszyński
Department of Mathematics and Informatics, Delft University of Technology, P.O. Box 356, 2600 AJ Delft, The Netherlands

(Received 9 March 1988; accepted for publication 26 October 1988)

It has been proved that if an involution $^+$ exists in a subring of continuous linear maps in a locally convex topological vector space $\mathscr{S}$ with the property $(\xi A)^+ = \xi^* A^+$, and $A^+A = 0$ iff $A = 0$, then there exists in $\mathscr{S}$ a Hermitian scalar product $\langle\,,\,\rangle$ such that $^+$ is the adjoint operation with respect to $\langle\;,\;\rangle$. The existence of a corresponding Gel'fand triple $\mathscr{S}\subset H\subset\mathscr{S}'$ has been shown.

I. INTRODUCTION

It is often silently assumed that the collection of observables describing measurable quantities concerning a physical system are represented as operators in an inner product (or scalar product) space, which can be seen as a member of a Gel'fand triple. However, there arises a natural question whether this geometrical structure is indeed determined by a structure of the collection of observables. And if it is so, then what are the conditions that imply this structure? Taking it one step further, one may ask what is an algebraic and topological characterization of abstract algebras over the field of complex numbers (which the observables are believed to constitute) to make them representable as the algebras of so-called extendable operators in a Gel'fand triple.[1] In this general approach it is necessary to find the underlying representation space. However, the usual technique, extending the Gel'fand–Naimark–Segal construction is in many cases not applicable, unless the space of positive functionals (if they can be defined at all) is rich enough. Therefore in our paper we consider a first approach to this general problem, assuming from the very beginning that our algebra already consists of operators in a locally convex topological vector space. Under this assumption we prove that the algebraic condition providing the desired Gel'fand triple is the existence of an involution in the algebra, such that for any operator A, $A^+A = 0$ if and only if $A = 0$, i.e., the involution is proper.[2]

This kind of result is known[3] for operators in Banach spaces, and the proof presented by Kakutani and Mackey can be partially adapted to our "unbounded" situation. However, a limitation impossible to overcome is connected with the necessity of our assumption that the involution $^+$ is antilinear when restricted to the complex numbers. Namely, the continuity of the associated automorphism of the field of complex numbers does not follow from our condition, even in the infinite-dimensional case, as it happens for Banach spaces.[3]

From the physical point of view the existence of an involution in the algebras of observables has been well established. Although there is no direct interpretation of the operation of involution, it is clear that it may be connected with the time relations in a physical system, with the time reversibility, and with representation of field operators in the quantum field theory. Also the property of being proper can find an easy physical justification, e.g., nilpotent observables can be excluded from the description of the physical system.

It is worthwhile to mention that the extendable operator *-algebras, even though their involution is proper, need not be Rickart *-rings.[2] Thus our result provides mathematically interesting and tangible examples of a very general type of *-ring, yet closely connected with the Hilbert space operator algebras.

Now we shall introduce some definitions needed.

Let H be a Hilbert space, and let $\mathbf{L}_H$ denote the lattice of all closed subspaces of H. Then $\mathbf{L}_H$ is orthocomplemented in a natural way. However, if instead of $\mathbf{L}_H$ we consider the (not complete) lattice $\mathbf{L}_H^{\text{fin}}$ of finite-dimensional subspaces of H, then the orthocomplementation can be defined only in a relative way.

Definition 1.1: A lattice $\mathbf{L}$ of subspaces of a vector space $\mathscr{S}$ is called relatively orthocomplemented if (a) every interval in $\mathbf{L}$,

$$[0,L\,] = \{Q\in\mathbf{L}\colon 0\leqslant Q\leqslant L\},$$

is an orthocomplemented lattice with the (relative) orthocomplementation map

$$\perp_L\colon [0,L\,]\to[0,L\,],$$

i.e., if $L_1\leqslant L_2\leqslant L$, then $\perp_L(L_2)\leqslant\perp_L(L_1)$, and $\perp_L(\perp_L(K)) = K$, $\perp_L(K)\wedge K = 0$, $\perp_L(K)\vee K = L$ for every $K\in[0,L]$; and (b) the family of relative orthocomplementations $\{\perp_L\}_{L\in\mathbf{L}}$ is compatible, i.e., if $L_1\leqslant L_2$, then for every $L\in[0,L_1]$,

$$\perp_{L_2}(L)\wedge L_1 = \perp_{L_1}(L).$$

We say that $L_1,L_2\in\mathbf{L}$ are orthogonal to each other if there exists $L\in\mathbf{L}$ with $L_1,L_2\leqslant L$, and $L_1\leqslant\perp_L(L_2)$. Then we write $L_1\perp L_2$.

In what follows we shall consider lattices $\mathbf{L}$ of subspaces of a vector space $\mathscr{S}$ such that $\mathbf{L}_{\mathscr{S}}^{\text{fin}}\subset\mathbf{L}$.

For relatively orthocomplemented lattices we have the following extension of the von Neumann theorem.[4]

Theorem 1.2: Let $\mathscr{S}$ be a vector space over $\mathbb{C}^1$, and let $\mathbf{L}$ be a relatively orthocomplemented lattice of linear subspaces of $\mathscr{S}$ such that (a) $\mathbf{L}$ contains all finite-dimensional subspaces of $\mathscr{S}$, i.e., $\mathbf{L}_{\mathscr{S}}^{\text{fin}}\subset\mathbf{L}$; and (b) if $M,N\in\mathbf{L}$ and one of them is finite dimensional, then $M+N = M\vee N$. Then for every nonzero $s_0\in\mathscr{S}$ there exists an involutive automorphism Θ of the field of complex numbers $\mathbb{C}^1$, and a Θ-symmetric Θ-semisesquilinear form $\langle\;,\;\rangle$ on $\mathscr{S}\times\mathscr{S}$ such that $\langle s_0,s_0\rangle = 1$, and $\langle x,y\rangle = 0$ if and only if $\langle x\rangle\perp\langle y\rangle$, where $\langle z\rangle$ is the (one-dimensional) subspace spanned by $z\in\mathscr{S}$. [We recall that $\langle\,,\,\rangle$ is a nondegenerate Θ-semisesquilinear form if $\Theta(\langle x,y\rangle) = \langle y,x\rangle$, $\langle\alpha x,\beta,y\rangle = \alpha\langle x,y\rangle\Theta(\beta)$, and $\langle x,x\rangle = 0$ iff $x = 0$.]

 0022-2488/89/030664-04$02.50

Proof: Let $F \in \mathbf{L}_{\mathscr{S}}$ be finite dimensional. Thus $F \in \mathbf{L}$. If $L \in \mathbf{L}_{\mathscr{S}}$ and $L \leqslant F$, then $L \in \mathbf{L}$, too. We have

$$F = \perp_F(L) \vee L = \perp_F(L) + L.$$

Let $s_0 \in \mathscr{S}$. Denote by $\mathscr{F}$ the class of finite-dimensional linear subspaces of $\mathscr{S}$ such that $s_0 \in K$ for all $K \in \mathscr{F}$, and $\dim K \geqslant 3$. By the von Neumann theorem (see Ref. 4, p. 61, Theorem 4.6) there exists an involutive automorphism Θ_F of $\mathbb{C}^1$ and a Θ_F-semisesquilinear form $\langle\ ,\ \rangle_F$ in $\mathscr{S} \times \mathscr{S}$ for every $F \in \mathscr{F}$ such that for any $L \leqslant F$,

$$\perp_F(L) = \{u \in F: \langle u,x \rangle_F = 0 \text{ for all } x \in L\},$$

and $\langle s_0, s_0 \rangle_F = 1$. Now let $F_1, F_2 \in \mathscr{F}$ with $F_1 \leqslant F_2$. Then $\langle\ ,\ \rangle_{F_2}|_{F_1 \times F_1}$ gives the same orthocomplementation on $[0,F_1]$ as $\langle\ ,\ \rangle_{F_1}$, and $\langle s_0,s_0 \rangle_{F_1} = \langle s_0,s_0 \rangle_{F_2} = 1$. By the uniqueness $\langle\ ,\ \rangle_{F_2}|_{F_1 \times F_1} = \langle\ ,\ \rangle_{F_1}$ and $\Theta_{F_1} = \Theta_{F_2}$.

Since for any $L_1, L_2 \in \mathscr{F}$, $L_1 \vee L_2 \in \mathscr{F}$, there exist a unique automorphism Θ on $\mathbb{C}^1$ and a unique Θ-semisesquilinear form $\langle\ ,\ \rangle$ on the whole $\mathscr{S} \times \mathscr{S}$ which are extensions of all Θ_F and $\langle\ ,\ \rangle_F$, respectively. □

Remark 1.3: If $\mathscr{S}$ is a Banach space, $\dim \mathscr{S} \geqslant \infty$, and $\mathbf{L}$ consists of closed subspaces, then $\mathscr{S}$ is isometrically isomorphic with a Hilbert space, with the scalar product $\langle\ ,\ \rangle$, and Θ is the complex conjugation. However, the problem of continuity of Θ is still open for general locally convex spaces, e.g., for Fréchet spaces.[3–5]

II. RING CHARACTERIZATION OF GEL'FAND TRIPLES

Let us consider the triple $\Phi \subset H \subset \Phi'$ (the so-called Gel'fand triple), where Φ is a locally convex topological vector space, Φ' is its strong dual, H is a Hilbert space, and the embeddings are continuous injections. If $A \in L(\Phi)$, then A can be regarded as a (in general unbounded) linear operator in the Hilbert space H, with a dense domain $\mathbf{D}(A)$ containing Φ. If $\Phi \subset \mathbf{D}(A^*)$, $A^*: \Phi \to \Phi$, and $A^*|_\Phi$ is continuous in Φ (a number of conditions providing this property can be found in Ref. 1), then we can define $A^+ = A^*|_\Phi$. Then $A^+ \in L(\Phi)$. The elements $A \in L(\Phi)$, for which the involution $A \to A^+$ is well defined, form a subalgebra $\mathfrak{A}$ of $L(\Phi)$, the so-called extendable maps algebra (cf. Ref. 1). The involution $^+$ has the following properties.

(i) $(\lambda A)^+ = \lambda^* A^+$, for any $\lambda \in \mathbb{C}^1$, where λ^* is the complex conjugation of λ.

(ii) $(A+B)^+ = A^+ + B^+$ and $(AB)^+ = B^+A^+$.

(iii) $A^+A = 0$ if and only if $A = 0$.

Since all elements of $\mathfrak{A}$ have the common dense invariant domain Φ, then the above properties easily follow from the properties of the operation of taking the adjoint of an operator in a Hilbert space.

Suppose now that we are given a locally convex topological vector space $\mathscr{S}$. It is of interest whether the space $\mathscr{S}$ can be included in a suitable Gel'fand triple. In quantum mechanics, common invariant domains are seen through the collection of unbounded operators (observables) defined on them. The space Φ has the interpretation of the test function space or, in other words, of the initial state space, in the language of Dirac's formalism of quantum mechanics. Therefore it seems essential to reconstruct the Gel'fand triple structure out of the properties of the algebra of observables.

We present now our main result solving this problem, which seems to be interesting also in a wider mathematical context (cf. Refs. 5 and 6).

Theorem 2.1: Let $\mathscr{S}$ be a locally convex topological vector space, and let $\mathfrak{A} \subset L(\mathscr{S})$ be an algebra over $\mathbb{C}^1$ with the following properties. (a) $\mathfrak{A}$ contains all elements of $L(\mathscr{S})$ with rank 1. (b) An involution $^+$ is defined in $\mathfrak{A}$ with properties (i)–(iii).

Then for every nonzero $s_0 \in \mathscr{S}$ there exists on $\mathscr{S} \times \mathscr{S}$ a sesquilinear nondegenerate form $\langle\ ,\ \rangle$ separately continuous in its arguments and such that $^+$ is the conjugation operation in $\mathfrak{A}$ with respect to $\langle\ ,\ \rangle$, i.e., for every $s,z \in \mathscr{S}$ and $A \in \mathfrak{A}$, $\langle As,z \rangle = \langle s,A^+z \rangle$, and $\langle s_0,s_0 \rangle = 1$.

Proof: We shall prove the theorem in a few steps using several lemmas. The idea of the proof is similar to that in Ref. 3; however, some points must be proved independently. To make the paper self-consistent we give here the full proof.

Let $\mathscr{C} \subset \mathfrak{A}$. Denote

$$\mathscr{C}^r = \{A \in \mathfrak{A}: CA = 0 \text{ for all } C \in \mathscr{C}\},$$
$$\mathscr{C}^l = \{A \in \mathfrak{A}: AC = 0 \text{ for all } C \in \mathscr{C}\}.$$

Following Ref. 3 we say that a set $\mathscr{C} \subset \mathfrak{A}$ is a right annihilator if $\mathscr{C}^{lr} = (\mathscr{C}^l)^r = \mathscr{C}$.

Lemma 2.2: A subset $\mathscr{C} \subset \mathfrak{A}$ is a right annihilator if and only if there exists in $\mathbf{L}_{\mathscr{S}}$ a closed subspace M such that

$$\mathscr{C} = \{A \in \mathfrak{A}: A\mathscr{S} \subset M\}.$$

Proof: Let $\mathscr{C} \subset \mathfrak{A}$ and $\mathscr{C} = \mathscr{C}^{lr}$. Define the following (closed) subspace of $\mathscr{S}$:

$$\phi(\mathscr{C}) = \bigcap_{B \in \mathscr{C}^l} \operatorname{Ker} B = M_{\mathscr{C}}. \tag{1}$$

Then for all $A \in \mathscr{C}$ and all $B \in \mathscr{C}^l$ we have $BA = 0$ and $BA\mathscr{S} = \{0\}$, i.e., $A\mathscr{S} \subset \operatorname{Ker} B$. Thus $A \in \mathscr{C}$ implies $A\mathscr{S} \subset M_{\mathscr{C}}$, that is,

$$\mathscr{C} \subset \{A \in \mathfrak{A}: A\mathscr{S} \subset M_{\mathscr{C}}\}.$$

Now let $A \in \mathfrak{A}$ be such that $A\mathscr{S} \subset M_{\mathscr{C}}$. Then for every $B \in \mathscr{C}^l$, $BA = 0$; thus $A \in \mathscr{C}^{lr} = \mathscr{C}$. Hence $\mathscr{C} \supset \{A \in \mathfrak{A}: A\mathscr{S} \subset M_{\mathscr{C}}\}$. This proves the sufficiency of the condition.

Now let $M \in \mathbf{L}_{\mathscr{S}}$. Define

$$v(M) = \{A \in \mathfrak{A}: A\mathscr{S} \subset M\} = \mathscr{C}_M. \tag{2}$$

Naturally $v(M) \subset \mathscr{C}^{lr}$. Suppose that there exists $A \in \mathscr{C}^{lr}$ not in $v(M)$, i.e., there exists $x_0 \in \mathscr{S}$ such that $Ax_0 \notin M$. Since M is closed, then there exists a continuous linear functional $f \in \mathscr{S}'$ such that $f(Ax_0) = 1$ and $f|_M = 0$ (see Ref. 7). Define a linear operator C by the formula $Cs = f(s)Ax_0$ for every $s \in \mathscr{S}$. Then clearly $C \in \mathfrak{A}$. For every $P \in v(M)$ we have $CPs = f(Ps)Ax_0 = 0$, i.e., $CP = 0$ [since by the definition of $v(M)$, $Ps \in M$]. Thus $C \in \mathscr{C}^l$. Hence $CA = 0$ and $CAx_0 = 0 = f(Ax_0)Ax_0 = Ax_0$. There is contradiction. Thus $A \in v(M)$, and eventually

$$v(M) = \mathscr{C}_M = (\mathscr{C}_M)^{lr} = \mathscr{C}^{lr}. \qquad \square$$

Lemma 2.3: The lattice $\mathbf{L}$ of right annihilators in the algebra $\mathfrak{A}$ is orthocomplemented. The orthocomplementation map is given by the formula

$$\mathbf{L} \ni \mathscr{C} \to \perp(\mathscr{C}) = \mathscr{C}^{+r} \in \mathbf{L}, \text{ where } \mathscr{C}^+ = \{C^+: C \in \mathscr{C}\}.$$

Proof: First note that if $\mathscr{C}$ is a right annihilator, so is

$\perp(\mathscr{C})$ by property (ii). Further, since $\mathscr{C}^{lr}=\mathscr{C}$ and $A^{++}=A$, we have $\perp(\perp(\mathscr{C}))=\mathscr{C}$. Finally, $\perp(\mathscr{C})\wedge\mathscr{C}=\{0\}$ because of property (iii). □

Lemma 2.4: The maps defined in Eqs. (1) and (2) have the properties

(a) $\phi\cdot v=\mathrm{id}_{\mathbf{L}_{\mathscr{S}}}$,

(b) $v\cdot\phi=\mathrm{id}_{\mathbf{L}}$.

Proof: (a) Let $M\in\mathbf{L}_{\mathscr{S}}$. Compute

$$M_1=(\phi\cdot v)(M)=\phi(\{C\in\mathfrak{A}\colon C\mathscr{S}\subset M\})=\bigcap_{B\in v(M)^l}\operatorname{Ker}B.$$

Clearly $M\subset M_1$. Indeed, if $m\in M$ and $BC=0$ for every $C\in v(M)$, then, in particular for $C=\mathrm{Proj}_m$ (the projection onto the one-dimensional subspace $\langle m\rangle$, with $\mathrm{Proj}_m m=m$), we have $Bm=0$; thus $m\in\operatorname{Ker}B$. This holds for all $B\in v(M)^l$. To prove that $M_1\subset M$ assume that there exists

$$m\in\bigcap_{B\in v(M)^l}\operatorname{Ker}B,$$

with $m\notin M$. Then again there exists a functional $f\in\mathscr{S}'$ such that $f|_M=0$ and $f(m)=1$. Set $Bs=f(s)m$. Then $B\in\mathfrak{A}$, and for every $C\in v(M)$ we have $BCs=f(Cs)m=0$. Thus $B\in v(M)^l$. Hence $Bm=0=f(m)m=m$, which yields a contradiction. Thus $m\in M$. In other words, $(\phi\cdot v)(M)=M$, i.e., $\phi\cdot v=\mathrm{id}_{\mathbf{L}_{\mathscr{S}}}$.

(b) Let $\mathscr{C}\subset\mathfrak{A}$ be a right annihilator. Compute

$$\mathscr{C}_1=(v\cdot\phi)(\mathscr{C})=v\Big(\bigcap_{B\in\mathscr{C}^l}\operatorname{Ker}B\Big)=\Big\{C\in\mathfrak{A}\colon C\mathscr{S}\subset\bigcap_{B\in\mathscr{C}^l}\operatorname{Ker}B\Big\}.$$

For each $C\in\mathscr{C}$ we have $BC=0$ for $B\in\mathscr{C}^l$; thus $C\mathscr{S}\operatorname{Ker}B$, i.e., $\mathscr{C}\subset\mathscr{C}_1$. On the other hand, if $C\in\mathscr{C}_1$, then for all $B\in C^l$, $BC\mathscr{S}=\{0\}$, i.e., $C\in\mathscr{C}^{lr}=\mathscr{C}$. Hence $\mathscr{C}_1=\mathscr{C}$, and $v\cdot\phi=\mathrm{id}_{\mathbf{L}}$. □

Notice that the collection $\mathbf{L}$ of all right annihilators in $\mathfrak{A}$ forms a lattice with the usual operation of inclusion and the meets and joints defined as follows: for $\mathscr{C}_1,\mathscr{C}_2\in\mathbf{L}$,

$$\mathscr{C}_1\wedge\mathscr{C}_2=\mathscr{C}_1\cap\mathscr{C}_2\quad\text{and}\quad\mathscr{C}_1\vee\mathscr{C}_2=(\mathscr{C}_1\cup\mathscr{C}_2)^{lr}.$$

On the other hand, the collection $\mathbf{L}_{\mathscr{S}}$ of closed subspaces of $\mathscr{S}$ is also a lattice with the usual lattice operations $M_1\wedge M_2=M_1\cap M_2$ and $M_1\vee M_2=\overline{(M_1+M_2)}$ (the closure).

Lemma 2.5: The maps ϕ and v defined in (1) and (2) are lattice isomorphisms.

Proof: Let $\mathscr{C}_1,\mathscr{C}_2\in\mathbf{L}$, and let $\mathscr{C}_1\leqslant\mathscr{C}_2$. Then

$$M_1=\phi(\mathscr{C}_1)=\bigcap_{B\in\mathscr{C}_1^l}\operatorname{Ker}B\subset M_2=\bigcap_{B\in\mathscr{C}_2^l}\operatorname{Ker}B.$$

Similarly, for $M_1,M_2\in\mathbf{L}_{\mathscr{S}}$ with $M_1\leqslant M_2$ we have

$$v(M_1)=\{C\in\mathfrak{A}\colon C\mathscr{S}\subset M_1\}\subset\{C\in\mathfrak{A}\colon C\mathscr{S}\subset M_2\}=v(M_2).$$

Thus ϕ and v are mutually inverse order-preserving maps between $\mathbf{L}$ and $\mathbf{L}_{\mathscr{S}}$, and hence they are lattice isomorphisms.

As the corollary to the above lemmas we have the following.

Lemma 2.6: The lattice $\mathbf{L}_{\mathscr{S}}$ is orthocomplemented, with $\perp(M)=\phi(\perp(v(M)))$.

Now it follows from Theorem 1.2 that there exists an automorphism Θ of the field $\mathbb{C}^1$, $\Theta^2=\mathrm{id}$, and a Θ-semisesquilinear form $\langle\ ,\ \rangle$ on $\mathscr{S}\times\mathscr{S}$, such that the orthocomplementation in $\mathbf{L}_{\mathscr{S}}$ is given by the formula

$$\perp(L)=\{u\in\mathscr{S}\colon\langle u,x\rangle=0\text{ for all }x\in L\},$$

where $L\in\mathbf{L}_{\mathscr{S}}$.

Since the kernels of the linear maps $\mathscr{S}\ni s\to\langle s,z\rangle$ are the closed subspaces $\perp(\langle z\rangle)\in\mathbf{L}_{\mathscr{S}}$ for every $z\in\mathscr{S}$, then these maps are continuous. Hence the form $\langle\ ,\ \rangle$ is continuous in the first argument. At the same time Θ-linear map $j\colon\mathscr{S}\to\mathscr{S}'$ is well defined by the formula

$$j(z)(s)=\langle s,z\rangle,\text{ with }s,z\in\mathscr{S}.\tag{3}$$

It is easy to observe that j is an injection.

Lemma 2.7: Let $s,z\in\mathscr{S}$, and let P_s and P_z denote the projections on $\langle s\rangle$ and $\langle z\rangle$, respectively (obviously $P_s,P_z\in\mathfrak{A}$). Then for any $A\in\mathfrak{A}$, $\langle As,z\rangle=0$ if and only if $P_z^+AP_s=0$ (equivalently $P_s^+A^+P_z=0$).

Proof: Assume that $\langle As,z\rangle=0$. Take $x\in\mathscr{S}$. Then $\langle AP_sx,P_zx\rangle=0$, and thus

$$AP_s\mathscr{S}\subset\bigcap_{Q\in\{P_z^+\}^{rl}}\operatorname{Ker}Q=\perp(P_z\mathscr{S}).$$

In particular, since $P_z^+\in\{P_z^+\}^{rl}$, we have

$$P_z^+AP_s=0.$$

Conversely, if $P_z^+AP_s=0$, then $AP_s\in\{P_z^+\}^r$, and for any $B\in\{P_z^+\}^{rl}$, $BAP_s=0$. That is, $AP_s\mathscr{S}\subset\operatorname{Ker}B$. Eventually, $\langle As,z\rangle=0$. □

Lemma 2.8: For every $A\in\mathfrak{A}$, denoting by A' the algebraically transposed operator A, i.e., $A'\in L(\mathscr{S}')$, we have

$$A'\cdot j=j\cdot A^+,$$

i.e., $\langle As,z\rangle=\langle s,A^+z\rangle$, for all $s,z\in\mathscr{S}$.

Proof: Let $z\in\mathscr{S}$, and $A\in\mathfrak{A}$. Define $V_1^0=\operatorname{Ker}(A'\cdot j)(z)$ and $V_2^0=\operatorname{Ker}(j\cdot A^+)(z)$. From Lemma 2.7 it follows that $V_1^0=V_2^0$. Thus the functionals $(A'\cdot j)(z)$ and $(j\cdot A^+)(z)$ are proportional. Without loss of generality we can assume that the algebra $\mathfrak{A}$ has the unit $\mathbf{I}$. Because then $\mathbf{I}^+=\mathbf{I}$ and $\mathbf{I}'=\mathbf{I}_{\mathscr{S}}$, we have, for some $\xi,\xi'\in\mathbb{C}^1$,

$$(\mathbf{I}+A)'\cdot j(z)=\xi j\cdot(\mathbf{I}+A^+)(z)=\xi j(z)+\xi j\cdot A^+(z)$$
$$=j(z)+\xi' j\cdot A^+(z).$$

Thus if $j(z)$ and $j\cdot A^+(z)$ are linearly independent, then $\xi=\xi'=1$. If, however, $j(z)$ and $j\cdot A^+(z)$ are linearly dependent, then let us take an element $B\in\mathfrak{A}$ which has that property. Then $B'\cdot j(z)=j\cdot B^+(z)$, and moreover

$$(B+A)'\cdot j(z)=j\cdot B^+(z)+\xi' j\cdot A^+(z)$$
$$=\xi'' j\cdot B^+(z)+\xi'' j\cdot A^+(z),$$

for some $\xi''\in\mathbb{C}^1$. But by the linear independence of $j\cdot B^+(z)$ and $j\cdot A^+(z)$, we have $\xi'=\xi''=1$. Thus for every $z\in\mathscr{S}$ and every $A\in\mathfrak{A}$, we get $(A'\cdot j)(z)=(j\cdot A^+)(z)$, that is, $\langle As,z\rangle=\langle s,A^+z\rangle$, for all $s,z\in\mathscr{S}$. □

Finally, for any $\xi\in\mathbb{C}^1$ and $A\in\mathfrak{A}$, by property (i) and Lemma 2.8 we have $(\xi A)^+=\xi^*A^+=\Theta(\xi)A^+$; thus

$\Theta(\xi) = \xi^*$, i.e., the automorphism Θ is the complex conjugation in $\mathbb{C}^1$. It follows that the form $\langle\ ,\ \rangle$ is Hermitian, continuous in both arguments, positive, and nondegenerate. This ends the proof of the theorem. □

Corollary 2.9: If the space $\mathscr{S}$ is bornological and reflexive (thus complete), then the injection j is a positive embedding $j: \mathscr{S} \to \mathscr{S}'$ in the sense (cf. Ref. 1) that (a) j is a strongly continuous antilinear map, and (b) j is positive, i.e., for every $s \in \mathscr{S}$, $j(s)(s) > 0$ iff $s \neq 0$.

Then there also exists a Hilbert space H, being the completion of $\mathscr{S}$ with respect to the norm $\|s\| = \langle s,s \rangle^{1/2}$, such that $\mathscr{S} \subset H \subset \mathscr{S}'$, and the embeddings are continuous. That is, there exists a Gel'fand triple such that the involutive structure of the algebra $\mathfrak{A}$ is compatible with the Hilbert space adjoint operation.[1-6]

III. CONCLUDING REMARKS

Our result follows from a rather strong assumption that $(\xi A)^+ = \xi^* A^+$. Without it not much is known. In particular, if the space $\mathscr{S}$ has the Schauder basis, then the construction analogous to that in Ref. 3 can be carried out. However, there are simple examples, such as the sequence spaces ω and ϕ, in which the result fails. There is no positive embedding $\omega \subset \phi = \omega'$; hence there is no adequate Gel'fand triple. Although $\phi \subset l_2 \subset \omega$ in a natural way, it is easy to give simple examples of a situation in which neither "direction" is possible. The problem needs further study (cf. Refs. 5 and 6). On the other hand, the problem of algebraic characterization of extendable map algebras also remains open.

Finally, let us observe that, if it happens that $\mathfrak{A} = L(\mathscr{S})$, i.e., the involution $^+$ is defined on the whole ring of continuous linear maps in $\mathscr{S}$, then $\mathscr{S}$ is a Hilbert space.

Theorem 3.1: Let $\mathscr{S}$ be a bornological reflexive l.c. topological vector space, which is metrizable or whose strong dual is metrizable. Assume that in the ring $L(\mathscr{S})$ of continuous linear maps in $\mathscr{S}$ there exists an involution $^+$ satisfying conditions (i)–(iii). Then $\mathscr{S}$ is topologically isomorphic to a Hilbert space H, and the involution $^+$ is given by the adjoint operation in the set of bounded linear operators in H.

Proof: Since the space $\mathscr{S}$ is reflexive, then each $U \in L(\mathscr{S}')$ is of the form $U = A'$, for some $A \in L(\mathscr{S})$. Let $g \in \mathscr{S}'$; then there exists $U \in (\mathscr{S}')$ such that $g \in Uj(\mathscr{S})$. But then $g = (U \cdot j)(s) = (j \cdot A^+)(s)$, for some $s \in \mathscr{S}$, and $A \in L(\mathscr{S})$. Thus $g \in j(\mathscr{S})$, i.e., the injection $j: \mathscr{S} \to \mathscr{S}'$ is surjective. Now by the closed graph theorem[7] j gives rise to a homeomorphism, since j^{-1} is closed. In particular, the embeddings in the Gel'fand triple $\mathscr{S} \subset H \subset \mathscr{S}'$ are also homeomorphisms. The result follows. □

[1] P. Kruszyński, "Algebras of extendable unbounded operators," EUT-report 84-WSK-04, Eindhoven, 1984.

[2] S. K. Berberian, *Baer*-Rings* (Springer, New York, 1972).

[3] S. Kakutani and G. W. Mackey, "Ring and lattice characterizations of complex Hilbert space," Am. Math. Soc. **52**, 727 (1946).

[4] V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, NJ, 1968), Vol. I.

[5] L. Schwartz, "Sous-espaces hilbertienes d'espaces vectoriels topologiques et noyaux associes (noyaux reproduisants)," J. Analyse Math. **13**, 115 (1964).

[6] P. Sørjonen, "Lattice-theoretical characterizations of inner product spaces," Studia Sci. Math. Hungar. **19**, 141 (1984).

[7] H. H. Schaefer, *Topological Vector Spaces* (Macmillan, New York, 1966).

Quantum inverse scattering method for a nonlinear N-wave resonance interaction system

Huan-Qiang Zhou
Center of Theoretical Physics, CCAST (World Laboratory), Beijing 100080, People's Republic of China and Institute of Solid State Physics, Sichuan Normal University, Chengdu 610066, People's Republic of China

Lin-Jie Jiang
Institute of Solid State Physics, Sichuan Normal University, Chengdu 610066, People's Republic of China

(Received 16 August 1988; accepted for publication 2 November 1988)

The quantum inverse scattering method is used for the study of a nonlinear N-wave resonance interaction system. The Yang–Baxter relations are solved to get various commutation relations for the scattering data operators. The energy spectrum of the quantum Hamiltonian for the model is determined and the existence of the quantum bound states is analyzed. In the classical limit, the corresponding r matrix is found, and the well-established correspondence between the classical solitons and the quantum bound states is examined. Finally, the integrability of the same model but with both fermion fields and boson fields is discussed.

I. INTRODUCTION

There has been considerable recent interest in the study of completely integrable nonlinear evolution equations exhibiting soliton behaviors. In particular, the discovery of the inverse scattering method makes it possible to find a wide class of two-dimensional nonlinear evolution equations, such as the KdV equation, the nonlinear Schrödinger equation, and the sine–Gordon equation.[1–4] As is well known, the Hamiltonian systems corresponding to these equations have an infinite number of conservation laws that are in involution with each other. Recent developments show that the quantum version of the inverse scattering method may be established (for review, see Ref. 5). Now this method has been successfully applied to study many completely integrable models in $(1+1)$-dimensional quantum field theory and in two-dimensional lattice statistics. Nevertheless, it provides a unified and elegant formulation for many techniques previously developed in different branches of one-dimensional mathematical physics, including Baxter's commuting transfer matrix technique in two-dimensional lattice statistics,[6] the Bethe–Ansatz technique for the $(1+1)$-dimensional quantum field theory,[7] and the construction of infinite number of conservation laws in completely integrable models.[8] Therefore, the quantum inverse scattering method (QISM) has been a power tool for getting explicit results from completely integrable systems. In a recent paper, Kulish[9] studied the quantum, nonlinear, three-wave resonance interaction model in the framework of the quantum inverse scattering method. At the classical level, its integrability is well known for a long time.[10] The same model but with two fermion fields and one boson field has also been carried out by Wang and Pu.[11] All these results show that the quantum, nonlinear, three-wave interaction models with three choices of statistics proposed by Ohkuma and Wadati[12] are completely integrable. In the present paper, we show that the quantum N-wave resonance interaction system is also completely integrable via QISM. To do so, the Yang–Baxter relations are solved and the energy spectrum of the Hamiltonian is determined. Further, the same model but with a different choice of statistics is discussed.

II. THEORY

The system discussed in the present paper is given by the Hamiltonian

$$H=\int\left[\sum_{j<k} v_{jk}\frac{w_{jk}^{+}}{i}\frac{\partial}{\partial x}w_{jk}\right.$$
$$\left.+\sum_{i<j<k}\epsilon_{ijk}\left(w_{ij}w_{jk}w_{ik}^{+}+w_{ik}w_{jk}^{+}w_{ij}^{+}\right)\right]dx\,. \quad (1)$$

Here, $w_{ij}(x)$ and $w_{ij}^{+}(x)$ are boson fields that satisfy the usual equal time commutation relations,

$$[w_{ij}(x),w_{kl}(y)]=[w_{ij}^{+}(x),w_{kl}^{+}(y)]=0\,,$$
$$[w_{ij}(x),w_{kl}^{+}(y)]=\delta_{ik}\delta_{jl}\delta(x-y), \quad (2)$$
$$i<j,\quad k<l,\quad i,j,k,l=1,2,\dots,N\,.$$

This system describes the fundamental processes represented by

$$w_{ij}+w_{jk}\Leftrightarrow w_{ik}\,.$$

For simplicity, we restrict ourselves to considering the case in which the bound states occur in all fields. For the group velocities v_{ij}, we have

$$(v_{jk}-v_{ik})(v_{ik}-v_{ij})>0,\quad i<j<k\,.$$

The corresponding equations of motion for the system can be cast into the Lax form (see Appendix A for the details). The auxiliary linear problem in QISM can be put in the form

$$\frac{\partial}{\partial x}T(x,x_0|\lambda)=:L(x,\lambda)T(x,x_0|\lambda):,\quad T(x_0,x_0|\lambda)=I\,, \quad (3)$$

with

$$L(x,\lambda)=i\lambda\sum_{l}a_l e_{ll}$$
$$+\sqrt{c}\sum_{l<m}\sqrt{a_l-a_m}\,(w_{lm}e_{lm}+w_{lm}^{+}e_{ml}),$$
$$a_1>a_2>\dots>a_N,\quad c<0\,,$$

where λ is the spectral parameter, and $w_{ij}(x)=w_{ij}(x,t=0)$. Here e_{lm} is an $N\times N$ matrix, with elements

 0022-2488/89/030668-05$02.50

being zero except that the element of lth row and mth column is equal to unity. In terms of the parameters from (3), the group velocities v_{ij} and the coupling constants ϵ_{ijk} in (1) are given by

$$v_{ij} = \frac{b_i - b_j}{a_i - a_j}, \quad \epsilon_{ijk} = \frac{(v_{ij} - v_{jk}) l_{ij} l_{jk} l_{ik}}{ic\beta_{ik}}. \tag{4}$$

In order to obtain the commutation relations for the scattering data operators, we rewrite Eq. (3) in a lattice form

$$T_{j+1}(\lambda) = :L_j(\lambda) T_j(\lambda):, \quad T_j(\lambda) \equiv T(x_j, x_0|\lambda), \tag{5}$$

where

$$L_j(\lambda) = \sum_l (1 + i\lambda\Delta a_l) e_{ll} + \sqrt{c} \sum_{l<m} \sqrt{a_l - a_m} (w_{lmj} e_{lm} + w^+_{lmj} e_{ml}).$$

Here Δ is the small lattice spacing and $w_{lmj} = w_{lm}(x_j)$. In Eq. (5), we have neglected the terms of order Δ^2. Thus the Yang–Baxter relations are given by

$$R(\lambda - \mu)(L_j(\lambda) \otimes L_j(\mu)) = (L_j(\mu) \otimes L_j(\lambda)) R(\lambda - \mu). \tag{6}$$

It is easy to check that the R matrix has the form

$$R(\lambda - \mu) = \frac{-ic}{\lambda - \mu - ic} \sum_{lm} e_{ll} \otimes e_{mm} + \frac{\lambda - \mu}{\lambda - \mu - ic} \sum_{lm} e_{lm} \otimes e_{ml}. \tag{7}$$

We note that the Hilbert space of quantum states of the system under study is the tensor product of $N(N-1)/2$ Fock spaces for boson fields,

$$H = \underset{j<k}{\otimes} H_{jk}, \quad j,k = 1,2,\dots,N,$$

with the pseudovacuum being defined by $w_{jk}(x)|0\rangle = 0$, $j<k$, $j,k = 1,2,\dots,N$. The expectation value of $L_j(\lambda) \otimes L_j(\mu)$ between the pseudovacuum is

$$W(\lambda,\mu) = \sum_{lm} (1 + i\lambda\Delta a_l + i\mu\Delta a_m) e_{ll} \otimes e_{mm} + \Delta c \sum_{l<m} (a_l - a_m) e_{lm} \otimes e_{ml}. \tag{8}$$

To make it possible to take the continuum limit, let us introduce a normalized monodromy matrix $T(\lambda)$ defined by

$$T(\lambda) = \lim_{N\to\infty} V^{-N}(\lambda) L_N(\lambda) \cdots L_{-N+1}(\lambda) V^{-N}(\lambda), \tag{9}$$

where

$$V(\lambda) = \sum_l (1 + i\lambda\Delta a_l) e_{ll}. \tag{10}$$

Then, the Yang–Baxter relations for $T(\lambda)$ can be written as

$$R_+(\lambda - \mu)(T(\lambda) \otimes T(\mu)) = (T(\mu) \otimes T(\lambda)) R_-(\lambda - \mu), \tag{11}$$

where

$$\begin{aligned} R_+(\lambda - \mu) &= U_+^{-1}(\mu,\lambda) \frac{R(\lambda - \mu)}{\lambda - \mu} U_+(\lambda,\mu), \\ R_-(\lambda - \mu) &= U_-(\mu,\lambda) \frac{R(\lambda - \mu)}{\lambda - \mu} U_-^{-1}(\lambda,\mu), \end{aligned} \tag{12}$$

with

$$\begin{aligned} U_+(\lambda,\mu) &= \lim_{N\to\infty} W^{-N}(\lambda,\mu)(V^N(\lambda) \otimes V^N(\mu)), \\ U_-(\lambda,\mu) &= \lim_{N\to\infty} (V^N(\lambda) \otimes V^N(\mu)) W^{-N}(\lambda,\mu). \end{aligned} \tag{13}$$

After a tedious but straightforward calculation, we have

$$\begin{aligned} R_\pm(\lambda - \mu) = {} & \frac{1}{\lambda - \mu} \sum_l e_{ll} \otimes e_{ll} \\ & \pm i\pi\delta(\lambda - \mu) \sum_{lm} \eta_{lm} e_{ll} \otimes e_{mm} \\ & + \frac{\lambda - \mu + ic}{(\lambda - \mu + i\epsilon)^2} \sum_{l<m} e_{lm} \otimes e_{ml} \\ & + \frac{1}{\lambda - \mu - ic} \sum_{l>m} e_{lm} \otimes e_{ml}, \end{aligned} \tag{14}$$

$$\eta_{lm} = \begin{cases} 1, & l<m, \\ 0, & l=m, \\ -1, & l>m. \end{cases}$$

Here the δ function appears as a result of the formula

$$\lim_{N\to\infty} \exp(ixN)/(x - i\epsilon) = 2i\pi\delta(x).$$

From this we obtain the commutation relations for the scattering data operators. In doing so, we use the notation

$$T(\lambda) = \sum_l A_l e_{ll} + \sum_{l<m} (B_{lm} e_{lm} + C_{lm} e_{ml}). \tag{15}$$

The explicit results are given in Appendix B.

Let us now discuss the scattering states for the model. From the Neumann series for A_l and B_{lm}, we have

$$A_l|0\rangle = |0\rangle, \quad B_{lm}|0\rangle = 0, \quad l<m, \quad l,m = 1,2,\dots,N. \tag{16}$$

The commutation relations between A_l and C_{mn} show that the state

$$\prod_{l<m} \prod_{\alpha=1}^{n_{lm}} C_{lm}(\lambda_\alpha^{(lm)})|0\rangle, \quad l,m = 1,2,\dots,N, \tag{17}$$

is an eigenstate of A_l. The corresponding eigenvalues a_l are given by

$$\begin{aligned} a_l(\lambda) = {} & \prod_{i=1}^{l-1} \prod_{\alpha=1}^{n_{il}} \frac{\lambda - \lambda_\alpha^{(il)} + ic}{\lambda - \lambda_\alpha^{(il)}} \\ & \times \prod_{i<l<j} \prod_{\alpha=1}^{n_{ij}} \frac{(\lambda - \lambda_\alpha^{(ij)} + ic)(\lambda - \lambda_\alpha^{(ij)} - ic)}{(\lambda - \lambda_\alpha^{(ij)})^2} \\ & \times \prod_{i=l+1}^{N} \prod_{\alpha=1}^{n_{li}} \frac{\lambda - \lambda_\alpha^{(li)} - ic}{\lambda - \lambda_\alpha^{(li)}}, \quad l = 1,2,\dots,N. \end{aligned} \tag{18}$$

On the other hand, $A_l(\lambda)$ are the generating functionals for an infinite number of conservation laws in the model. They have the following asymptotic expansions:

$$A_l(\lambda) = 1 + \tau_l/\lambda + \mu_l/\lambda^2 + \cdots, \quad \tau_l = icN_l,$$

$$\mu_l = ic\left(\sum_{i=1}^{l-1} \frac{P_{il}}{\beta_{il}} - \sum_{i=l+1}^{N} \frac{P_{li}}{\beta_{li}}\right) - \frac{c^2}{2}(N_l^2 + N_l) - c^2 \sum_{i=1}^{l-1} N_i$$

$$- \sum_{l<i<j} \frac{l_{li} l_{ij} l_{lj}}{\beta_{li}\beta_{lj}} V_{lij} + \sum_{i<l<j} \frac{l_{il} l_{lj} l_{ij}}{\beta_{il}\beta_{lj}} V_{ilj}$$
$$- \sum_{i<j<l} \frac{l_{ij} l_{jl} l_{il}}{\beta_{jl}\beta_{il}} V_{ijl} , \tag{19}$$

where

$$N_l = \int \left(\sum_{i=1}^{l-1} w_{il}^+ w_{il} - \sum_{i=l+1}^{N} w_{li}^+ w_{li} \right) dx,$$
$$P_{ij} = \int \frac{w_{ij}^+}{i} \frac{\partial w_{ij}}{\partial x} dx,$$
$$V_{ijk} = \int (w_{ij} w_{jk} w_{ik}^+ + w_{ik} w_{jk}^+ w_{ij}^+) dx ,$$

as λ is large. From this we can reconstruct the well-known conserved quantities, i.e., charges, momentum, and Hamiltonian:

$$N_l = \frac{\tau_l}{ic}, \quad l = 1,2,...N,$$
$$P = \frac{1}{ic} \sum_{l=2}^{N} \beta_{1l} \left(\mu_l - \frac{1}{2}\tau_l^2 - \frac{ic}{2}\tau_l - ic \sum_{m=1}^{l-1} \tau_m \right),$$
$$H = \frac{1}{ic} \sum_{l=2}^{N} v_{1l}\beta_{1l} \left(\mu_l - \frac{1}{2}\tau_l^2 - \frac{ic}{2}\tau_l - ic \sum_{m=1}^{l-1} \tau_m \right). \tag{20}$$

Expanding the eigenvalues for $A_l(\lambda)$ in terms of the inverse powers of λ, we have

$$a_l(\lambda) = 1 + \frac{\alpha_l}{\lambda} + \frac{\beta_l}{\lambda^2} + \cdots,$$
$$\alpha_l = ic \left(\sum_{i=1}^{l-1} n_{il} - \sum_{i=l+1}^{N} n_{li} \right),$$
$$\beta_l = ic \left(\sum_{i=1}^{l-1} \sum_{\alpha=1}^{n_{il}} \lambda_\alpha^{(il)} - \sum_{i=l+1}^{N} \sum_{\alpha=1}^{n_{li}} \lambda_\alpha^{(li)} \right)$$
$$+ \frac{ic}{2} \left(\alpha_l + 2 \sum_{i=1}^{l-1} \alpha_i \right) + \frac{\alpha_l^2}{2}, \quad l = 1,2,...,N. \tag{21}$$

Comparing with Eqs. (19) and (20), we get

$$N_l = \sum_{i=1}^{l-1} n_{il} - \sum_{i=l+1}^{N} n_{li}, \quad l = 1,2,..,N,$$
$$P = \sum_{i<j} \beta_{ij} \sum_{\alpha=1}^{n_{ij}} \lambda_\alpha^{(ij)}, \quad H = \sum_{i<j} v_{ij}\beta_{ij} \sum_{\alpha=1}^{n_{ij}} \lambda_\alpha^{(ij)} . \tag{22}$$

This implies that Eq. (17) is the eigenstate for an infinite number of conservation laws in the model.

As is well known, a classical soliton is a manifestation of the quantum bound state. This fact has been rigorously shown by Wadati and Sakagami[13] in the nonlinear Schrödinger model. This hints that the existence of a classical soliton should be intimately related to the condition of the bound states in the quantum variant of the system. In our case, $N(N-1)/2$ sets of operators,

$$A_l(\lambda), \quad A_m(\lambda), \quad B_{lm}(\lambda), \quad \text{and} \quad C_{lm}(\lambda) ,$$

satisfy the same commutation relations as the scattering data operators of the nonlinear Schrödinger model. As was pointed out by Kulish,[9] for the three-wave interaction model, when reducing to the subspace $H_{l,l+1}$ $(l = 1,2,...,N-1)$, the corresponding set gives just the nonlinear Schrödinger model, respectively. However, this is not true for other fields. This result can also be seen from the eigenvalues for A_l $(l = 1,2,...,N)$. Using the results for nonlinear Schrödinger model with attraction, we conclude that the spectral parameters of the many-particle bound state form a string given by

$$\lambda_\alpha^{(l,l+1)} = \lambda^{(l,l+1)} - ic(m^{(l,l+1)} + 1 - 2\alpha)/2,$$
$$\alpha = 1,2,...,m^{(l,l+1)}, \quad \mathrm{Im}\,\lambda^{(l,l+1)} = 0 . \tag{23}$$

The bound states involving other fields can be constructed by using the "contracting" technique proposed by Wadati and Ohkuma.[12] For example, contracting a pair, say $\lambda_\alpha^{(l,l+1)}$ and $\lambda_\beta^{(l+1,l+2)}$, out of the strings $\{\lambda_\alpha^{(l,l+1)}\}_{\alpha=1}^{m^{(l,l+1)}}$ and $\{\lambda_\alpha^{(l+1,l+2)}\}_{\alpha=1}^{m^{(l+1,l+2)}}$, we obtain a bound state formed by a connected string

$$\{\cdots\lambda_{\alpha+1}^{(l,l+1)}\lambda_{\alpha-1}^{(l,l+1)}\cdots,$$
$$\cdots\lambda_{\beta+1}^{(l+1,l+2)}\lambda_{\beta-1}^{(l+1,l+2)}\cdots\lambda_\alpha^{(l,l+1)} + \lambda_\beta^{(l+1,l+2)}\}$$

involving the $w_{l,l+2}$ field.

This procedure may be repeated to have a series of bound states that include a $w_{l,l+2}$ field as well as $w_{l,l+1}$ and $w_{l+1,l+2}$ fields. Further, the bound states involving a $w_{l,l+3}$ field can be constructed by contracting the strings formed by $w_{l,l+1}$, $w_{l+1,l+2}$, and $w_{l+2,l+3}$ fields, and so on. It is easy to check that the binding energy of the bound state is zero. This must be a consequence of the fact that the system is linear dispersive.

For completeness, let us here come back to the classical theory. According to Izergin and Korepin,[14] the quasiclassical limit is given by

$$R \underset{\hbar\to 0}{\to} \sum_{lm} e_{lm} \otimes e_{ml} \left(\sum_{lm} e_{ll} \otimes e_{mm} - i\hbar r \right). \tag{24}$$

Here, r is the classical r matrix, $r = [-ic/(\lambda-\mu)] \Sigma_{lm} e_{lm} \otimes e_{ml}$. The corresponding classical Yang–Baxter relations for the normalized monodromy matrix become

$$\{T(\lambda) \otimes T(\mu)\} = r_+(\lambda-\mu) T(\lambda) \otimes T(\mu)$$
$$- T(\lambda) \otimes T(\mu) r_-(\lambda-\mu) , \tag{25}$$

with

$$r_\pm(\lambda-\mu) = -\frac{c}{\lambda-\mu} \sum_l e_{ll} \otimes e_{ll}$$
$$\pm ic\pi\delta(\lambda-\mu) \sum_{l<m} (e_{lm} \otimes e_{ml}$$
$$- e_{ml} \otimes e_{lm}).$$

Here the Poisson brackets of the tensor product of two matrices A and B are defined by

$$\{A \otimes B\}_{kl,mn} = \{A_{km}, B_{ln}\}. \tag{26}$$

The classical commutation relations for the scattering data are given in Appendix C. It is easy to see that the correspondence between quantum and classical commutation relations follows the usual principle:

$$\{a,b\} = i[a,b]. \tag{27}$$

Using a generalization of Faddeev's method for the nonlinear Schrödinger model, we can establish the correspondence between classical solitons and quantum bound states: the conditions for classical soliton and quantum bound state are the same.

III. CONCLUSION

To conclude let us turn to discuss the integrability of the same model but with a different choice of statistics. If we regard $\{w_{1l},...w_{l-1,l}, w_{l,l+1},...,w_{lN}\}$ as fermion fields rather than boson fields for fixed l ($l = 1,2,...,N$), then the system remains completely integrable. It should be noted that the exchange between w_{lm} and $w_{N+1-m,N+1-l}$ does not affect the physics of the system. Thus we obtain $N/2$ $(N+1/2)$ new completely integrable systems for even (odd) N, respectively. For these systems, the Lax pair operators are supermatrices. The Yang–Baxter relations should be understood in the graded sense[15]:

$$\mathscr{R}(\lambda-\mu)\mathscr{L}_i(\lambda)\underset{s}{\otimes}\mathscr{L}_j(\mu)$$
$$=\mathscr{L}_j(\mu)\underset{s}{\otimes}\mathscr{L}_i(\lambda)\mathscr{R}(\lambda-\mu). \tag{28}$$

Here by $\underset{s}{\otimes}$ we mean the Grassmann direct product defined by

$$(A\underset{s}{\otimes}B)_{kl,mn}=(-1)^{[P(k)+P(m)]P(l)}A_{km}B_{ln}.$$

To transform the Lax pair operator $\mathscr{L}_j^{(\alpha)}(\lambda)$ into the standard supermatrix form, we now introduce a similarity transformation

$$\mathscr{L}_j^{(\alpha)\prime}(\lambda)=S_\alpha\mathscr{L}_j^{(\alpha)}(\lambda)S_\alpha^{-1}, \tag{29}$$

with

$$S_\alpha=\sum_l e_{ll}-e_{N+1-\alpha,N+1-\alpha}-e_{NN}$$
$$+e_{N+1-\alpha,N}+e_{N,N+1-\alpha},$$
$$\alpha=\begin{cases}1,2,...,N/2, & \text{for even } N,\\ 1,2,...,(N+1)/2, & \text{for odd } N,\end{cases}$$

for a fixed α, respectively. Then, the R matrix satisfying the Yang–Baxter relations (28) for the transformed Lax pair operators reads

$$\mathscr{R}(\lambda-\mu)=\frac{-ic}{\lambda-\mu-ic}\sum_{lm}e_{ll}\otimes e_{mm}$$
$$+\frac{\lambda-\mu}{\lambda-\mu-ic}\sum_{lm}(-1)^{P(l)P(m)}e_{lm}\otimes e_{ml}, \tag{30}$$

with $P(l)=0$, $l=1,2,...,N-1$, and $P(N)=1$. Further details will be considered elsewhere.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor B. L. Zhang for his interest in this work. Thanks are also due to Dr. Xiao-Yu Kuang, Dr. Jian-Gang Tang, and Dr. Zhuan Xiong for their help and useful advice.

APPENDIX A: DERIVATION OF THE LAX PAIR FOR THE SYSTEM (1)

The traditional basis for applying the inverse scattering method is to represent the equations of motion for the system in the Lax form:

$$\frac{\partial}{\partial t}T(x,y|\lambda)=M(x,\lambda)T(x,y|\lambda),$$
$$\frac{\partial}{\partial x}T(x,y|\lambda)=L(x,\lambda)T(x,y|\lambda). \tag{A1}$$

The compatibility condition should be consistent with the equations of motion:

$$L_t-M_x+[L,M]=0, \tag{A2}$$

where L and M are $N\times N$ matrices depending on the spectral parameter λ and the dynamical variables. In our case, we assume

$$\begin{aligned}L(x,\lambda)&=i\lambda\sum_l a_le_{ll}+\sum_{lm}p_{lm}(x)e_{lm},\\ M(x,\lambda)&=-i\lambda\sum_{lm}b_{lm}e_{lm}+\sum_{lm}q_{lm}(x)e_{lm}.\end{aligned} \tag{A3}$$

Substituting (A3) into (A2), we have

$$(a_lb_{lm}-b_{lm}a_m)\lambda^2$$
$$+\left[a_lq_{lm}-q_{lm}a_m+\sum_{n=1}^N(b_{ln}p_{nm}-p_{ln}b_{nm})\right]i\lambda$$
$$+\frac{\partial p_{lm}}{\partial t}-\frac{\partial q_{lm}}{\partial x}+\sum_{n=1}^N(p_{ln}q_{nm}-q_{ln}p_{nm})=0.$$

Obviously, the coefficients in the same powers of λ in the above equation must be zero. Thus

$$\begin{aligned}&a_lb_{lm}-b_{lm}a_m=0,\\ &a_lq_{lm}-q_{lm}a_m+\sum_{n=1}^N(b_{ln}p_{nm}-p_{ln}b_{nm})=0,\end{aligned} \tag{A4}$$

$$\frac{\partial p_{lm}}{\partial t}-\frac{\partial q_{lm}}{\partial x}+\sum_{n=1}^N(p_{ln}q_{nm}-q_{ln}p_{nm})=0. \tag{A5}$$

From this we conclude

$$b_{lm}=b_l\delta_{lm},\quad q_{lm}=-v_{lm}p_{lm} \tag{A6}$$

$$\frac{\partial p_{lm}}{\partial t}+v_{lm}\frac{\partial p_{lm}}{\partial x}=-\sum_{n=1}^N(v_{ln}-v_{nm})p_{ln}p_{nm}. \tag{A7}$$

Here we have used the notation $v_{lm}=(b_l-b_m)/(a_l-a_m)$. Now choosing

$$p_{lm}=\begin{cases}l_{lm}w_{lm}, & l<m,\\ 0, & l=m,\\ l_{ml}w_{ml}^+, & l>m,\end{cases} \tag{A8}$$

where

$$l_{lm}=\sqrt{c\beta_{lm}},\quad \beta_{lm}=a_l-a_m,\quad l<m,\quad l,m=1,2,...,N,$$

and substituting (A8) into (A6) and (A7), we have

$$q_{lm}=\begin{cases}-v_{lm}l_{lm}w_{lm}, & l<m,\\ 0, & l=m,\\ -v_{lm}l_{ml}w_{ml}^+, & l>m,\end{cases} \tag{A9}$$

and

$$\frac{\partial w_{lm}}{\partial t} + v_{lm}\frac{\partial w_{lm}}{\partial x}$$
$$= -\sum_{n=1}^{l-1}\frac{(v_{nl}-v_{nm})l_{nl}l_{nm}}{l_{lm}} w_{nl}^{+} w_{nm} - \sum_{n=l+1}^{m-1}\frac{(v_{ln}-v_{nm})l_{ln}L_{nm}}{l_{lm}} w_{ln} w_{nm} - \sum_{n=m+1}^{N}\frac{(v_{ln}-v_{mn})l_{ln}l_{mn}}{l_{lm}} w_{ln} w_{mn}^{+}. \quad \text{(A10)}$$

Comparing the above with the equations of motion, we immediately obtain the relations given in Eq. (4). It is worthwhile to note that for real ϵ_{klm}, l_{lm} must be pure imaginary. If so, the constant c must be negative definite.

APPENDIX B: THE COMMUTATION RELATIONS FOR THE SCATTERING DATA OPERATORS

The commutation relations are

$$[A_l(\lambda),A_m(\mu)] = 0, \quad l,m = 1,2,\ldots,N,$$
$$[B_{lm}(\lambda),B_{lm}(\mu)] = [C_{lm}(\lambda),C_{lm}(\mu)] = 0,$$
$$B_{lm}(\lambda)A_l(\mu) = [(\lambda-\mu+ic)/(\lambda-\mu+i\epsilon)]A_l(\mu)B_{lm}(\lambda),$$
$$B_{lm}(\lambda)A_m(\mu) = [(\lambda-\mu-ic)/(\lambda-\mu-i\epsilon)]A_m(\mu)B_{lm}(\lambda),$$
$$A_l(\lambda)C_{lm}(\mu) = [(\lambda-\mu-ic)/(\lambda-\mu-i\epsilon)]C_{lm}(\mu)A_l(\lambda),$$
$$A_m(\lambda)C_{lm}(\mu) = [(\lambda-\mu+ic)/(\lambda-\mu+i\epsilon)] \times C_{lm}(\mu)A_m(\lambda), \quad l<m,$$
$$[B_{lm}(\lambda),A_n(\mu)] = [B_{mn}(\lambda),A_l(\mu)] = 0,$$
$$[A_n(\lambda),C_{lm}(\mu)] = [A_l(\lambda),C_{mn}(\mu)] = 0,$$
$$B_{ln}(\lambda)A_m(\mu) = [(\lambda-\mu+ic)(\lambda-\mu-ic)/(\lambda-\mu+i\epsilon)^2] \times A_m(\mu)B_{ln}(\lambda) + 2c\pi\delta(\lambda-\mu) \times B_{lm}(\lambda)B_{mn}(\lambda),$$
$$A_m(\lambda)C_{ln}(\mu) = [(\lambda-\mu+ic)(\lambda-\mu-ic)/(\lambda-\mu+i\epsilon)^2] \times C_{ln}(\mu)A_m(\lambda) + 2c\pi\delta(\lambda-\mu) \times C_{lm}(\lambda)C_{mn}(\lambda), \quad l<m<n.$$

APPENDIX C: THE POISSON BRACKETS FOR THE CLASSICAL SCATTERING DATA

The Poisson brackets are

$$\{A_l(\lambda),A_m(\mu)\} = 0, \quad l,m = 1,2,\ldots,N,$$
$$\{B_{lm}(\lambda),B_{lm}(\mu)\} = \{C_{lm}(\lambda),C_{lm}(\mu)\} = 0,$$
$$\{B_{lm}(\lambda),A_l(\mu)\} = [-c/(\lambda-\mu+i\epsilon)]B_{lm}(\lambda)A_l(\mu),$$
$$\{B_{lm}(\lambda),A_m(\mu)\} = [c/(\lambda-\mu-i\epsilon)]B_{lm}(\lambda)A_m(\mu),$$
$$\{A_l(\lambda),C_{lm}(\mu)\} = [c/(\lambda-\mu-i\epsilon)]A_l(\lambda)C_{lm}(\mu),$$
$$\{A_m(\lambda),C_{lm}(\mu)\} = [-c/(\lambda-\mu+i\epsilon)]A_m(\lambda)C_{lm}(\mu), \quad l<m,$$
$$\{B_{lm}(\lambda),A_n(\mu)\} = \{B_{mn}(\lambda),A_l(\mu)\} = 0,$$
$$\{A_n(\lambda),C_{lm}(\mu)\} = \{A_l(\lambda),C_{mn}(\mu)\} = 0,$$
$$\{B_{ln}(\lambda),A_m(\mu)\} = 2ic\pi\delta(\lambda-\mu)B_{lm}(\lambda)B_{mn}(\mu),$$
$$\{A_m(\lambda),C_{ln}(\mu)\} = 2ic\pi\delta(\lambda-\mu)C_{lm}(\lambda)C_{mn}(\lambda), \quad l<m<n.$$

[1]C. S. Gardner, G. M. Green, M. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[2]V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. **8**, 226 (1974).
[3]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973).
[4]V. E. Zakharov, S. V. Manakov, S. P. Novikov, and L. P. Pitaevskii, *Teoriya Solitonov* (Nauka, Moscow, 1980).
[5]H. B. Thacker, Rev. Mod. Phys. **53**, 253 (1982); L. D. Faddeev, Sov. Sci. Rev. Math. Phys. C **1**, 107 (1980); P. P. Kulish and E. K. Sklyanin, *Lecture Notes in Physics*, Vol. 151, edited by J. Hieterinta and C. Montonen (Springer, Berlin, 1982).
[6]R. J. Baxter, Ann. Phys. (NY) **70**, 323 (1972).
[7]H. Bethe, Z. Phys. **71**, 105 (1931).
[8]M. Lüscher, Nucl. Phys. B **117**, 475 (1976); H. Q. Zhou and J. G. Tang, Phys. Rev. B (to be published); H. Q. Zhou, Z. Xiong, and L. J. Jiang, J. Phys. A **21**, 3385 (1988).
[9]P. P. Kulish, Physica **18**, 360 (1986).
[10]D. J. Kaup, A. Reiman, and A. Bers, Rev. Mod. Phys. **51**, 275 (1979); V. E. Zakharov and S. V. Manakov, JETP Lett. **18**, 243 (1973).
[11]Y. P. Wang and F. C. Pu, J. Math. Phys. **29**, 400 (1988).
[12]K. Ohkuma and M. Wadati, J. Phys. Soc. Jpn. **53**, 2899 (1984).
[13]M. Wadati and M. Sakagami, J. Phys. Soc. Jpn. **53**, 1933 (1984).
[14]A. G. Izergin and V. E. Korepin, Sov. J. Part. Nucl. **13**, 207 (1982).
[15]A. G. Izergin and P. P. Kulish, Teor. Mat. Fiz. **44**, 189 (1980).

Phase shifts in the collision of massive particles

B. G. Sidharth
B. M. Birla Planetarium, Adarsh Nagar, Hyderabad-500 463, India

(Received 13 August 1985; accepted for publication 26 October 1988)

In this paper phase shifts for inverse power potentials and their superpositions are considered. These potentials govern a large number of collision processes involving massive particles. But they are problematic as they do not conform to the norms obeyed by standard potentials: The usual rules for estimating the error of the phase shifts break down and an abnormally large number of phase shifts have to be computed, even if they are small. A sufficiency condition is deduced for the applicability of the Born approximation which shows that, for a given potential strength and l, surprisingly, the Born approximation is good at low energies and bad at high energies. The magnitude and direction of the error committed are also estimated. These conclusions are then verified in a number of special cases, for instance, the inverse fourth power potential used in the scattering of electrons by atoms and the Lennard–Jones potential used in the scattering of beams of molecular hydrogen with mercury atoms.

I. INTRODUCTION

It is well known that inverse power potentials govern the collisions of massive particles (molecular beams, for instance) and other important processes. As another example, the potential A/r^4 features in the scattering of electrons by atoms, while a $1/r^6$ term gives the Van der Waals interaction that appears in potentials of the Lennard–Jones type, viz., $A_1/r^{s_1} + A_2/r^{s_2}$, which govern, among other things, processes involving noble atoms and which are also relevant in nuclear physics in the context of ion collisions. These potentials have been studied by several authors and used in a variety of physical problems.[1–7]

Unfortunately, because of the singularity at the origin, such potentials stand apart and are problematic[6]; they do not conform to the norms obeyed by standard potentials. For instance, the usual criterion for the applicability of the Born approximation breaks down as we will see. Also the usual rules for estimating the error of the calculated phase shifts[8] do not apply. And, finally, even if the phase shifts are small, an abnormally large number of them—of the order of a few hundred[9]—have to be computed, owing to the slow convergence of the scattering amplitude series.

For such potentials, we will, among other things, exploit the symmetry of the radial Schrödinger equation under scale transformations to deduce not only a sufficient condition for the validity of the Born approximation, but also an estimate of the magnitude and sign of the error involved. Thus the Born approximation can indeed by used to compute the large number of small phase shifts required. These deductions will then be verified by comparison with some known results.

II. THE PHASE SHIFTS

Our starting point is the radial Schrödinger equation[10]

$$u'' + [K^2 - l(l+1)/r^2 - \lambda U(r)]u = 0,$$
$$u(0) = 0, \quad u(r) \to (1/K)\sin(Kr - l\pi/2 + \delta_l), \quad r \to \infty, \tag{1}$$

where λ is a parameter, and where primes denote differentiation with respect to r.

In integral form, (1) is given by[10]

$$u(r) = (r\cos\delta_l)j_l(Kr) + \lambda \int_0^\infty G(r,r')U(r')u(r')dr', \tag{2}$$

where

$$G(r,r') = \begin{cases} (1/K)\cdot(Kr)j_l(Kr)\cdot(Kr')n_l(Kr'), & r \leqslant r', \\ (d/K)\cdot(Kr')j_l(Kr')\cdot(Kr)n_l(Kr), & r \geqslant r', \end{cases}$$

while the phase shifts δ_l are given by

$$\sin\delta_l = -\lambda K \int_0^\infty rj_l(Kr)u(r)U(r)dr, \tag{3}$$

where $j_l(Kr)$ and $n_l(Kr)$ are the spherical Bessel and Neumann functions.

Further, the kernel $G(r,r')$ has the bound[11]

$$|G(r,r')/(rr')^{1/2}| = 1/t_l. \tag{4}$$

In the first instance, we consider Eq. (1) for the potential

$$U(r) = Ar^{-s} \tag{5}$$

with the following observations. For convergence at ∞ of the integrals in (2) and (3), we require that $s > 1$. If $s > 2$, the potential is singular. In this case A must necessarily be positive, that is, the potential must be repulsive. If $A < 0$, the scattering problem is indeterminate as both linearly independent solutions of (1) vanish at the origin[12] (cf. also Ref. 1). In fact, the scattering problem is meaningless because the spectrum of the operator would be unbounded below and bound states with arbitrarily large negative energy would exist.

Specific forms for the wave function near the origin have been considered for different types of singular repulsive potentials by various authors.[13] However, the general small r behavior of the wave function for such potentials is given by the following lemma.

Lemma: If

$$r^s U(r) \to 0+ \quad \text{as} \quad r \to 0, \quad \text{for some } s > 2,$$

 0022-2488/89/030673-05$02.50

then, for the solution of (1) that vanishes at the origin,

$$u(r)/r^M \to 0 \quad \text{as } r\to 0 \tag{6}$$

for arbitrarily large M.

Proof: First we observe that in $(0,E)$, where E can be chosen arbitrarily small, $u(r)$ can have no zero other than at $r=0$. One way of seeing this is by comparing (1) with the equation

$$V'' - [M(M-1)/r^2]V = 0, \tag{7}$$

where M can be any number greater than 1.

By Sturm's Lemma[14] it follows that between two successive zeros of $u(r)$ in $(0,E)$, that is, in the open interval $(0,E)$, there must lie at least one zero of a function V satisfying (7). Choosing $V=r^M$, this is seen to be impossible.

Without loss of generality, we can choose $u(r)>0$ in $0<r<E$. (Physically speaking the repulsive singular potential pushes out the wave function.)

We further observe that, as $u(r)\to 0$ when $r\to 0$, it also follows that

$$ru'(r)\to 0 \quad (r\to 0), \tag{8}$$

since, by the mean value theorem,

$$u(r) = u(0) + ru'(\theta r), \quad 0<\theta<1,$$

or

$$\theta u(r) = \theta r u'(\theta r) \to 0 \quad \text{as } \theta r \to 0.$$

Next, from (1) and (7), we obtain

$$vu' - uv' = vu' - uv'|_0^r = \int_0^r \left[Ar^{-s} + \frac{l(l+1)}{r^2} - \frac{M(M-1)}{r^2} - K^2 \right] uv\, dr,$$

in view of (8). So for $r<E$, remembering that Ar^{-s} dominates the other terms in the square brackets on the right side, if $v>0$,

$$(u/v)' > 0,$$

that is, u/v is an increasing function. So u/v decreases as $r\to 0$ and, since it is greater than 0,

$$u/v \to b, \quad r\to 0, \quad \text{where } b \geqslant 0.$$

We choose $v=r^m$ and it follows that

$$u/r^{M-\epsilon} \to 0 \quad \text{as } r\to 0.$$

Remembering that M can be made arbitrarily large, the lemma follows. □

The lemma ensures the convergence at $r=0$ of the integrals in (2) and (3). Thus, for the potentials given by (5), with $A>0$ and $s>2$, the integrals in (2) and (3) are convergent at infinity and at $r=0$.

To proceed, we replace in (1), λ by λ' and K by K', to get

$$u'' + [K'^2 - l(l+1)/r^2 - \lambda' A r^{-s}]u = 0. \tag{9}$$

We next make the transformation $r=a\rho$ in (9). This gives

$$\frac{d^2u}{d\rho^2} + \left[a^2K'^2 - \frac{l(l+1)}{\rho^2} - \lambda' a^{2-s} A\rho^{-s} \right] u = 0. \tag{10}$$

Equation (10) is similar to (9) except that aK' and $\lambda' a^{2-s}$ replace K' and λ'. As u satisfies the same condition at $r=\rho=0$, viz., $u(0)=0$, it follows that

$$\delta_l(K';\lambda') = \delta_l(aK';\lambda' a^{2-s}). \tag{11}$$

Equation (11) represents in integral form the consequence of the symmetry of (9) under the transformation $r=a\rho$.

To obtain a condition for this symmetry in differential form, we differentiate both sides of (11) with respect to a to get

$$K' \frac{\partial \delta_l}{\partial (aK')} + \frac{\partial}{\partial a}(\lambda' a^{2-s}) \frac{\partial \delta_l}{\partial(\lambda' a^{2-s})} = 0,$$

which leads to

$$\lambda(s-2)\frac{\partial \delta_l}{\partial \lambda} - K\frac{\partial \delta_l}{\partial K} = 0, \tag{12}$$

where $K=aK'$ and $\lambda = \lambda' a^{2-s}$.

To put (12) in a simpler form, we make the substitution, $K^{s-2} \equiv \mu$. This gives

$$\lambda \frac{\partial \delta_l}{\partial \lambda} - \mu \frac{\partial \delta_l}{\partial \mu} = 0. \tag{13}$$

It is easy to verify that the general solution of (13) is

$$\delta_l = f(\lambda\mu) = f(\lambda K^{s-2}), \tag{14}$$

where f is an arbitrary function. In fact, (14) is an expression of the condition (11).

We would next like to expand the wave function $u(r)$ as a power series in λ. For this, we first rewrite (2) as

$$u(r) = (r\cos\delta_l) j_l(Kr) + \lambda \int_\epsilon^R G_l(r,r')U(r')u(r')dr' + \mu_1(\epsilon) + \nu_1(R),$$

where

$$\mu_1(\epsilon) \equiv \lambda \int_0^\epsilon G_l(r,r')U(r')u(r')dr',$$

$$\nu_1(R) \equiv \lambda \int_R^\infty G_l(r,r')U(r')u(r')dr'.$$

The function $\mu_1(\epsilon)$ can be made arbitrarily small and by suitably choosing ϵ in view of (4) and the small r behavior of $u(r)$ as given in (6). So also $\nu_1(R)$ can be made arbitrarily small for a suitable choice of R, because of the large r behavior of $u(r)$, as given in (1), the bound given in (4) and the fact that $U(r)$, as given by (5), falls off faster than $1/r$ as $r\to\infty$.

Similarly, Eq. (3) can be written as

$$\sin\delta_l = -\lambda K \int_\epsilon^R r j_l(Kr)u(r)U(r)dr + \mu_2(\epsilon) + \nu_2(R), \tag{15}$$

where, for the same reasons, $\mu_2(\epsilon)$ and $\nu_2(R)$ can be made arbitrarily small.

We also observe that $G_l(r,r')$, which is a combination of spherical Bessel and Neumann functions, is uniformly continuous in (r,r') in (ϵ,R). Thus the expansion

$$u(r) = u^{(0)}(r) + \lambda u^{(1)}(r) + \lambda^2 u^{(2)}(r) + \cdots + \lambda^n u^{(n)}(r) + \cdots \tag{16}$$

converges absolutely and uniformly with respect to r (see Ref. 15), and so

$$u(r) - u^{(0)}(r) \to 0 \quad \text{as } \lambda \to 0,$$

uniformly with respect to r, and term by term integration of the series (16) can be made when this expression is substituted in (15). Thus we get

$$\tan \delta_l(\lambda) = -K\lambda \int_\epsilon^R [rj_l(Kr)]^2 U(r)dr + O(\lambda^2) + \mu_2(\epsilon) + \nu_2(R),$$

where $\mu_2(\epsilon)$ and $\nu_2(R)$ can be made arbitrarily small independently of n in (16), by choosing ϵ suitably small and R sufficiently large. That is, we finally get

$$\tan \delta_l = -K\lambda B_l + O(\lambda^2), \tag{17}$$

where

$$B_l \equiv \int_0^\infty [rj_l(Kr)]^2 U(r)dr. \tag{18}$$

When the potential strength λ is small enough for λ^2 to be neglected, we recover from (17), $\bar{\delta}_l$, the Born approximation for the phase shift:

$$\delta_l \approx \bar{\delta}_l \equiv -K\lambda B_l, \tag{19}$$

where B_l is defined in (18).

In particular, if

$$U(r) = Ar^{-s},$$

it is known that (Ref. 1)

$$\bar{\delta}_l = \frac{A\pi}{2^s}(\lambda K^{s-2}) \frac{\Gamma(s-1)\Gamma(l-s/2+\frac{3}{2})}{[\Gamma(s/2)]^2\Gamma(l+s/2+\frac{1}{2})} \quad (2l>s-3). \tag{20}$$

Equations (19) and (20) show up one aspect of the irregular behavior of inverse power potentials: It is well known that the Born approximation (19) is valid, in general, for high energies, $K \gg 1$ and/or weak potentials, $\lambda \ll 1$. But (20) shows that for inverse power potentials with $s>2$, the Born approximation is particularly bad at high energies because $|\bar{\delta}_l| \to \infty$ as $K \to \infty$.

To obtain a sufficiency condition for the Born approximation for inverse power potentials, we observe that, as required by (14), $\bar{\delta}_l$ is indeed a function of λK^{s-2} and can be considered as the first term in the expansion of δ_l in powers of λK^{s-2}. In fact, combining (14), (19), and (20), we can write

$$\delta_l = \bar{\delta}_l + O([\lambda K^{s-2}]^2), \tag{21}$$

where $\bar{\delta}_l$ is given by (20). So the Born approximation for the phase shift is good if $|\lambda K^{s-2}| \ll 1$. So, when $s>2$, as in physical problems, the Born approximation is not good at high energies. Rather, for a given potential strength and any fixed l, it gives very good results at low energies! Happily, for these potentials, at high energies, the semiclassical approximation is good and has been successfully used (cf. Ref. 7).

To sum up, (20) and (21) show that for the potential λAr^{-s} and energy K,

$$\delta_l \equiv \delta_l(K;\lambda) = \frac{A\pi}{2^s} \frac{\Gamma(s-1)\Gamma(l-s/2+\frac{3}{2})}{[\Gamma(s/2)]^2\Gamma(l+s/2+\frac{1}{2})} \times (\lambda K^{s-2}) + O([\lambda K^{s-2}]^2)[2l>(s-3)]. \tag{22}$$

From (22), it is seen that when $[\lambda K^{s-2}]^2$ can be neglected, $\delta_l/\delta_{l+1} \to 1$ as $l \to \infty$, so that the phase shifts indeed fall off very slowly and the convergence of the scattering amplitude series is very slow as remarked earlier.

The generalization to potentials of the type

$$U(r) = U_1(r) + U_2(r) + \cdots + U_n(r) \equiv A_1 r^{-s_1} + \cdots + A_n r^{-s_n} \tag{23}$$

is immediate. As remarked earlier, near the origin, the repulsive part of $U(r)$ must dominate.

Exactly as before, via the transformation, $r = a\rho$, we deduce this time

$$\delta_l(K';\lambda_1';\lambda_2';\ldots;\lambda_n') = \delta_l(aK';\lambda_1' a^{2-s_1};\lambda_2' a^{2-S_2};\ldots;\lambda_n' a^{2-s_n}),$$

which leads to

$$K\frac{\partial \delta_l}{\partial K} - (s_1-2)\lambda_1 \frac{\partial \delta_l}{\partial \lambda_1} - \cdots - (s_n-2)\lambda_n \frac{\partial \delta_l}{\partial \lambda_n} = 0.$$

It is easily verified that the general solution is

$$\delta_l = f(\lambda_1 K^{s_1-2};\lambda_2 K^{s_2-2};\ldots;\lambda_n K^{s_n-2}), \tag{24}$$

where f is an arbitrary function of n independent variables.

We now use the fact that the Born approximations are additive, that is, if $\bar{\delta}_{l(i)}$ denotes the Born approximation for the phase shifts of the potential $U_i(r)$ defined in (23), then

$$\sum_i \bar{\delta}_{l(i)} = \bar{\delta}_l = -K \int_0^\infty [rj_l(Kr)]^2 \Big[\sum_i \lambda_i U_i(r)\Big] dr. \tag{25}$$

So, for potentials of the type (23), Eqs. (20), (24), and (25) yield

$$\delta_l \approx \sum_i A_i \frac{\pi}{2^{s_i}} \frac{\Gamma(s_i-1)\Gamma(l-s_i/2+\frac{3}{2})}{[\Gamma(s_i/2)]^2\Gamma(l+s/2+\frac{1}{2})} \times (\lambda_i K^{s_i-2}) + O([\lambda K^{s-2}]^2),$$
$$2l > \max(s_i-3), \quad \lambda K^{s-2} = \max(\lambda_i K^{s_i-2}). \tag{26}$$

III. ESTIMATE OF THE ERROR

For bounded potentials, or for those that behave like $1/r$ at the origin, the error can be estimated easily enough. An estimate for the magnitude of the error for potentials of the type (23) can also be deduced without difficulty. Equations (23) and (26) show that neglecting terms of the order of $(\lambda K^{s-2})^2$, the phase shift is of the order (λK^{s-2}). That is, the error is of the order of the square of the phase shift and so the percentage of error is given by

$$|\epsilon_l| \approx 100|\bar{\delta}_l|\% . \tag{27}$$

It is more difficult to determine the direction of the error. We develop the integral equation (2) as a perturbational series:

$$\begin{aligned}\frac{u(r)}{\cos\delta_l} = rj_l(Kr) + \lambda \Big[&Krn_l(Kr)B_l + Krj_l(Kr)\int_r^\infty \\ &\times r'^2 n_l(Kr')j_l(Kr')U(r')dr' \\ &- Krn_l(Kr)\int_r^\infty \{r'j_l(Kr')\}^2 U(r')dr' \Big] \\ &+ O(\lambda^2) .\end{aligned} \tag{28}$$

For simplicity, we first assume that, as $r\to\infty$, the dominant component of the potential (23), viz., $-A^2/r^s$, $s=\min(s_i)$, is attractive. For large r, invoking the asymptotic forms of the spherical Bessel functions and also using $U(r) = -A^2/r^s$, some manipulation with (28) yields

$$\frac{u(r)}{\cos\delta_l} \to rj_l(Kr) + [rn_l(Kr)] \times\left[-\bar{\delta}_l - \frac{A^2\lambda}{2K(s-1)}\cdot\frac{1}{r^{s-1}} + O\left(\frac{1}{r^s}\right)\right] \tag{29}$$

where δ_l is given by (19). We next compare (29) with the asymptotic form of $u(r)$ as given in (1), viz.,

$$\frac{u(r)}{\cos\delta_l} \to rj_l(Kr) + rn_l(Kr)[-\tan\delta_l] .$$

Remembering that for attractive potentials, $U(r)<0$, the δ_l are positive, and assuming that the sign of the phase shift is determined by the dominant component of the potential, as indeed is the case for the small phase shifts under consideration, we can see that when λK^{s-2} is small and

$$\tan\delta_l \approx \delta_l ,$$

the first Born approximation is less than the true phase shift δ_l.

If we include repulsive potentials with negative phase shifts, it is still true that

$$|\bar{\delta}_l| < \delta_l ;$$

that is, finally, using (27),

$$\delta_l - \bar{\delta}_l \sim \pm|\epsilon_l|\cdot|\delta_l|/100 = \pm|\delta_l|^2 , \tag{30}$$

depending on whether the phase shift is positive or negative.

IV. VERIFICATION

We will now verify Eqs. (14), (22), and (26) and other conclusions. We will start with the two cases where exact solutions are known.

(I) The repulsive inverse square potential, $U(r) = \lambda A^2/r^2$. It is well known that (cf. Refs. 1 and 10)

$$\delta_l = (\pi/2)\beta\lambda ,$$

where β is independent of λ and K, in agreement with Eq. (22).

(II) The inverse fourth power potential, $\lambda A/r^4$. For arbitrary $l>1$, and low energy, it is known that (cf. Ref. 2)

$$\tan\delta_l = \{\pi A/[(2l+3)(2l+1)(2^l-1)]\}\lambda K^2 + O(K^4) ,$$

again in agreement with (22).

When $l=0$, the condition on l as incorporated in (22) is violated. However, we can check with the master equation, (14).

$$\delta_0 = -\sqrt{A}\cdot\sqrt{\lambda K^2} ,$$

in agreement with (14).

(III) We next note that for large l, it can be deduced from (26), that the following recurrence relation holds:

$$(2l+1+s)\delta_{l+1} + 2(s-2)\delta_l = (2l+1-s)\delta_{l-1}, \quad s=\min(s_i), \quad l \text{ large} . \tag{31}$$

The relation (31) has been deduced independently.[17] However, it must be stressed that (26) is valid for any $l>\frac{1}{2}\max(s_i-3)$.

(IV) We next use the data for the scattering of hydrogen molecule beams by mercury atoms.[18] For this problem, the Lennard–Jones potential is used:

$$U(r) = (4B/\sigma^2)[(\sigma/r)^{12} - (\sigma/r)^6] ,$$

with $B=125$ and $\sigma = 2.91\times10^{-8}$ cm.

Introducing $A=K\sigma$, the sufficiency condition for the Born approximation, viz., $\lambda K^{s-2}\ll1$, reads $BA^4\ll1$, $BA^{10}\ll1$. In fact, this sufficiency condition is badly violated in the specific examples we will consider, where A is taken to be 3 and 5. Instead, we will work with the much weaker condition

$$|\delta_l \ll 1 ,$$

exploiting the fact that this is what the sufficiency condition implies. Tables I and II show that Eqs. (26) and (28) and a pocket calculator give figures that compare very favorably with calculations using the IBM 704 computer.

TABLE I. Phase shifts for a Lennard–Jones potential. ($B=125$, $A=3$.)

l	δ_l from (19)	From Ref. 14
11	0.062 (6.2)[a]	0.064 ($\pm$3.1)
12	0.040 (4)	0.041 ($\pm$4.9)
13	0.027 (2.7)	0.027 ($\pm$7.4)
14	0.019 (1.9)	0.019 ($\pm$10.5)
15	0.014 (1.4)	0.013 ($\pm$15.4)

[a] Figures in parentheses indicate corresponding percentage of error with sign.

TABLE II. Phase shifts for a Lennard–Jones potential. ($B = 125$, $A = 5$.)

l	δ_l from (19)	δ_l from Ref. 14
15	0.105 (10.5)[a]	0.110 (± 1.8)
16	0.077 (7.7)	0.079 (± 2.5)
17	0.06 (6)	0.06 (± 3.3)
18	0.04 (4)	0.05 (± 4)
19	0.03 (3)	0.04 (± 5)

[a] Figures in parentheses indicate corresponding percentage of error with sign.

V. CONCLUSIONS

For inverse power potentials or their superpositions, the usual criterion for the validity of the Born approximation does not hold. Instead, the Born approximation is valid for low energies, as shown by Eq. (26). The percentage of error involved is given by (27), while its direction is given by (28).

ACKNOWLEDGMENTS

The author is very grateful to Professor L. Fonda for useful discussions. He would also like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste. The author is also indebted to the referee, in the light of whose comments the paper has been rewritten.

[1] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Clarendon, Oxford, 1965), and references therein.
[2] B. H. Bransden, *Atomic Collision Theory* (Benjamin, New York, 1970), and references therein.
[3] M. R. C. McDowell and J. P. Coleman, *Introduction to the Theory of Ion Atom Collisions* (North-Holland, Amsterdam, 1970), and references therein.
[4] T. F. O'Malley, Phys. Rev. **130**, 1020 (1963); Phys. Rev. A **134**, 1188 (1964).
[5] Y. V. Martynenko and O. B. Firsov, Sov. Phys. JETP **17**, 154 (1963).
[6] B. R. Levy and J. B. Keller, J. Math. Phys. **14**, 54 (1963).
[7] K. W. Ford and J. A. Wheeler, Ann. Phys. (NY) 7, 259, 287 (1959).
[8] W. Kohn, Rev. Mod. Phys. **26**, 292 (1954).
[9] M. S. Child, "Scattering theory," Lecture Notes, Winter College on Lasers in Atomic and Molecular Physics, ICTP, Trieste, 1981.
[10] P. Roman, *Advanced Quantum Theory* (Addison–Wesley, Reading, MA, 1965).
[11] *Tables for Spherical Bessel Functions* (Columbia U.P., New York, 1947).
[12] W. M. Frank, D. J. Land, and R. M. Spector, Rev. Mod. Phys. **43**, 36 (1971).
[13] T. Regge and V. de Alfaro, *Potential Scattering* (North-Holland, Amsterdam, 1965).
[14] E. C. Titchmarsh, *Eigen Function Expansions* (Oxford U. P., London, 1962), Part I.
[15] I. G. Petrovsky, *Lectures on the Theory of Integral Equations* (Mir, Moscow, 1971).
[16] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3 (Pergamon, London, 1958).
[17] B. G. Sidharth, Nuovo Cimento A **46**, 419 (1978).
[18] R. B. Bernstein, J. Chem. Phys. **33**, 795 (1960).

A family of electrovac colliding wave solutions of Einstein's equations

Wei Li[a)]
Department of Mathematics and Computer Science, Clarkson University, Potsdam, New York 13676
Frederick J. Ernst
Department of Mathematics and Computer Science and Department of Physics, Clarkson University, Potsdam, New York 13676

(Received 1 July 1988; accepted for publication 9 November 1988)

Beginning with any colliding wave solution of the vacuum Einstein equations, a corresponding electrified colliding wave solution can be generated through the use of a transformation due to Harrison [J. Math. Phys. **9**, 1744 (1968)]. The method, long employed in the context of stationary axisymmetric fields, is equally applicable to colliding wave solutions. Here it is applied to a large family of vacuum metrics derived by applying a generalized Ehlers transformation to solutions published recently by Ernst, García, and Hauser (EGH) [J. Math. Phys. **28**, 2155, 2951 (1987); **29**, 681 (1988)]. Those EGH solutions were themselves a generalization of solutions first derived by Ferrari, Ibañez, and Bruni [Phys. Rev. D **36**, 1053 (1987)]. Among the electrovac solutions that are obtained is a charged version of the Nutku–Halil [Phys. Rev. Lett. **39**, 1379 (1977)] metric that possesses an *arbitrary* complex charge parameter.

I. INTRODUCTION

During recent years there has been a renewal of interest in exact solutions of the Einstein equations, which can be interpreted as describing the collision of gravitational plane waves. Modern solution-generating techniques are particularly fruitful in providing new exact solutions to this fairly old problem. In this paper we shall presume that the reader is already familiar with the basic features of colliding wave solutions, as described, for example, in the classic papers of Szekeres,[1] Kahn and Penrose,[2] and Nutku and Halil.[3]

A. The electrovac field equations

When one is discussing colliding gravitational plane waves, it is convenient to employ null coordinates u and v, in terms of which the line element assumes the form

$$ds^2 = 2g_{uv}\, du\, dv + \sum_{a,b=1}^{2} h_{ab}\, dx^a\, dx^b, \tag{1.1}$$

where g_{uv} and h_{ab} are functions of u and v alone, and the complex potential equations[4] assume the form

$$\begin{aligned}(f/\rho)\{\partial_u(\rho\mathscr{E}_v)+\partial_v(\rho\mathscr{E}_u)\} &= \{\mathscr{E}_u+2\Phi^*\Phi_u\}\mathscr{E}_v+\{\mathscr{E}_v+2\Phi^*\Phi_v\}\mathscr{E}_u,\\ (f/\rho)\{\partial_u(\rho\Phi_v)+\partial_v(\rho\Phi_u)\} &= \{\mathscr{E}_u+2\Phi^*\Phi_u\}\Phi_v+\{\mathscr{E}_v+2\Phi^*\Phi_v\}\Phi_u,\end{aligned} \tag{1.2}$$

where

$$f := \mathrm{Re}(\mathscr{E}) + \Phi^*\Phi. \tag{1.3}$$

The field ρ, defined by

$$\rho^2 := \det h, \tag{1.4}$$

satisfies the wave equation

$$\rho_{uv} = 0, \tag{1.5}$$

which means that it is expressible as the sum of a function of u and a function of v. The null coordinates u and v have frequently been chosen so that $\rho = 1 - u^2 - v^2$, but other choices are possible. We shall, however, always choose the coordinates so that $\rho(0,0) = 1$, $\rho_u(0,0) = \rho_v(0,0) = 0$, and the second derivatives exist at $u = v = 0$.

Once one has solved Eqs. (1.2) for $\mathscr{E}$ and Φ, one can construct the complete metric tensor and electromagnetic field by well-known methods. We can immediately identify $h_{22} = -f$. On the other hand, $h_{12} = h_{21} = -f\omega$, where ω may be determined up to an additive constant by integrating the equation

$$d\omega = \rho f^{-2} * [d\chi + 2\,\mathrm{Im}(\Phi^*\, d\Phi)], \tag{1.6}$$

where $\chi := \mathrm{Im}(\mathscr{E})$ is the twist potential, and where the two-dimensional duality operator $*$ has the property

$$*du = du, \quad *dv = -dv. \tag{1.7}$$

The remaining component h_{11} is then determined using Eq. (1.4).

The complete line element (1.1) involves one additional metric tensor component g_{uv}. This component can be determined by quadratures from $\mathscr{E}$ and Φ using the relations

$$\begin{aligned}\rho_u\gamma_u &= \tfrac{1}{2}\rho_{uu} + \rho\{\tfrac{1}{4}f^{-2}|\mathscr{E}_u + 2\Phi^*\Phi_u|^2 - f^{-1}|\Phi_u|^2\},\\ \rho_v\gamma_v &= \tfrac{1}{2}\rho_{vv} + \rho\{\tfrac{1}{4}f^{-2}|\mathscr{E}_v + 2\Phi^*\Phi_v|^2 - f^{-1}|\Phi_v|^2\},\end{aligned} \tag{1.8}$$

where

$$e^{2\gamma} := f g_{uv}. \tag{1.9}$$

B. SU(2,1) invariance

The field equations that we have been reviewing possess an intrinsic SU(2,1) symmetry, which was described long ago by Kinnersley.[5] For readers who are not familiar with that work we shall briefly review the analysis, using, how-

[a)] Permanent address: Institute of Modern Physics, Northwest University, Xian, People's Republic of China.

 0022-2488/89/030678-05$02.50

ever, a variational principle instead of the field equations themselves.

The variational principle from which the complex potential equations (1.2) can be derived has long been known to have the form

$$\delta \int \rho\left\{\frac{1}{4} f^{-2}(d\mathscr{E}+2\Phi^* d\Phi) *(d\mathscr{E}+2\Phi^* d\Phi)^* - f^{-1} d\Phi *d\Phi^*\right\}=0 \tag{1.10}$$

when expressed in terms of $\mathscr{E}$ and Φ. The structure of the Lagrangian density is strikingly similar to the quantities that enter the equations (1.8) that determine the field γ.

Not every solution of the complex potential equations (1.2) gives rise to a bona fide colliding wave metric. It must be possible to join the metric across the null hypersurfaces $u=0$ and $v=0$ to Petrov type N (or more degenerate) plane wave solutions. The constants k and l introduced by Ernst, García, and Hauser[6] in connection with their colliding wave condition can be generalized in the case of electrovac fields. In terms of $\mathscr{E}(u,v)$ and $\Phi(u,v)$ we shall define constants

$$\begin{aligned} k &:= \tfrac{1}{4} f(0,0)^{-2}|\mathscr{E}_v(0,0)+2\Phi(0,0)^*\Phi_v(0,0)|^2 \\ &\quad - f(0,0)^{-1}|\Phi_v(0,0)|^2, \\ l &:= \tfrac{1}{4} f(0,0)^{-2}|\mathscr{E}_u(0,0)+2\Phi(0,0)^*\Phi_u(0,0)|^2 \\ &\quad - f(0,0)^{-1}|\Phi_u(0,0)|^2, \end{aligned} \tag{1.11}$$

which will play the same role as the analogous constants in the vacuum case.

We shall find it convenient to study the SU(2,1) invariance of general expressions of the form

$$L_{\alpha\beta} := \tfrac{1}{4} f^{-2}[\mathscr{E}_\alpha + 2\Phi^*\Phi_\alpha][\mathscr{E}_\beta + 2\Phi^*\Phi_\beta]^* - f^{-1}\Phi_\alpha \Phi_\beta^*. \tag{1.12}$$

Following Kinnersley[5] we shall write

$$\mathscr{E} = (U-W)/(U+W), \quad \Phi = V/(U+W), \tag{1.13}$$

where, of course, only ratios of U, V, and W are significant, and f is easily seen to be given by

$$f = (U^*U + V^*V - W^*W)/|U+W|^2.$$

In the numerator of this expression we observe the manifestly SU(2,1) invariant combination

$$\kappa := U^*U + V^*V - W^*W. \tag{1.14}$$

A little additional work suffices to demonstrate that $L_{\alpha\beta}$ can be cast into the manifestly SU(2,1) invariant form

$$\begin{aligned} L_{\alpha\beta} &= \kappa^{-2}[U^*U_\alpha + V^*V_\alpha - W^*W_\alpha][U^*U_\beta \\ &\quad + V^*V_\beta - W^*W_\beta]^* - \kappa^{-1}[U_\alpha U_\beta^* \\ &\quad + V_\alpha V_\beta^* - W_\alpha W_\beta^*]. \end{aligned} \tag{1.15}$$

Thus any SU(2,1) transformation (with constant coefficients) of the fields U, V, and W will preserve the field equations (1.2) that determine $\mathscr{E}$ and Φ and the field equation (1.8) that determines γ. In particular, the determination of the transformed metric tensor component g_{uv} is trivial, since according to Eq. (1.9) the product of f and g_{uv} is invariant. Moreover, since the values of the constants k and l are not changed when SU(2,1) transformations are executed, such transformations will *always* yield bona fide colliding wave solutions when bona fide colliding wave seed metrics are employed!

Kinnersley[5] described the SU(2,1) group we have been discussing in terms of five simple classes of transformation:

gauge,

$$\begin{aligned} (U+W) &\to (U+W), \\ V &\to V + a(U+W), \\ (U-W) &\to (U-W) - 2a^*V - aa^*(U+W); \end{aligned} \tag{1.16}$$

gauge,

$$\begin{aligned} (U+W) &\to (U+W), \\ V &\to V, \\ (U-W) &\to (U-W) + i\alpha(U+W); \end{aligned} \tag{1.17}$$

duality rotation, rescaling,

$$\begin{aligned} (U+W) &\to b(U+W), \\ V &\to (b^*/b)V, \\ (U-W) &\to (1/b^*)(U-W); \end{aligned} \tag{1.18}$$

Ehlers,

$$\begin{aligned} (U+W) &\to (U+W) + i\beta(U-W), \\ V &\to V, \\ (U-W) &\to (U-W); \end{aligned} \tag{1.19}$$

Harrison,

$$\begin{aligned} (U+W) &\to (U+W) - 2c^*V - cc^*(U-W), \\ V &\to V + c(U-W), \\ (U-W) &\to (U-W). \end{aligned} \tag{1.20}$$

Here a, b, and c are arbitrary complex parameters and α and β are arbitrary real parameters.

The *Ehlers transformation* and the *Harrison charging transformation* (Ref. 7) were known for quite a while before Kinnersley identified them as members of an SU(2,1) internal symmetry group. Perhaps the most significant thing about these SU(2,1) transformations is the fact that they *always* map colliding wave solutions into colliding wave solutions, whereas, because asymptotic flatness was not always preserved, the corresponding transformations were relatively unproductive in connection with the generation of useful new stationary axisymmetric fields.

When one restricts attention to vacuum-preserving transformations, one obtains the *generalized Ehlers transformation*, under which

$$-i\mathscr{E} \to [\alpha(-i\mathscr{E})+\beta]/[\gamma(-i\mathscr{E})+\delta], \tag{1.21}$$

where α, β, γ, and δ are new parameters such that

$$\alpha\delta - \beta\gamma = 1. \tag{1.22}$$

This is equivalent to transformations of classes 4, 2, and 3 applied in succession, and has been discussed by Ernst, García, and Hauser (EGH).[6]

It should be mentioned that a transformation that was employed recently by Chandrasekhar and Xanthopoulos[8] to construct a charged version of the Nutku–Halil colliding wave solution[3] with a special value of the charge corresponds to the discrete SU(2,1) transformation

$$
\begin{aligned}
&U \to -V,\\
&V \to U, \qquad (1.23)\\
&W \to W.
\end{aligned}
$$

When a vacuum solution ($V=0$) is subjected to this transformation, the result is a solution with $\mathscr{E} = -1$ and $\Phi = \xi$, where $\xi := U/W$ is a solution of the associated complex potential equation

$$(\xi\xi^* - 1)(1/\rho)\{\partial_u(\rho\xi_v) + \partial_v(\rho\xi_u)\} = 4\xi^*\xi_u\xi_v. \qquad (1.24)$$

We shall see that this result can be absorbed naturally into a more complete result, characterized by an arbitrary charge value, obtained when a more general SU(2,1) transformation is applied to the same Nutku–Halil seed metric.

II. HARRISON'S CHARGING TRANSFORMATION

A. The FIB/EGH metric

Our objective will be to derive a new family of electrovac colliding wave solutions of the Einstein equations by applying a Harrison charging transformation to the three-parameter EGH solution[6] of the Einstein vacuum field equations. It will be recalled that the EGH solution was itself a generalization of a two-parameter solution derived earlier by Ferrari, Ibañez, and Bruni (FIB).[9] Among the solutions that will be obtained will be a charged version of the famous Nutku–Halil vacuum solution.[3]

Throughout the rest of this paper we shall employ null coordinates such that

$$\rho(u,v) = 1 - u^2 - v^2. \qquad (2.1)$$

The EGH metric was conveniently described in terms of auxiliary coordinates x and y such that

$$x := u\sqrt{1-v^2} + v\sqrt{1-u^2}, \quad y := u\sqrt{1-v^2} - v\sqrt{1-u^2}, \qquad (2.2)$$

and parameters n, ν, and ν'. Everything was expressed quite elegantly in terms of certain functions

$$
\begin{aligned}
T(n,\nu,\nu') :=\ & \frac{1}{2}\sqrt{1-x^2}\left\{(p+p')\left(\frac{1-x}{1+x}\right)^{n/2} + (p-p')\left(\frac{1+x}{1-x}\right)^{n/2}\right\}\\
&+ \frac{i}{2}\sqrt{1-y^2}\left\{(q+q')\left(\frac{1-y}{1+y}\right)^{n/2} + (q-q')\left(\frac{1+y}{1-y}\right)^{n/2}\right\}, \qquad (2.3)
\end{aligned}
$$

where

$$p := \cos\nu, \quad q := \sin\nu, \quad p' := \cos\nu', \quad q' := \sin\nu'. \qquad (2.4)$$

For example, the metric components h_{ab} were given by

$$
\begin{aligned}
h_{11} &= \rho^{1+n}|T(n+1,\nu',\nu)/T(n,\nu,\nu')|^2,\\
h_{22} &= \rho^{1-n}|T(n-1,\nu',\nu)/T(n,\nu,\nu')|^2, \qquad (2.5)\\
h_{12} &= -\rho\,\mathrm{Im}\{T(n-1,\nu',\nu)T(n+1,\nu',\nu)^*\}\\
&\quad \times(|T(n,\nu,\nu')|^2)^{-1},
\end{aligned}
$$

while the complex $\mathscr{E}$ potential was given by

$$\mathscr{E} = -\rho^{1-n}T(n-2,\nu,\nu')^*/T(n,\nu,\nu')^*, \qquad (2.6)$$

and g_{uv} was given by

$$g_{uv} = -\rho^{(n^2-1)/2}|T(n,\nu,\nu')|^2/\sqrt{(1-u^2)(1-v^2)}. \qquad (2.7)$$

Since we shall need the expressions later, we recall that for the EGH metric Kinnersley's H-potential[10] was given by

$$
\begin{aligned}
H_{11} &= -\rho^{1+n}[T(n+2,\nu,\nu')/T(n,\nu,\nu')]^*,\\
H_{22} &= -\rho^{1-n}[T(n-2,\nu,\nu')/T(n,\nu,\nu')]^*,\\
H_{12} &= i(1-n)z - 2\{x\,\mathrm{Im}\,T(n,\nu',\nu) \qquad (2.8)\\
&\qquad + iy\,\mathrm{Re}\,T(n,\nu',\nu)\}/T(n,\nu,\nu')^*,\\
H_{21} &= H_{12} - 2iz.
\end{aligned}
$$

For vacuum metrics it is possible to define an alternative E potential[11] by

$$E := (\rho + ih_{12})/h_{22}. \qquad (2.9)$$

In the case of the EGH metric this potential is given by

$$E(n,\nu,\nu') = \rho^n T(n+1,\nu',\nu)/T(n-1,\nu',\nu). \qquad (2.10)$$

There is another matrix potential H' that is of considerable importance in executing Harrison's charging transformation. This H potential is the one associated with a metric obtained from the EGH metric by applying a Kramer–Neugebauer involution $\mathscr{E} \leftrightarrow -E^*$ (Ref. 12).

Recall that the metric tensor h is expressible in the form

$$h = \frac{\rho}{F}\begin{pmatrix} F^2+\omega^2 & \omega \\ \omega & 1\end{pmatrix}, \qquad (2.11)$$

where $f = -h_{22} = -\rho/F$. The real part of the so-called *self-duality relation*[13]

$$\rho^{-1}h\epsilon * dH = -i\,dH, \qquad (2.12)$$

yields the relation

$$\rho^{-1}h\epsilon * dh = -d(\mathrm{Im}\,H). \qquad (2.13)$$

Consider now a Kramer–Neugebauer involution $\mathscr{E} \leftrightarrow -E^*$. The metric tensor h' of the resulting space-time is given by

$$h' = -\frac{\rho}{f}\begin{pmatrix} f^2+\chi^2 & \chi \\ \chi & 1\end{pmatrix}. \qquad (2.14)$$

Naturally, there is an H potential H' associated with this new space-time and the relation

$$\rho^{-1}h'\epsilon * dh' = -d(\mathrm{Im}\,H'). \qquad (2.15)$$

holds, too. Finally, direct evaluation of the left-hand side of this equation shows that

$$
\begin{aligned}
d(\mathrm{Im}\,H'_{11}) &= -\rho f^{-2}\,\mathrm{Im}[\mathscr{E}^{*2} * d\mathscr{E}], \qquad (2.16)\\
\mathrm{Im}\,H'_{22} &= \omega, \qquad (2.17)\\
d(\mathrm{Im}\,H'_{12}) &= dz + \tfrac{1}{2}\rho f^{-2} * d(\mathscr{E}\mathscr{E}^*), \qquad (2.18)\\
\mathrm{Im}\,H'_{21} &= \mathrm{Im}\,H'_{12} - 2z, \qquad (2.19)
\end{aligned}
$$

a result that will be of considerable significance in carrying out the Harrison transformation.

The effect of a Kramer–Neugebauer involution upon the EGH solution was discussed by Ernst, García, and Hauser.[6] Such a transformation maps one EGH solution into another. Bearing in mind that the value of $\operatorname{Im} H'_{22} = \omega$ is already known, it is apparent that the new solution is obtained from the original EGH solution by the substitutions

$$n \to 1 - n, \quad \nu \to \nu', \quad \nu' \to \nu + \pi. \tag{2.20}$$

In particular, the H potential H' is given by

$$\begin{aligned} H'_{11} &= -\rho^{2-n}[T(n-3,\nu',\nu)/T(n-1,\nu',\nu)]^*, \\ H'_{22} &= -\rho^{n}[T(n+1,\nu',\nu)/T(n-1,\nu',\nu)]^*, \\ H'_{12} &= inz + 2\{x \operatorname{Im} T(n-1,\nu,\nu') \\ &\quad + iy \operatorname{Re} T(n-1,\nu,\nu')\}/T(n-1,\nu',\nu)^*, \\ H'_{21} &= H'_{12} - 2iz. \end{aligned} \tag{2.21}$$

B. A more general seed solution

In the interest of achieving a greater degree of generality we shall subject the EGH metric to a generalized Ehlers transformation (1.21) before applying the Harrison charging transformation. Under such an Ehlers transformation, the f and χ potentials transform as follows:

$$f \to f/[\gamma^2 f^2 + (\delta + \gamma\chi)^2], \tag{2.22}$$

$$\chi \to \chi/[\gamma^2 f^2 + (\delta + \gamma\chi)^2]. \tag{2.23}$$

Moreover, the matrix potential H' undergoes the simple $SL(2,R)$ transformation

$$H' \to SH'S^T, \tag{2.24}$$

where

$$S = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \tag{2.25}$$

In summation, we find that the more general vacuum metric that we shall use as a seed solution has the $\mathscr{E}$ potential

$$\mathscr{E} = i\,\frac{i\alpha\rho^{1-n}T(n-2,\nu,\nu')^* + \beta T(n,\nu,\nu')^*}{i\gamma\rho^{1-n}T(n-2,\nu,\nu')^* + \delta T(n,\nu,\nu')^*}, \tag{2.26}$$

from which

$$f = -\rho^{1-n}\left|\frac{T(n-1,\nu',\nu)}{i\gamma\rho^{1-n}T(n-2,\nu,\nu')^* + \delta T(n,\nu,\nu')^*}\right|^2 \tag{2.27}$$

follows immediately. We shall also need the H'-matrix elements

$$\begin{aligned} H'_{11} = \{ &-\alpha^2\rho^{2-n}T(n-3,\nu',\nu)^* \\ &- \beta^2\rho^n T(n+1,\nu',\nu)^* \\ &+ 2\alpha\beta\,[i(n-1)zT(n-1,\nu',\nu)^* \\ &+ 2x \operatorname{Im} T(n-1,\nu,\nu') \\ &+ 2iy \operatorname{Re} T(n-1,\nu,\nu')]\}/T(n-1,\nu',\nu)^* \end{aligned} \tag{2.28}$$

and

$$\begin{aligned} H'_{22} = \{ &-\gamma^2\rho^{2-n}T(n-3,\nu',\nu)^* \\ &- \delta^2\rho^n T(n+1,\nu',\nu)^* \\ &+ 2\gamma\delta[i(n-1)zT(n-1,\nu',\nu)^* \\ &+ 2x \operatorname{Im} T(n-1,\nu,\nu') \\ &+ 2iy \operatorname{Re} T(n-1,\nu,\nu')]\}/T(n-1,\nu',\nu)^*. \end{aligned} \tag{2.29}$$

C. Determination of the new metric

When Harrison's charging transformation is applied to a vacuum metric, the resulting complex potentials assume the form

$$\mathscr{E} = \mathscr{E}_0/(1 - |c|^2\mathscr{E}_0), \quad \Phi = c\mathscr{E}_0/(1 - |c|^2\mathscr{E}_0), \tag{2.30}$$

where the subscript 0 distinguishes potentials of the vacuum seed metric from those of the final electrovac solution. It follows immediately that

$$f = f_0/(|1 - |c|^2\mathscr{E}_0|^2), \quad \chi = \chi_0/(|1 - |c|^2\mathscr{E}_0|^2). \tag{2.31}$$

Of course, $h_{22} = -f$ and $h_{12} = -f\omega$, where ω must be determined from the twist potential χ using the relation

$$*d\omega = \rho f^{-2}\{d\chi + 2 \operatorname{Im}(\Phi^* \, d\Phi)\}, \tag{2.32}$$

which, in the present case, assumes the form

$$*d\omega = \rho f^{-2}\{d\chi_0 - |c|^4 \operatorname{Im}(\mathscr{E}_0^{*2}\, d\mathscr{E}_0)\}. \tag{2.33}$$

The first term on the right-hand side is readily expressed in terms of the original ω_0 field. In principle, the second term could be evaluated by direct substitution, but that is not the easy way to construct the field ω. Instead, we take advantage of the abundance of potentials, the values of which we already know, writing the integral of Eq. (2.33) in the form

$$\omega = \operatorname{Im}\{H'^{0}_{22} + |c|^4 H'^{0}_{11}\}. \tag{2.34}$$

Into this expression we can substitute the matrix elements listed in Eqs. (2.29) and (2.30), thereby constructing ω, from which $h_{12} = h_{21}$ may easily be calculated. The most difficult component to calculate is h_{11}, but simplifications analogous to those that occurred in connection with the original EGH metrics occur here also.

The final expressions for the metrical fields of our new electrovac colliding wave metrics may be expressed as follows:

$$h = \frac{\rho}{N}\begin{pmatrix} K & L \\ L & M \end{pmatrix}, \tag{2.35}$$

where

$$K = |A|^2, \tag{2.36}$$

$$L = \operatorname{Im}(AB^*), \tag{2.37}$$

$$M = |B|^2, \tag{2.38}$$

$$N = \operatorname{Re}(AB^*). \tag{2.39}$$

The fields A and B are simple generalizations of the corresponding fields introduced by Ernst, García, and Hauser.[6] Here they have the values

$$\begin{aligned} A = \rho^{n(n+2)/4}\{ &(\gamma^2 + |c|^4\alpha^2)\rho^{2(1-n)}T(n-3,\nu',\nu) \\ &+ (\delta^2 + |c|^4\beta^2)T(n+1,\nu',\nu) \\ &+ 2|c|^2\rho^{1-n}T(n-1,\nu',\nu) \\ &+ 2i(\gamma\delta + |c|^4\alpha\beta)(n-1)\rho^{-n}zT(n-1,\nu',\nu) \\ &- 4(\gamma\delta + |c|^4\alpha\beta)\rho^{-n}[x \operatorname{Im} T(n-1,\nu,\nu') \\ &- iy \operatorname{Re} T(n-1,\nu,\nu')]\}, \end{aligned} \tag{2.40}$$

$$B = \rho^{n(n-2)/4}T(n-1,\nu',\nu), \tag{2.41}$$

which obviously reduce to the EGH values in the appropriate limit. It should be remarked that the ratio

$$E := A/B \tag{2.42}$$

does *not* satisfy the usual Ernst equation in the case of electrovac fields.

The expressions for M and N are particularly simple. One finds that

$$M = \rho^{n(n-2)/2} |T(n-1,\nu',\nu)|^2, \tag{2.43}$$

$$N = \rho^{n^2/2} |i(\gamma - i|c|^2\alpha)\rho^{1-n} T(n-2,\nu,\nu')^* + (\delta - i|c|^2\beta) T(n,\nu,\nu')^*|^2. \tag{2.44}$$

The EGH expression for g_{uv} was

$$g_{uv} = -\rho^{-1/2} \frac{N}{\sqrt{(1-u^2)(1-v^2)}}. \tag{2.45}$$

The same relation holds here, with N given by the preceding equation, as can easily be seen from the fact that the product $h_{22} g_{uv}$ is invariant under any SU(2,1) transformation.

III. ELECTRIFIED NUTKU–HALIL SOLUTION

For simplicity let us now turn off the generalized Ehlers transformation by setting $\alpha = \delta = 1$ and $\beta = \gamma = 0$. Our expressions for the fields A and B then simplify to

$$A = \rho^{n(n+2)/4} \{T(n+1,\nu',\nu) + 2|c|^2 \rho^{1-n} T(n-1,\nu',\nu) + |c|^4 \rho^{2(1-n)} T(n-3,\nu',\nu)\}, \tag{3.1}$$

$$B = \rho^{n(n-2)/4} T(n-1,\nu',\nu). \tag{3.2}$$

In particular, when $n = 0$, we obtain the electrified Nutku–Halil solution

$$A = T(1,\nu',\nu) + 2|c|^2 \rho T(-1,\nu',\nu) + |c|^4 \rho^2 T(-3,\nu',\nu), \tag{3.3}$$

$$B = T(-1,\nu',\nu). \tag{3.4}$$

The metrical fields are in this case given by

$$K = |T(1,\nu',\nu) + 2|c|^2 \rho T(-1,\nu',\nu) + |c|^4 \rho^2 T(-3,\nu',\nu)|^2, \tag{3.5}$$

$$L = \mathrm{Im}\{[T(1,\nu',\nu) + |c|^4 \rho^2 T(-3,\nu',\nu)] T(-1,\nu',\nu)^*\}, \tag{3.6}$$

$$M = |T(-1,\nu',\nu)|^2, \tag{3.7}$$

$$N = |T(0,\nu,\nu') + |c|^2 \rho T(-2,\nu,\nu')|^2. \tag{3.8}$$

This is a generalization of the electrified version of the Nutku–Halil solution found by Chandrasekhar and Xanthopoulos,[8] where our complex charge parameter c was restricted to values such that $|c| = 1$.

Since c always occurs in the combination $|c|^2\rho$, one might be tempted to rescale ρ, but this means relinquishing the usual relation $\rho = 1 - u^2 - v^2$. The situation is similar to the situation one encounters in the Schwarzschild solution, where one may "eliminate" the mass parameter m by rescaling all lengths. One can do it, and for calculational purposes one often does do it, but to do so can be misleading, giving one the false impression that the parameter m is unessential.

IV. CONCLUSIONS

Since one may construct new colliding wave solutions by applying SU(2,1) transformations to the Ernst potentials associated not only with the Killing vector ∂_{x^2} but with any other Killing vector, it is apparent that an endless supply of colliding wave solutions can be constructed quite mechanically. Whereas in the stationary axisymmetric case the maintenance of asymptotic flatness limited the success of early attempts to apply solution-generating techniques, in the case of colliding plane waves the same techniques are guaranteed to be fruitful.

Just how far one should go working out new solutions in this way is debatable. We believe a more useful endeavor will be to attempt to get a better grasp of the *general* problem, and this will oblige one to employ more up-to-date techniques, generally involving Riemann–Hilbert problems in one form or another. Indeed it is likely to be through the application of such techniques that the general initial value problem for colliding gravitational plane waves will eventually be solved. Then we shall be able to dispense with the inverted approach which has heretofore been employed in deriving new colliding wave solutions; i.e., working backward from the region of collision to the regions in which there are separate incident plane waves, a procedure that likely has given us a very lopsided perspective on the collision of gravitational plane waves.

ACKNOWLEDGMENT

This research was supported in part by Grant No. PHY-8605958 from the National Science Foundation.

[1] P. Szekeres, J. Math. Phys. **13**, 286 (1970).
[2] K. Khan and R. Penrose, Nature (London) **229**, 1575 (1971).
[3] Y. Nutku and M. Halil, Phys. Rev. Lett. **39**, 1379 (1977).
[4] F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[5] W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[6] F. J. Ernst, Alberto García D., and I. Hauser, J. Math. Phys. **28**, 2155, 2951 (1987); **29**, 681 (1988).
[7] B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[8] S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **410**, 311 (1987).
[9] V. Ferrari, J. Ibañez, and M. Bruni, Phys. Rev. D **36**, 1053 (1987).
[10] W. Kinnersley, J. Math. Phys. **18**, 1529 (1977).
[11] S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London Ser. A **396**, 55 (1984).
[12] D. Kramer and G. Neugebauer, Commun. Math. Phys. **10**, 132 (1968).
[13] I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980).

Dissipation in nonlinear response

Laurent P. Lévy and Andrew T. Ogielski
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

(Received 13 April 1988; accepted for publication 2 November 1988)

It is shown that the measurable nonlinear susceptibilities of third and higher orders probe only a fraction of equilibrium fluctuations, and in general can give only lower bounds on the equilibrium correlation functions. This contrasts with the linear and quadratic response, which completely determines the corresponding correlations by general fluctuation–dissipation theorems. The exactly solvable one-dimensional kinetic Ising model illustrates which fluctuations give rise to dissipation; only two-*magnon* processes of short wavelength and opposite momenta contribute to the third-order nonlinear susceptibilities, while two- and three-*magnon* processes contribute to the corresponding equilibrium correlation functions.

I. INTRODUCTION

Any thermodynamic relation between the driven response of a system and its equilibrium properties is of great practical value; the spectral density of fluctuations, the form of the correlation functions, becomes accessible experimentally by measuring the dynamical response to external fields (i.e., the susceptibilities). For the linear part of the response, the fluctuation–dissipation theorem (FDT) provides this essential connection.[1] For step-driven nonlinear processes,[2] the relaxation toward the equilibrium state can also be related to equilibrium correlation functions. If the equilibrium state is invariant under time reversal (an assumption which excludes systems with no well-defined time-reversed state, i.e., $p-n$ junctions, etc.), a more interesting generalization of the FDT relates third-order correlation functions to the second-order nonlinear susceptibility χ_2.[3] Most observable quantities are such that the response changes sign if the forces are reversed. In such symmetric systems, all odd equilibrium correlations and all even susceptibilities are zero. The most interesting generalizations of the FDT therefore involve the third-order response. Unfortunately, all attempts in this direction so far have been unsuccessful.[4]

In order to understand the source of this failure, we first consider the one-dimensional kinetic Ising model for which all the correlation functions and nonlinear susceptibilities can be calculated exactly. This model has a zero temperature critical point, and the nonlinearities arise from the change in the thermodynamic functions induced by the external forces, as well as the change of the correlation functions induced by the external field. The contributions of the fluctuations of the two-spin correlations to the third-order dissipation are dominated by two-*magnon* processes of very short wavelength and opposite momenta.[5] On the other hand, the third-order fluctuations involve two- and three-*magnon* processes of all wavelengths. This illustrates that nonlinear response is only sensitive to a fraction of the fluctuations, and only permits the setting of a lower bound on the spectrum of third-order fluctuations.

In more complex systems, the same conclusion can be reached using the Kubo formalism. The third-order susceptibilities are not sufficient to determine four-spin correlation functions. More precisely, some of the fluctuations do not contribute to the third-order dissipative response. The only circumstance where a third-order FDT exists is when one of the fields couples to the system with no dissipation; it is then possible to extend the second-order FDT to the third order. A susceptibility measures the response to the application of external forces. Thus only causal processes can contribute to susceptibilities. On the other hand, a correlation probes both causal and noncausal responses, since it is a measure of the probability that an observable quantity has a value M, knowing it had or will have a value M_0 at earlier or later times, respectively. This is why the measurements of nonlinear susceptibilities permit the obtaining of only a lower bound on the spectrum of equilibrium fluctuations. This thermodynamic inequality generalizes the fluctuation–dissipation theorem to nonlinear processes. It is particularly useful to probe the onset of long-range order in physical problems where higher-order correlations diverge.[6]

To reconstruct the correlation functions from the susceptibilities, an additional difficulty arises because of the zero-frequency anomalies[7] associated with their static limit. In higher-order correlation functions, singularities also occur whenever a mixing frequency vanishes. However, all these anomalies can be removed with a proper definition of the correlation functions.

This paper is organized as follows: In Sec. II we give the necessary definitions, and generalize the fluctuation–dissipation theorem to the second-order processes in the framework of Kubo's theory of nonlinear response. A simple procedure to include the singular contribution to the correlation functions is then introduced. In Sec. III, the nonlinear response and the correlation functions of the one-dimensional kinetic Ising model are evaluated and analyzed, to illustrate the point of the paper: nonlinear susceptibilities allow us to set only a lower bound on the four-spin correlation functions. The general theory of third-order processes is developed in Sec. IV, where we argue that only some of the higher-order fluctuations contribute to dissipation.

II. NONLINEAR RESPONSE

If (i) the equilibrium state of the system can be described by a density matrix, and (ii) the effect of the external fields can be included in a Hamiltonian $H = -\Sigma m^\alpha h_\alpha(t)$ linear in the external forces $h_\alpha(t)$, the nonlinear response of the system can be derived from a perturbation expansion of the time-dependent density matrix $\rho(t)$ in power of the external fields $h_\alpha(t)$, which are applied adiabatically from

 0022-2488/89/030683-06$02.50

$t = -\infty$.[8] This procedure enables us to parametrize each component α $(1,...,x)$ of the response $\langle M_\alpha(t)\rangle = \mathrm{Tr}[\rho(t)m_\alpha]$ to the x applied forces h_β in terms of the generalized susceptibilities χ,

$$\langle M^\alpha(t)\rangle = m_0^\alpha + \sum_{n=1} \frac{1}{n!}\int \chi^{\alpha\beta_1\cdots\beta_n}(t,t_1,...,t_n) \times h_{\beta_1}(t_1)\cdots h_{\beta_n}(t_n)dt_1\cdots dt_n, \quad (1)$$

which are quantities intrinsic to the *equilibrium* state of the system. Susceptibilities are observable functions of time (i.e., real) and can be defined in terms of the equilibrium-retarded Green's functions G_r,

$$\chi^{\alpha\beta_1\cdots\beta_n}(t,t_1\cdots t_n) = P_n G_r^{\alpha\beta_1\cdots\beta_n}(t,t_1\cdots t_n), \quad (2)$$

where we have used the symbol P_n for the sum over all possible permutations of $\{\beta_1,t_1\}\cdots\{\beta_n,t_n\}$. In equilibrium, the system is invariant under time translation, so χ and G_r depend only on the relative times, $\tau = t - t_1,...,\tau_{n-1} = t_{n-1} - t_n$. G_r coincides with the aftereffect function ϕ,

$$\phi^{\alpha\beta_1\cdots\beta_n}(\tau,\tau_1\cdots\tau_{n-1}) = (i/\hbar)^n\langle[[m^\alpha(\tau+\cdots+\tau_{n-1}), m^{\beta_1}(\tau_1+\cdots+\tau_{n-2})],...,m^{\beta_n}]\rangle, \quad (3)$$

for positive values of $\tau\cdots\tau_{n-1}$, and is zero otherwise. In Eq. (3), quantities in brackets $[m^\alpha, m^{\beta_1}]$ denote the commutators of the operators m^α and m^{β_1}. By convention, the thermodynamic average of a lower case quantity is performed with the equilibrium density matrix $\langle a\rangle = \mathrm{Tr}(\rho^0 a)$ and the operator $a(t) = e^{-iH_0t}ae^{iH_0t}$, where H_0 is the equilibrium Hamiltonian. Explicitly, the lowest-order susceptibilities are

$$\chi^{\alpha\beta_1}(t,t_1) = \chi^{\alpha\beta_1}(\tau) = \phi^{\alpha\beta_1}(\tau)\theta(\tau), \quad (4a)$$

$$\chi^{\alpha\beta_1\beta_2}(\tau,\tau_1) = \phi^{\alpha\beta_1\beta_2}(\tau,\tau_1)\theta(\tau)\theta(\tau_1) + \phi^{\alpha\beta_2\beta_1}(\tau+\tau_1,-\tau_1)\theta(\tau+\tau_1)\theta(-\tau_1), \quad (4b)$$

where $\theta(\tau) = 1$ if $\tau > 0$, and is zero otherwise. The knowledge of correlation functions or aftereffect functions ϕ is sufficient to determine the susceptibilities χ. However, the converse is not true; while the susceptibilities χ are causal quantities and vanish in all the time quadrants such that $t < \max(t_1,...,t_n)$, the aftereffect functions are nonzero at all times. This is one of the reasons why it is generally not possible to reconstruct from χ the complete spectrum of fluctuations.

If the operators $m_\beta(t)$ are eigenstates of the time-reversal operator T, they transform as $T\{m_\beta(t)\} = \epsilon_\beta m_\beta(-t)$, where the eigenvalues $\epsilon_\beta = \pm 1$. This permits us to evaluate ϕ for negative times when it is known at all times $\tau_i > 0$,

$$\phi^{\alpha\beta_1\cdots\beta_n}(-\tau,-\tau_1\cdots-\tau_{n-1}) = \epsilon_\alpha\epsilon_{\beta_1}\cdots\epsilon_{\beta_n}\phi^{\alpha\beta_1\cdots\beta_n}(\tau,\tau_1\cdots\tau_{n-1})^*. \quad (5)$$

Whereas this identity is not necessary to establish the linear FDT, it is essential in its generalization to higher-order susceptibilities.

We proceed to show that *at all orders* the correlation functions C_n,

$$C^{\alpha\beta_1\cdots\beta_n}(t,t_1\cdots t_n) = \langle m^\alpha(t)m^{\beta_1}(t_1)\cdots m^{\beta_n}(t_n)\rangle, \quad (6)$$

can be related to linear combinations of the aftereffect functions ϕ_n. For a set of $n+1$ thermodynamic quantities $m^\alpha(t), m^{\beta_1}(t)\cdots m^{\beta_n}(t)$, we can write a distinct correlation function for each of the $(n+1)!$ possible permutations of operators. The invariance of the trace under circular permutations,

$$C^{\beta_n\alpha\beta_1\cdots\beta_{n-1}}(t_n,t,t_1\cdots t_{n-1}) = T_n C^{\alpha\beta_1\cdots\beta_n}(t,...,t_n), \quad (7)$$

reduces the number of independent correlation functions C_n to $n! = (n+1)!/n$.[9] In Eq. (7), the operator $T_n = \exp(i\hbar\beta\,\partial/\partial t_n)$ translates t_n by $i\hbar\beta$. In Fourier space, T_n is the constant $\exp(-\hbar\beta\omega_n)$. Time translation invariance imposes the relation $T_0T_1\cdots T_n = 1$ within the operator T_i. The aftereffect functions are completely antisymmetric and therefore satisfy n Jacobi symmetry relations. For example, we have at the second order,

$$\phi^{\alpha\beta_1\beta_2}(\tau,\tau_1) + \phi^{\beta_1\alpha\beta_2}(-\tau,\tau+\tau_1) = 0, \quad (8a)$$

$$\phi^{\alpha\beta_1\beta_2}(\tau,\tau_1) + \phi^{\beta_1\beta_2\alpha}(\tau_1,-\tau-\tau_1) + \phi^{\beta_2\alpha\beta_1}(-\tau-\tau_1,\tau) = 0. \quad (8b)$$

These relations reduces the number of independent functions ϕ_n to $n!$. The linear system [Eq. (3)] defining the aftereffect function, in terms of correlation functions, can therefore be inverted, since except for discrete resonances which will be analyzed later, it is nonsingular. At the first and second orders, we find

$$(T_1 - 1)C^{\alpha\beta_1}(\tau) = \phi^{\alpha\beta_1}(\tau), \quad (9a)$$

$$(T_0-1)(T_1-1)(T_2-1)C^{\alpha\beta_1\beta_2}(\tau,\tau_1) = T_0\phi^{\alpha\beta_1\beta_2}(\tau,\tau_1) + \phi^{\beta_2\alpha\beta_1}(-\tau-\tau_1,\tau) + T_2^{-1}\phi^{\beta_1\beta_2\alpha}(\tau_1,-\tau-\tau_1). \quad (9b)$$

At first order, the only singularity in the correlation functions occurs at zero frequency ($T_1 = 1$ at $\omega = 0$). Additional resonances are present at higher orders when a mixing frequency becomes static (for $n = 2$, $T_1 = 1$ at $\omega_1 - \omega = 0$).

The linear FDT follows from Eq. (9a), since

$$\phi(\tau) = \phi(\tau)\theta(\tau) - \phi(-\tau)\theta(-\tau) = \chi(\tau) - \chi(-\tau).$$

At the second order, it is possible to express the aftereffect function ϕ in terms of the causal response functions χ, if the equilibrium state of the system is invariant under time reversal. To this effect, we divide the τ, τ_1 plane in six sectors, defined by

$$1 = \theta(\tau)\theta(\tau_1) + \theta(-\tau)\theta(\tau+\tau_1) + \theta(-\tau)\theta(-\tau_1) + \theta(\tau)\theta(-\tau-\tau_1) + \theta(\tau_1)\theta(-\tau-\tau_1) + \theta(-\tau_1)\theta(\tau+\tau_1). \quad (10)$$

After multiplication of Eq. (10) by the aftereffect function ϕ, we can rewrite each of the six terms on the right-hand side in terms of retarded Green's functions, using the symmetry relations (8) and the invariance under time reversal (5). Explicitly,

$$\phi^{\alpha\beta_1\beta_2}(\tau,\tau_1) = G_r^{\alpha\beta_1\beta_2}(\tau,\tau_1) + G_r^{\beta_1\alpha\beta_2}(-\tau,\tau+\tau_1) + \epsilon G_r^{\alpha\beta_1\beta_2}(-\tau,-\tau_1)^* - \epsilon G_r^{\beta_1\alpha\beta_2}(\tau,-\tau-\tau_1)^*$$

$$- (G_r^{\beta_1\beta_2\alpha}(\tau_1, -\tau-\tau_1) - \epsilon G_r^{\alpha\beta_2\beta_1}(-\tau-\tau_1,\tau_1)^*)$$
$$- (\epsilon G_r^{\beta_1\beta_2\alpha}(-\tau_1,\tau+\tau_1)^* - G_r^{\alpha\beta_2\beta_1}(\tau+\tau_1,-\tau_1)). \quad (11)$$

This expression greatly simplifies using the definition [Eq. (2)] of the susceptibilities in terms of the retarded Green's functions. Since the second-order response is nonzero only if $\epsilon = \epsilon_\alpha \epsilon_{\beta_1} \epsilon_{\beta_2} = 1$, it is convenient to introduce the time-reversal symmetric susceptibility

$$2\tilde{\chi}(\tau,\tau_1) = \chi(\tau,\tau_1) + \epsilon\chi(-\tau,-\tau_1)^*$$

whose Fourier transform is the real part of the susceptibility. With this convention, we find

$$\phi^{\alpha\beta_1\beta_2}(\tau,\tau_1) = 2(\tilde{\chi}^{\alpha\beta_1\beta_2}(\tau,\tau_1) - \tilde{\chi}^{\beta_1\beta_2\alpha}(\tau_1,-\tau-\tau_1)). \quad (12)$$

Quantum mechanically, there are $n!$ possible ways to symmetrize correlation functions, many of which correspond to observable quantities. For the sake of simplicity, only the completely symmetrized correlation function,

$$C_s^{\alpha\beta_1\beta_2}(\tau,\tau_1) = (1/3!)P_n C^{\alpha\beta_1\beta_2}(t,t_1,t_2), \quad (13)$$

is considered. If ω and ω_1 are the Fourier variables conjugated to τ and τ_1, Eqs. (9b), (12), and (13) permit us to establish the second-order FDT,[10]

$$\begin{aligned} C_s^{\alpha\beta_1\beta_2}(\omega,\omega_1) &= 2\,\mathrm{Re}(E(\omega,\omega_1)\chi^{\alpha\beta_1\beta_2}(\omega,\omega_1) \\ &\quad + E(-\omega_1,\omega-\omega_1)\chi^{\beta_2\alpha\beta_1}(-\omega_1,\omega-\omega_1) \\ &\quad + E(\omega_1-\omega,-\omega)\chi^{\beta_1\beta_2\alpha}\chi(\omega_1-\omega,-\omega)), \end{aligned} \quad (14)$$

where the spectral function E,

$$E(\omega,\omega_1) = -\hbar^2\left(\frac{1+e^{\beta\hbar\omega} + (e^{\beta\hbar(\omega-\omega_1)} + e^{\beta\hbar\omega_1})/2}{3(1-e^{\beta\hbar(\omega-\omega_1)})(1-e^{\beta\hbar\omega_1})}\right), \quad (15)$$

smears out the microscopic contributions to the macroscopic response at high frequencies. In the classical limit, the spectral function E becomes $1/\beta^2\omega(\omega_1-\omega)$ (See Refs. 3 and 11). This second-order FDT differs from the linear FDT in two important respects: (a) the determination of the correlation functions requires the knowledge of all independent components of the susceptibility tensor and not of $\chi^{\alpha\beta_1\beta_2}(\omega,\omega_1)$ alone, and (b) the dissipative response is described at the second order by the real part of the susceptibility.

It is appropriate here to discuss the analytic properties of the generalized susceptibilities. The retarded Green's functions are causal responses of all the time variables; in Fourier space, they are analytic functions of the complex frequencies, $\omega,...,\omega_{n-1}$, in the corresponding upper-half planes. On the other hand, generalized susceptibilities are causal only in the argument (τ) where the actual physical observation takes place. Analyticity is thus only warranted in the upper-half plane of the detected frequency ω. It follows that the usual Kramers–Kronig relation relating the real and imaginary parts of the response function holds for the first argument, but not for the others.[12]

The FDT (14) does not apply when any of the frequencies ω,ω_1, or $\omega-\omega_1$ vanishes, since correlation functions can remain finite at large times. To include these singular contributions, we redefine the correlation functions in terms of regular functions.[7] Explicitly, the correlation function C',

$$\begin{aligned} C'(\omega,\omega_1) &= C(\omega,\omega_1) - 2\pi[B(\omega)\delta(\omega_1) + D(\omega_1)\delta(\omega)] \\ &\quad + 4\pi^2 D\delta(\omega)\delta(\omega_1), \end{aligned} \quad (16)$$

is regular in the entire ω, ω_1 plane. The functions $B(\tau)$ and $D(\tau_1)$ are the asymptotic forms of the correlation functions as τ_1 or τ becomes, respectively, infinite. Similarly, $D = C(\pm\infty,\pm\infty)$. In the literature, a principle requiring the damping of correlation functions at large time[13] is often invoked to express B and D in terms of first-order correlation functions. Explicitly, the identity $B(\omega) = D(\omega) = P_3[C_{\alpha\beta_1}(\omega)\langle m_{\beta_2}\rangle]$ may in some circumstances be used. In random systems like spin glasses, this principle of damping of correlation functions does not hold, so one must exercise great care in the evaluation of B and D.

Let

$$\begin{aligned} U(\omega,\omega') &= (1-\exp(\beta\hbar\omega))(1-\exp(\beta\hbar\omega')) \\ &\quad \times (1-\exp(\beta\hbar(\omega-\omega'))). \end{aligned}$$

Since the singularities in C are delta functions which coincide with the zeros of U, UC is a regular function. Multiplying Eq. (14) by U, we have $UC = UC'$, since UC and UC' have no singularities. Since C' is regular, we perform, after division by U, a principal value (P) decomposition of the right-hand side, and obtain the correct form of the second-order FDT,

$$\begin{aligned} C_s^{\alpha\beta_1\beta_2}(\omega,\omega_1) &= 2\mathrm{P}\,\mathrm{Re}(E(\omega,\omega_1)\chi^{\alpha\beta_1\beta_2}(\omega,\omega_1) \\ &\quad + E(-\omega_1,\omega-\omega_1)\chi^{\beta_2\alpha\beta_1}(-\omega_1,\omega-\omega_1) \\ &\quad + E(\omega_1-\omega,-\omega)\chi^{\beta_1\beta_2\alpha}\chi(\omega_1-\omega,-\omega)), \end{aligned} \quad (17)$$

for all frequencies. Before considering the general third-order relations, it is useful to gain some insight with a simple example where the third-order response can be calculated exactly.

III. THE THIRD-ORDER RESPONSE OF THE 1-D KINETIC ISING MODEL

We consider a one-dimensional Ising system described by the Hamiltonian $H = -J\Sigma s_i s_{i+1} - \Sigma h_i s^i$, where the spin variables $s_i = \pm 1$. The Kubo formulation of nonlinear response cannot be used as such to discuss the dynamics of this system; the adiabatic response excludes any possible dissipation, unless a thermalization bath is explicitly included in the Hamiltonian. A spin-boson Hamiltonian has been introduced to circumvent this difficulty, in the context of the two-level dynamics coupled to a dissipative bath.[14] An alternative approach is to use stochastic dynamics, which evaluates the *isothermal* response of the system in the presense of a random noise. This isothermal response differs in several ways from Kubo's theory, and we will contrast the relevant results. The starting point here is a master equation for the probability P that the spin configuration of the N spins in the system is $\sigma(\sigma = \{s_1\cdots s_N\})$ at time t,

$$\frac{d}{dt}P(\sigma,t) = \Gamma(\sigma|\sigma')P(\sigma',t). \quad (18)$$

For single spin-flip dynamics,[15] the spin-flip probability Γ between configuration σ' and σ is zero, unless they differ by the sign of one spin,

$$\Gamma_0(\sigma|\sigma') = \sum_i \Gamma_0^i = \sum_i s_i s'_i \delta^i(\sigma|\sigma')\omega_i(\sigma'), \tag{19}$$

where the delta function is zero if the configuration σ and σ' differ by a spin other than i. In what follows, we choose the spin-flip probability $w_i(\sigma)$,

$$w_i(\sigma) = (1/\tau)[1 - s_i \tanh(\beta h_i(\sigma))], \tag{20}$$

consistent with detail balance, and giving in the high temperature limit a single relaxation time τ. In Eq. (20), the local field $h_i(\sigma) = h_i + J(s_{i-1} + s_{i+1})$.

With this choice, the local magnetization $\langle s_i \rangle$ and the two-spin correlation function $\langle s_i s_j \rangle$, which are the probability averages $\Sigma s_i P(\sigma,t)$ and $\Sigma s_i s_j P(\sigma,t)$, respectively, satisfy the equations of motion,

$$\tau \frac{d}{dt}\langle s_i \rangle = -\langle s_i \rangle + \langle \tanh(\beta h_i(\sigma)) \rangle, \tag{21a}$$

$$\tau \frac{d}{dt}\langle s_i s_j \rangle = -2\langle s_i s_j \rangle + \langle s_i \tanh(\beta h_j(\sigma)) \rangle + \langle s_j \tanh(\beta h_i(\sigma)) \rangle. \tag{21b}$$

In a one-dimensional system, this dynamics can be solved exactly at all orders in the external field $h(t)$. To wit, the thermodynamic field,

$$\tanh(\beta h_i(\sigma)) = x_i + y_i(s_{i-1} + s_{i+1}) + z_i s_{i-1} s_{i+1}, \tag{22}$$

can be computed to third order in the external field $h(t)$ by expanding the coefficients x, y, and z in powers of h,

$$\begin{aligned} x &= (1 - \tfrac{1}{2}\gamma^2)\beta h + (-2 + 4\gamma^2 - 3\gamma^4)[(\beta h)^3/6],\\ y &= (\gamma/2)(1 - (1 - \gamma^2)(\beta h)^2),\\ z &= -(\gamma^2/2)\beta h + (\gamma^2/6)(4 - 3\gamma^2)(\beta h)^3, \end{aligned} \tag{23}$$

where $\gamma = \tanh 2\beta J$. If the driving field is inhomogeneous, x, y, and z depend on the site i. The third-order response is then governed by the coupled dynamical equations,

$$\tau \frac{d}{dt}\langle s_i \rangle = -[(1 - 2y)\langle s_i \rangle - x - z\langle s_{i-1} s_{i+1} \rangle], \tag{24a}$$

$$\tau \frac{d}{dt}\langle s_i s_j \rangle = -[2\langle s_i s_j \rangle - 2y(\langle s_i s_{j-1} \rangle + \langle s_i s_{j+1} \rangle) - 2x\langle s_i \rangle - 2z\langle s_i s_{j-1} s_{j+1} \rangle], \tag{24b}$$

where we have assumed translation invariance [i.e., $h(t)$ is spatially uniform]. This isothermal dynamics explicitly includes the derivatives of the thermodynamic potential in x, y, and z. In Kubo's theory, the adiabatic response is computed between the same equilibrium states, and no change in the thermodynamic functions arises. This difference between adiabatic and isothermal response is very well understood at the linear level.[7] For nonlinear response, the usual continuation of the Green's function to imaginary time, used to establish the connection at the linear level, does not give a simple picture.

Replacing the correlation function $\langle s_i s_j \rangle$ in Eq. (24a) by its equilibrium value $a^{|j-i|}$ (where $a = \tanh \beta J$), we recover at the first order the well-known response function of the 1-D kinetic Ising model,

$$\chi(t) = \frac{\beta N}{\tau \cosh 2\beta J}\,\theta(t)\exp\left(-(1-\gamma)\frac{t}{\tau}\right), \tag{25}$$

where N is the number of spins in the chain. At the lowest order in h, it is also straightforward[16] to express three-spin correlation functions in terms of the local magnetization response,

$$\langle s_i s_{j-1} s_{j+1} \rangle(t) \approx (a^2 + a^{|j-i-1|} - a^{|j-i+1|})\langle s_j \rangle(t), \tag{26}$$

where $i \neq j$. At the third order, we also have to consider the change in the two-spin correlation induced by the external field,

$$\langle s_i s_j \rangle(t) = \langle s_i s_j \rangle + \int_{-\infty}^{+\infty} \chi_2(t - t_1, t_1 - t_2) h(t_1) h(t_2)\, dt_1\, dt_2, \tag{27}$$

where χ_2 may be interpreted as a noise susceptibility. The forces driving the two-spin correlation function away from equilibrium include three-spin correlations, which have been expressed in terms of linear response in Eq. (26). Since the system has translational invariance, the driven response of $r_m = \langle s_i s_{i+m} \rangle$ with $m > 0$ is governed by the linear system,

$$\tau \frac{d}{dt} r_m = -2r_m + \gamma(r_{m-1} + r_{m+1}) + g_0(t) + a^m g_1(t), \tag{28a}$$

where the driving terms g_0 and g_1 are, respectively,

$$\begin{aligned} g_0(t) &= 2((1 - a^2)/(1 + a^2))\beta h(t)\langle s \rangle(t),\\ g_1(t) &= \gamma(a - a^{-1})(\gamma\beta h\langle s \rangle + [(1 - a^2)/(1 + a^2)](\beta h)^2). \end{aligned} \tag{28b}$$

In addition, probability conservation requires that $r_0(t) = 1$ at all times. Glauber[15] has investigated the general solution of Eq. (28a) in the absence of inhomogeneous terms ($g_0 = g_1 = 0$). The response associated to g_0 and g_1 can be inferred from his results, noting that any particular solution of Eq. (28a) must have the form

$$r_m^{(0,1)}(t) = a^m + \frac{1}{\tau}\int_{-\infty}^{t} K_m^{(0,1)}(t - t') g^{(0,1)}(t')\, dt', \tag{29}$$

where the kernels $K^{0,1}$ satisfy the homogeneous system. As can be checked by substitution, the kernels K^0 and K^1, appropriate to the driving forces g_0 and g_1, are

$$K_m^{(0)}(t) = e^{-2t/\tau}\left[I_0\left(2\gamma\frac{t}{\tau}\right) + 2\sum_{k=1}^{m-1} I_k\left(2\gamma\frac{t}{\tau}\right) + I_m\left(2\gamma\frac{t}{\tau}\right)\right], \tag{30a}$$

$$K_m^{(1)}(t) = e^{-2t/\tau}\sum_{k=1}^{\infty} a^l\left[I_{m-k}\left(2\gamma\frac{t}{\tau}\right) - I_{m+k}\left(2\gamma\frac{t}{\tau}\right)\right], \tag{30b}$$

where I_n is the modified Bessel function of the first kind. This dynamical response of the two-spin correlation functions can be checked against the static limit, $r_m = [(\chi h)^2 + a^m(1 + m(\chi h)^2)]/(1 + (\chi h)^2)$, obtained by transfer matrix methods. We stress that only r_2, and therefore the response functions $K_2^{(0,1)}$, contribute to the third-order susceptibility. On the other hand, it is shown below that all the response functions $K_m^{(0,1)}$ enter in the determination of the third-order fluctuations. It is instructive to evaluate the Fourier transform of the two-spin correlation susceptibility,

$$\begin{aligned}\chi_2(\omega,\omega_1) &= 2(\beta/\tau)((1-a^2)/(1+a^2))K_2^{(0)}(\omega)\chi_1(\omega_1)\\ &\quad + (\beta/\tau)\gamma(a-a^{-1})K_2^{(1)}\\ &\quad \times[\gamma\chi_1(\omega_1) + \beta((1-a^2)/(1+a^2))]. \qquad (31)\end{aligned}$$

In the low temperature and low frequency limit, this susceptibility diverges as $\exp(2\beta J)$. This exhibits the same behavior as the linear susceptibility, as expected from scaling. Similarly, it displays critical slowing down on a characteristic time scale $\tau/(2(1-\gamma))$, which is half the relaxation time of linear susceptibility.

The third-order response function χ_3 can be similarly expressed in terms of the susceptibilities χ_1 and χ_2 as [cf. Eq. (24a)]

$$\begin{aligned}\chi_3(\omega,\omega_1,\omega_2) &= \chi_1(\omega)((N\beta)^2(4\gamma^2 - 3\gamma^4 - 2)/6\\ &\quad - N\beta\gamma(1-\gamma^2)\chi_1(\omega_1)\\ &\quad - (\gamma^2/2)\chi_2(\omega_1,\omega_2)). \qquad (32)\end{aligned}$$

Note that the last term is nontrivial in all the frequency arguments, and it is analogous to the adiabatic susceptibility in Kubo's theory.

We now consider the correlation function $C_{ijkl}(t,t_1,t_2) = \langle s_i, s_j(t) s_k(t+t_1) s_l(t+t_1+t_2)\rangle$, which is the joint probability that spin i has value s_i, spin j the value s_j at time t later, and so on. For Markov processes, it is well-known that this joint probability is expressible in terms of conditional probabilities,

$$\begin{aligned}C_{ijkl}(t,t_1,t_2) &= \sum_{\sigma\cdots} P(\sigma,t)s_i P(\sigma|\sigma_1,t)s_j\\ &\quad \times P(\sigma_1|\sigma_2,t_1)s_k P(\sigma_2|\sigma_3,t_2)s_l, \qquad (33)\end{aligned}$$

where the sum is carried out over all spin configurations. As shown by Glauber,[15] $q_l(t_2) = \Sigma P(\sigma_2|\sigma_3,t_2)s_l$ may be regarded as the expectation value of spin l when the initial configuration of the spin is σ_2. It is given by

$$q_l(t) = e^{-t/\tau}\sum_m s_m I_{l-m}(t). \qquad (34)$$

Thus we only need consider $C_{ijkl}(t,t_1,0)$. For the sake of simplicity, we will also assume $t=0$ and focus on the frequency dependence of the second argument. The conditional probability $P(\sigma|\sigma_1,t)s_k s_l$ may also be interpreted as the probability that spin k and l have value $r_{kl}(t_1)$, knowing the spin was in configuration σ at $t_1 = 0$. If $k = l$, $r_{kk}(t)$ is unity at all times. This condition determines r_{kl} for $k \geqslant l$ (see Ref. 15),

$$\begin{aligned}r_{kl}(t) &= a^{k-l} + e^{-2t/\tau}\sum_{m>n}(s_m s_n - a^{m-n})\\ &\quad \times\left[I_{k-m}\left(\gamma\frac{t}{\tau}\right)I_{l-n}\left(\gamma\frac{t}{\tau}\right)\right.\\ &\quad \left. - I_{k-n}\left(\gamma\frac{t}{\tau}\right)I_{l-m}\left(\gamma\frac{t}{\tau}\right)\right]. \qquad (35)\end{aligned}$$

The spatial averaged correlation function $C(0,t,0) = \Sigma C_{ijkl}(0,t,0)$ is obtained by summing the equilibrium value of two- and four-spin correlation functions over all sites,

$$\begin{aligned}&C(0,t,0) - N^2\left(\frac{1+a}{1-a}\right)^2\\ &\quad = 2\sum_{p=1}^{\infty}\left(2\left[\frac{1+a}{1-a}\right]\left(\frac{1}{1-a}K_p^{(0)}(t) - K_p^{(2)}(t)\right)\right.\\ &\qquad \left. - \left[\frac{(1+a)^2 + 2a}{(1-a)^2}\right]K_p^{(1)}(t)\right), \qquad (36)\end{aligned}$$

where the functions $K^{(0,1)}$ have been defined in Eq. (30), while $K^{(2)}$ is obtained by substituting pa^p to a^p in Eq. (30b).[17] We see that the dynamic correlation functions involve all the fluctuations in the two-spin correlation functions, and not simply the fluctuation in $\langle s_i s_{i+2}\rangle$, which governs the dissipative nonlinear response. In a different language, the dissipative response is dominated by $\langle s(q,0)s(0,0)s(-q,t)s(0,t)\rangle$, where the wave vector $q\sim 2/l$ is at half the zone boundary. This can be viewed as the emission of two "magnons" of opposite momenta. By contrast, all wave vectors contribute to the correlation function $C(0,t,0)$, which describes only a subset of the third-order fluctuations.

IV. THIRD-ORDER FLUCTUATIONS AND NONLINEAR RESPONSE

In this last section, we examine the relation between third-order fluctuations and dissipation within Kubo's theory. We have seen in Sec. II that correlation functions and aftereffect functions are related at all orders.[4] We therefore focus on their connection to susceptibilities. First, we consider the expansion in powers of h of the second-order quantity,

$$\begin{aligned}&\langle[[M^\alpha(t),M^{\beta_1}(t_1)],M^{\beta_2}(t_2)]\rangle\\ &\quad = \sum_{\beta_3}\int dt_3 Y^{\alpha\beta_1\beta_2\beta_3}(t,t_1,t_2,t_3)h_{\beta_3}(t_3), \qquad (37)\end{aligned}$$

where M^α is the magnetization operator in the presence of the external field. Y appears here as a linear susceptibility of a second-order aftereffect function. Using the expansion of M in terms of h, we relate the aftereffect functions ϕ to Y,

$$\begin{aligned}\phi^{\alpha\beta_1\beta_2\beta_3}(\tau,\tau_1,\tau_2) &= Y^{\alpha\beta_1\beta_2\beta_3}(\tau,\tau_1,\tau_2)\\ &\quad + \epsilon Y^{\alpha\beta_1\beta_2\beta_3}(-\tau,-\tau_1-\tau_2)^*. \qquad (38)\end{aligned}$$

Since the correlation functions are linearly related to the functions ϕ,[18] we establish a thermodynamic identity between correlation functions and the susceptibilities Y,

$$C^{\alpha\beta_1\beta_2\beta_3}(\omega,\omega_1,\omega_2) = 2\,\mathrm{Im}\Big(\sum E(\Omega_i,\Omega_j,\Omega_k) \times Y(\Omega_i,\Omega_j,\Omega_k)\Big), \tag{39}$$

where the sum is carried out over the permutations of (α,ω), $(\beta_1,\omega_1-\omega)$, $(\beta_2,\omega_2-\omega_1)$, and $(\beta_3,-\omega_2)$. The frequency arguments $\Omega_i=\Omega$, $\Omega_j=\Omega+\Omega_1$, and $\Omega_k=\Omega+\Omega_1+\Omega_2$ are determined by the frequency Ω, Ω_1, Ω_2, associated with the first three indices of Y (for α $\Omega=\omega$, for β_1 $\Omega_1=\omega_1-\omega$ and so on). This identity is reminiscent of the second-order FDT established in Sec. II, with the caveat that the susceptibility Y is not expressible in terms of third-order response functions χ_3 alone. There are several possible decompositions of Y in terms of causal and noncausal functions. The decomposition we consider is based on the second-order theory, and permits us to extend to the third order the second-order FDT (Sec. II), if the fluctuation in one of the variables is small.[19] We define the quantity $Z(\tau,\tau_1,\tau_2)$,

$$Z(\tau,\tau_1,\tau_2) = Y(\tau,\tau_1,\tau_2) - \chi(\tau,\tau_1,\tau_2) + \chi(-\tau,\tau_1,\tau_2), \tag{40}$$

which is expressible as a linear combination of the causal (G_r) and noncausal functions g, h, k. We define g as

$$g(\tau,\tau_1,\tau_2) = \phi(\tau,\tau_1,\tau_2)\theta(\tau)\theta(\tau_1)\theta(-\tau-\tau_1-\tau_2),$$

and h and k are obtained by circular permutation of τ, τ_1, and τ_2 in the theta functions. These four functions are actually related by two relations,[20] so Z depends only on two independent functions. A necessary and sufficient condition for Z to vanish is that $g(\tau,\tau_1,\tau_2)=0$ at all times. This result follows from the actual decomposition of Y in terms of G_r, g, h, and k. In this instance, Eq. (40) becomes a genuine third-order FDT analogous to the second-order result [Eq. (17)]. Physically, $g=0$ when the fluctuations in one of the variables are negligible. In this instance, it is possible to factorize ϕ as

$$\phi(\tau,\tau_1,\tau_2) = \phi_1(\tau,\tau_1)\phi_2(\tau_2),$$

where the function ϕ_2 is causal and vanishes at negative times.

A four-spin correlation function can be interpreted as a power spectrum when ω_1 vanishes;

$$C(\omega,0,\omega_2) = \langle m(\omega)m(-\omega)m(\omega_2)m(-\omega_2)\rangle$$

is then a positive definite quantity. In the same circumstances, $\mathrm{Im}\{\chi(\omega,0,\omega_2)\}$ and $\mathrm{Im}\{\chi(\omega_2,0,\omega)\}$ measure the power absorbed at frequency ω and ω_2, respectively. On the physical grounds that the power spectrum of the fluctuations can only be greater than their contribution to dissipation, we argue that $\mathrm{Im}[Z(\omega,0,\omega_2)+Z(\omega,0,-\omega_2)]$ must be positive. In addition, the use of the symmetry relations of χ under time reversal can be used to transform Eq. (39) into the inequality

$$(4/\omega^2\omega_2)\mathrm{Im}[\chi(\omega,0,\omega_2)+\chi(\omega,0,-\omega_2)] + (4/\omega\omega_2^2)\,\mathrm{Im}[\chi(\omega_2,0,\omega) + \chi(\omega_2,0,-\omega)] \leqslant \beta^3 C_c(\omega,0,\omega_2), \tag{41}$$

where C_c is the connected part of the correlation function C. This inequality can be viewed as the extension of the FDT to the third-order response.

To conclude, we have shown explicitly in the context of the one-dimensional Ising model that the third-order dissipation is sensitive to only a fraction of the third-order fluctuation. Similar conclusions have been argued to hold in the context of Kubo's theory. When the third-order correlations could be interpreted as a power spectrum, we obtained a lower bound of this spectrum with the third-order response functions.

[1]H. Nyquist, Phys. Rev. **32**, 110 (1928); H. B. Callen and T. A. Welton, *ibid.* **83**, 34 (1951); R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).

[2]W. Bernard and H. B. Callen, Rev. Mod. Phys. **31**, 1017 (1959).

[3]G. F. Efremov, Zh. Eksp. Teor. Fiz. **55**, 2322 (1968) [Sov. Phys. JETP **28**, 1232 (1969)].

[4]R. L. Stratonovich, Zh. Eksp. Teor. Fiz. **58**, 1612 (1970) [Sov. Phys. JETP **31**, 864 (1970)]; G. N. Bochkov and Yu. E. Kuzovlev, *ibid.* **72**, 238 (1977) [**45**, 125 (1977)]; B. Hao, Physica (Utrecht) A **109**, 221 (1981).

[5]By magnons, we mean here the wave-vector dependent magnetization $s(q)=\Sigma s_k\exp(ikq)$. The range of wave vectors involved in the dissipative response is determined here by the (short) range of the interactions in this model.

[6]L. P. Lévy and A. T. Ogielski, Phys. Rev. Lett. **57**, 3288 (1986); L. P. Lévy, Phys. Rev. B **38**, 4963 (1988).

[7]P. C. Kwok and T. D. Schultz, J. Phys. C **2**, 1196 (1969).

[8]Sh. M. Kogan, Zh. Eksp. Teor. Fiz. **43**, 304 (1962) [Sov. Phys. JETP **16**, 217 (1963)]; R. L. Peterson, Rev. Mod. Phys. **39**, 69 (1967).

[9]For x distinct components $\alpha=1\cdots x$, there are x^{n+1} $[(n+1)!]$ possible correlation functions, which are all expressible in terms of the x^{n+1} distinct functions given in Eq. (6). The existence of the symmetry relations (7) decrease further the number of independent functions. To find the number of independent functions, in terms of which all correlation functions can be computed, is to find how many distinct ways there are to color the $n+1$ vertices on a regular polyhedron with x possible colors. Two distinct ways cannot be obtained by rotation (or permutation) of one into the other. The solution of this classical combinatorial problem is

$$N(x,n+1) = \frac{1}{n+1}\sum_{i=1}^{n+1}\chi^{(n+1,i)} = \sum_{d|n+1}\phi(d)x^{(n+1)/d}.$$

We have used the notation $(n+1,i)$ for the largest common divisor of $n+1$ and i. The Euler function ϕ counts the number of integers less than d and prime to d [$\phi(1)=1,\phi(2)=1,\phi(3)=2,\phi(4)=2,\ldots$]. The term d is a divisor of $n+1$.

[10]Let Ω, Ω_1, and Ω_2 be the Fourier variable conjugated to t, t_1, and t_2. The relation between the Fourier transforms of $\chi(t;t_1,t_2)$ and $\chi(\tau,\tau_1)$ is $\chi(\Omega,\Omega_1,\Omega_2)=\delta(\Omega+\Omega_1+\Omega_2)\chi(\Omega,\Omega+\Omega_1)$. It follows that $\omega\equiv\Omega$ and $\omega_1\equiv\Omega+\Omega_1$.

[11]The classical limit of this second-order FDT differs from the form given in Efremov's paper, which contains several typographical errors.

[12]Explicitly, $\mathrm{Re}\,\chi(\omega,\omega_1)=1/\pi\int\mathrm{Im}\,\chi(\omega',\omega_1)/(\omega'-\omega)$.

[13]A. I. Akhiezer and S. Peletminski, *Methods of Statistical Physics* (MIR, Moscow, 1972).

[14]A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger, Rev. Mod. Phys. **59**, 1 (1987).

[15]R. J. Glauber, J. Math. Phys. **4**, 294 (1963); M. Suzuki and R. Kubo, J. Phys. Soc. Jpn. **24**, 51 (1968).

[16]This result is easily obtained in the static limit by transfer matrix methods. It can be verified by substitution in the equations of motion of the three-spin correlation functions, using $\langle s_i s_j s_k s_l\rangle_e = a^{l-k+j-i}$, if $l>k>j>i$.

[17]This result was derived carrying out the summation over an infinite ring, and using the addition theorem $I_m(2x)=\Sigma I_{m+p}(x)I_p(x)$.

[18]Explicit formulas relating C to ϕ are given by Stratonovich (Ref. 4).

[19]A physical example is given by magneto-acoustic and magneto-optical effects, where the measurement of a third-order susceptibility gives a measurement of spin–spin correlation functions.

[20]These relations are

$$G_r(\tau,\tau_1,\tau_2)+\epsilon G_r^*(\tau,\tau_1,\tau_2) = -h(\tau_2,-\tau-\tau_1-\tau_2,\tau) - \epsilon h^*(\tau,-\tau-\tau_1-\tau_2,\tau_2),$$

$$g(\tau,\tau_1,\tau_2)+\epsilon g^*(\tau,\tau_1,\tau_2) = -k(\tau_2,-\tau-\tau_1-\tau_2,\tau) - \epsilon k^*(\tau_2,\tau_1,\tau).$$

Lyapunov exponents as a measure of the size of chaotic regions

Paul W. Cleary
Department of Mathematics, Monash University, Clayton, Victoria, 3168, Australia

(Received 22 April 1988; accepted for publication 2 November 1988)

Chaotic orbits in conservative dynamical systems are partitioned into a number of distinct regions by sets of cantori. This causes long time scale oscillations in the corresponding Lyapunov spectrums. A direct correlation between the magnitude of the maximal Lyapunov exponent and the size of the corresponding chaotic region in the surface of section appears to exist for the two-dimensional conservative systems examined.

I. INTRODUCTION

When exploring the local structure of a flow the only information available is the behavior of individual orbits over finite time intervals. An orbit can be characterized by its rate of divergence from adjacent orbits. More generally, the rates of growth of a volume element of phase space in various directions can be used to describe the orbits or attractors that pass through such volume elements. These rates of growth are measured by the Lyapunov exponents. They allow us not only to determine whether individual orbits are chaotic, but also to compare orbits and determine whether or not one is more chaotic than another. Furthermore, they provide a relatively simple method by which the dimensions of various attractors, arising from many dissipative systems, can be calculated.

The Lyapunov exponents of an m-dimensional mapping $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$, where $\mathbf{x}$ is an m-dimensional vector, are defined to be

$$\lambda_i = \log\Big[\lim_{n\to\infty} [j_i(n)]^{1/n}\Big], \quad i = 1,2,...,m, \tag{1}$$

where $j_i(n)$ is the magnitude of ith eigenvalue of the matrix

$$J_n = [J(x_n)J(x_{n-1})\cdots J(x_2)J(x_1)].$$

Here $J(x) = (\partial\mathbf{F}/\partial\mathbf{x})(x)$ is the Jacobian matrix of the mapping. The Lyapunov exponents are ordered by size,

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m,$$

and generally depend on the choice of the initial condition x_1. Any two initial points on the same orbit will give the same Lyapunov spectrum. Therefore any two points in the same connected chaotic region of phase space will produce the same values for the Lyapunov exponents, since there exists a single chaotic orbit that eventually densely fills this entire region.

The Lyapunov numbers were originally defined by Oseldec[1] and are given by Farmer *et al.*[2] The values λ_i, for $i = 1,2,...,m$, measure the rates of expansion of a volume element of phase space in the m principal directions. If $\lambda_i > 0$ then the volume element expands exponentially in the corresponding direction. If $\lambda_i = 0$ then the growth is linear and if $\lambda_i < 0$ then the volume element shrinks in that direction. The directions in which the Lyapunov exponents are calculated vary continuously as one moves around an orbit or attractor. An orbit is chaotic if the maximal Lyapunov exponent λ_1 is positive. Furthermore in a conservative system we have $\Sigma_{i=1}^m \lambda_i = 0$ and for a dissipative system $\Sigma_{i=1}^m \lambda_i < 0$, since the volume of a given element of phase space is always invariant for a conservative system and always decreasing for a dissipative system.

The previous definition of the Lyapunov exponents is not useful for their numerical calculation. The elements of the matrix J_n rapidly become very large, reflecting the exponential divergence of the orbits. After only a few iterations of any numerical scheme the magnitudes of the eigenvalues $j_i(n)$ become too large for computation. Several authors have examined this problem of numerically determining Lyapunov exponents for both systems of differential equations and chaotic time series. We adopt the notation of Benettin *et al.*,[3] who discussed the relationship between the maximal Lyapunov exponent and Kolmogorov entropy.

Consider an m-dimensional compact connected Riemannian manifold of at least class C^2. For $x \in M$, we denote the tangent space to M at x and the norm induced on it by the Riemannian metric on M by E_x and $\|\circ\|$, respectively. Let X be a vector field of class C^2 on M and T^t be the flow induced by X. Then for any t we have $T^t x = x(t)$, where $\{x(t): t\in\mathbb{R}\}$ is an integral curve or orbit of the vector field X such that $x(0) = x$. We also denote the tangent mapping of E_x onto $E_{T^t x}$, induced by the diffeomorphism T^t, by dT^t_x. Shimada and Nagashima[4] defined the k-dimensional Lyapunov exponents to be

$$\lambda(\mathbf{e}^k, x_0) = \lim_{t\to\infty} \ln \frac{\|dT^t_{x_0}\mathbf{e}_1 \wedge dT^t_{x_0}\mathbf{e}_2 \wedge \cdots \wedge dT^t_{x_0}\mathbf{e}_k\|}{\|\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \cdots \wedge \mathbf{e}_k\|}, \tag{2}$$

for $k = 1,2,...,m$, where $\mathbf{e}^k$ is a k-dimensional subspace of the tangent space E_{x_0} at x_0 and $\{\mathbf{e}_i: i = 1,2,...,k\}$ are bases for $\mathbf{e}^k$. Here $\wedge$ denotes an exterior product of tangent vectors. The exponent (2) represents the expansion rate of the k volume of the k-dimensional parallelepiped in the tangent space along the orbit starting at x_0. The tangent vectors $\mathbf{e}_i$, for $i = 1,2,...,m$, in the bases correspond to the vertices of the parallelepiped. These exponents do not depend on the choice of either bases or norm, but only on the initial point x_0 and the k-dimensional subspace $\mathbf{e}^k$.[1,2,4] The existence and finiteness of the limits involved in (2) were proved by Oseldec.[1] The one-dimensional exponents $\lambda(\mathbf{e}^1, x)$ take at most m values and correspond to the values obtained from the normal definition of Lyapunov exponents. The k-dimensional exponents $\lambda(\mathbf{e}^k, x)$ may take at most C^m_k distinct values.[4] These are equal to the sum of k distinct one-dimensional Lyapunov exponents. Benettin *et al.*[5] proved that if a basis set

 0022-2488/89/030689-07$02.50

$\{\mathbf{e}_i : i = 1,2,...,m\}$, for the tangent space, is chosen at random then each of the k-dimensional exponents $\lambda(\mathbf{e}^k,x)$ will converge to the maximal value of the corresponding set of C_k^m distinct possible values, for $k = 1,2,...,m$. Since the tangent mapping dT_x^t satisfies the chain rule $dT_{x_0}^{t+s} = dT_{x_s}^t \circ dT_{x_0}^s$, where $x_s = T^s x_0$, and the numerical integration of the differential equations gives a discrete mapping, the k-dimensional Lyapunov exponents are usually written as

$$\lambda(\mathbf{e}^k,x_0) = \lim_{n\to\infty} \frac{1}{n\tau} \sum_{j=0}^{n-1} \ln \frac{\| \wedge_i dT_{x_j}^{\tau} \mathbf{e}_i^j \|}{\| \wedge_i \mathbf{e}_i^j \|}, \tag{3}$$

where $\{\mathbf{e}_i^j : i = 1,2,...,k\}$ is a basis for the subspace $\mathbf{e}^k$ at the jth time step. Then $\mathbf{e}_i^j = dT_{x_0}^{j\tau}\, \mathbf{e}_i^0$. Gram–Schmidt reorthonormalization is performed at each time step to prevent the basis from becoming degenerate.

The Shimada and Nagashima algorithm using (3) can only be applied to dynamical systems where the evolution of the tangent vectors from $\mathbf{e}_i^j$ to $dT_{x_j}^{\tau}\, \mathbf{e}_i^j$ can be determined either by use of differential equations or discrete mappings. If instead only a time series is available, then it is necessary to numerically estimate a linearized flow map A^t on the tangent space from the observed data. Sano and Sawada[6] presented a technique for estimating A^t. Wolf *et al.*[7] also present an algorithm for calculating Lyapunov exponents from chaotic time series. Their method is similar to the Shimada and Nagashima approach and gives good results for the first few positive Lyapunov exponents.

In this paper we examine the behavior of the Lyapunov spectrums for a series of quartic polynomial potentials. We find that long term oscillations, not present for simpler systems such as the Hénon,[8] Rössler,[9] and Lorenz[10] systems, are introduced by the existence of cantori. We also investigate the apparent relationship between the maximal Lyapunov exponent and the area of the corresponding region in the surface of section.

II. CANTORI AND LONG PERIOD OSCILLATIONS

The Lyapunov spectrums of simple two- and three-dimensional dissipative systems such as the Hénon,[8] Rössler,[9] and Lorenz[10] systems and of various experimental time series have been studied previously.[4–7] We shall examine some properties of the Lyapunov exponents, not found in the above systems, using a class of quartic polynomial potentials whose dynamical behavior is quite complicated and varied. The Verhulst potentials are all of the form

$$V(x,z^2) = \frac{1}{2}(\omega_1^2 x^2 + \omega_2^2 z^2) - \left(\frac{A_1}{3} x^3 + A_2 x z^2\right) - \left(\frac{B_1}{4} x^4 + \frac{B_2}{2} x^2 z^2 + \frac{B_3}{4} z^4\right), \tag{4}$$

where the coefficients ω_1^2, ω_2^2, A_1, A_2 and B_1, B_2, B_3 are all real. It arises from truncating a Taylor series expansion of a general axisymmetric discrete-symmetric potential at fourth order.[11]

All the Lyapunov exponents belonging to any quasiperiodic orbit in a conservative dynamical system vanish. Since the system (1) is conservative only the Lyapunov spectrums for the chaotic orbits are calculated. Two Lyapunov exponents vanished for every chaotic orbit studied in all the Verhulst potentials. We therefore have $\lambda_4 = -\lambda_1$ and $\lambda_2 = \lambda_3 = 0$. The accuracy of the Lyapunov spectrum was checked by ensuring that $|\lambda_1 + \lambda_4|$, $|\lambda_2|$, and $|\lambda_3|$ were all less than 0.0005.

Three nonintegrable Verhulst potentials U_1, U_2, and U_3 are given in Table I. Figures 1–3 show the corresponding surfaces of section for the energy $E = 1000$. A series of homogeneous Verhulst potentials V_1, V_2, V_3, V_4, V_5, V_6, and V_7 also given in Table I, were found using the first two steps of a generalization of Painlevé analysis.[12,13] Their derivation and an examination of their properties will be presented elsewhere. The surfaces of section corresponding to these potentials are given in Figs. 4–10.

The surfaces of section in Figs. 2–4 possess substantial chaotic regions. These regions are not filled uniformly. They are partitioned into distinct areas by families of cantori (circular cantor sets). A cantori[14,15] is the remnant of an invariant curve with irrational winding number. It can be visualized as a circle with infinitely many gaps deleted by overlapping nearby island chains. The flux of orbits through such a cantori is typically very small and varies from one cantori to another according to the corresponding winding number. This causes the orbit to be restricted to a band between adjacent cantori sometimes for very long periods of

TABLE I. The maximal Lyapunov exponents for the principle chaotic orbits in a selection of the quartic polynomial potentials and the figure number of the corresponding surface of section.

	Potential	Fig.	Maximal Lyapunov exponent λ_1
U_1	$\frac{1}{2}(x^2 + \frac{1}{2}z^2) - (\frac{4}{9}x^3 + xz^2) + (\frac{1}{12}x^4 + \frac{1}{2}x^2z^2 + \frac{1}{4}z^4)$	1	0.05 ± 0.005
U_2	$\frac{1}{2}(x^2 + z^2) - (\frac{1}{3}x^3 + xz^2) + (\frac{1}{12}x^4 + \frac{1}{4}x^2z^2 + \frac{1}{12}z^4)$	2	0.252 ± 0.007
U_3	$\frac{1}{2}(x^2 + z^2) - (\frac{1}{3}x^3 + xz^2) + (\frac{1}{12}x^4 + \frac{1}{2}x^2z^2 + \frac{1}{8}z^4)$	3	0.13 ± 0.015
V_1	$\frac{8}{3}x^4 + 2x^2z^2 + \frac{1}{136}z^4$	4	0.996 ± 0.007
V_2	$6\frac{34}{163}x^4 + 2x^2z^2 + \frac{1}{66}z^4$	5	0.266 ± 0.008
V_3	$\frac{32}{5}x^4 + 2x^2z^2 + \frac{32}{1517}z^4$	6	0.155 ± 0.006
V_4	$17\frac{4}{21}x^4 + 2x^2z^2 + \frac{1}{21}z^4$	7	0.075 ± 0.004
V_5	$18\frac{2}{11}x^4 + 2x^2z^2 + \frac{32}{725}z^4$	8	0.082 ± 0.004
V_6	$24\frac{1}{7}x^4 + 2x^2z^2 + \frac{1}{28}z^4$	9	0.078 ± 0.004
V_7	$32\frac{1}{9}x^4 + 2x^2z^2 + \frac{1}{36}z^4$	10	0.081 ± 0.007

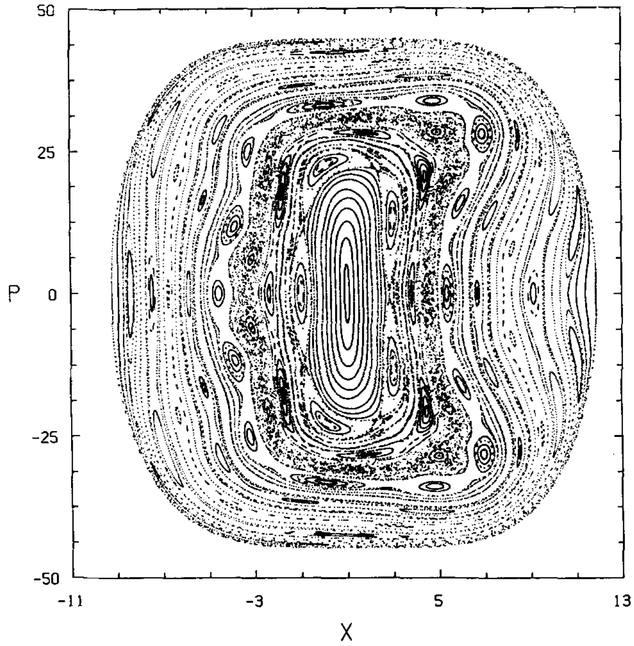


FIG. 1. Surface of section for the potential U_1 at energy $E = 1000$.

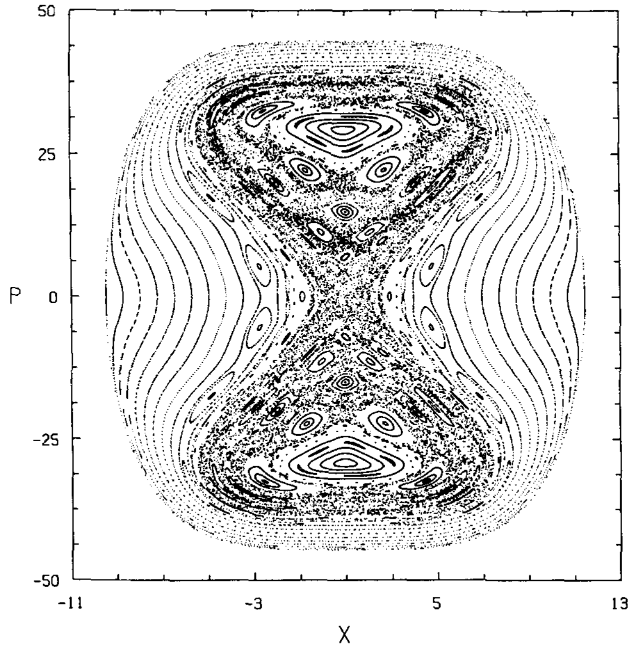


FIG. 3. Surface of section for the potential U_3.

time. Occasionally it diffuses through the cantori into another such region for a further period of time. After a long period of time all the accessible cantori will be visited and the chaotic region in the surface of section generally appears to be filled relatively evenly.

In each of these cantori bounded regions the chaotic orbit has a slightly different effective Lyapunov spectrum. The effect of occupying one area for an extended period of time and then moving to another causes very long time scale oscillations in the Lyapunov exponents. Figure 11(a) shows the maximal Lyapunov exponent for the potential U_1 over a large number of iterations of the numerical integration. For the first five million points the orbit has not been sufficiently sampled to provide good estimates of the Lyapunov exponents. After five million points the values of $|\lambda_4 + \lambda_1|$, $|\lambda_2|$, and $|\lambda_3|$ have all reached sufficiently small values to indicate that the maximal Lyapunov exponent is asymptotically converging to about 0.03. However, the maximal Lyapunov exponent then grows to 0.05 over the next five million points. For the next twenty million points it has an average value of about 0.05. Additionally, there are clear long time scale oscillations superimposed on this average value with a magnitude of 12% of the average value of the maximal Lyapunov exponent. Over this entire region $|\lambda_4 + \lambda_1|$, $|\lambda_2|$, and $|\lambda_3|$ all remain suitably small and give no indication of the existence

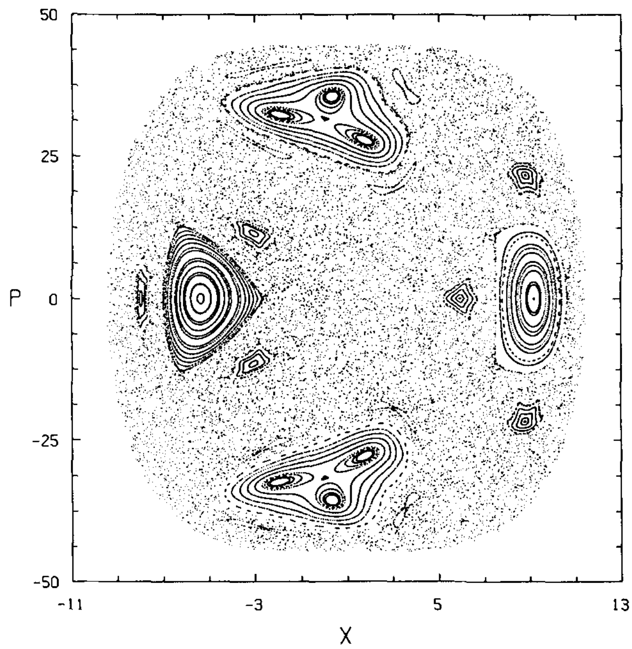


FIG. 2. Surface of section for the potential U_2.

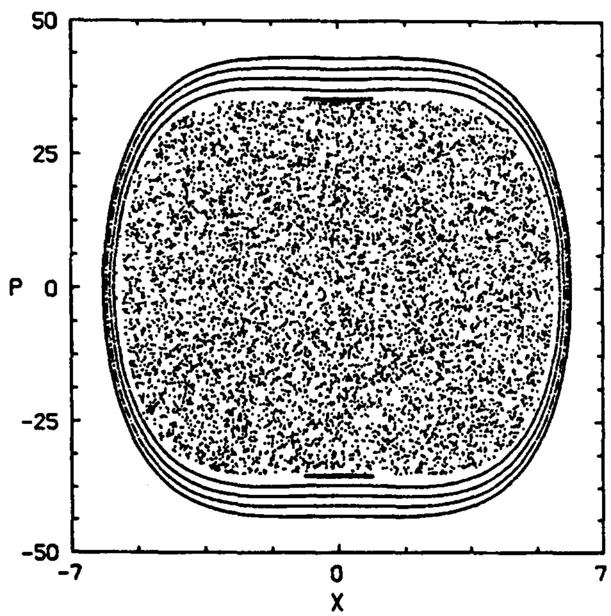


FIG. 4. Surface of section for the potential V_1.

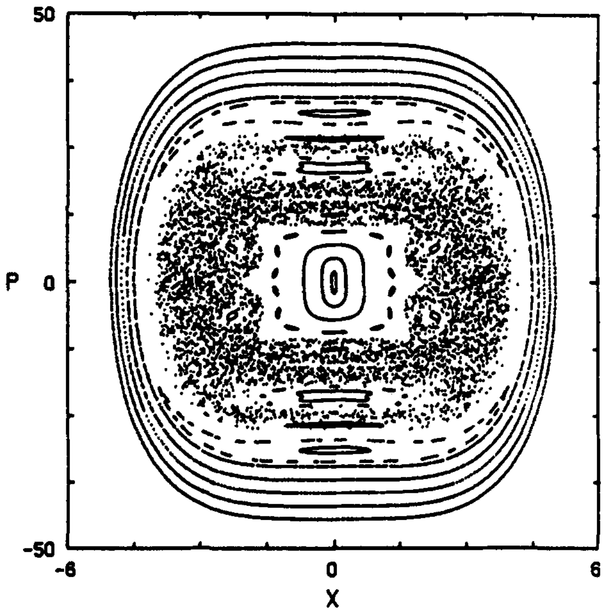


FIG. 5. Surface of section for the potential V_2.

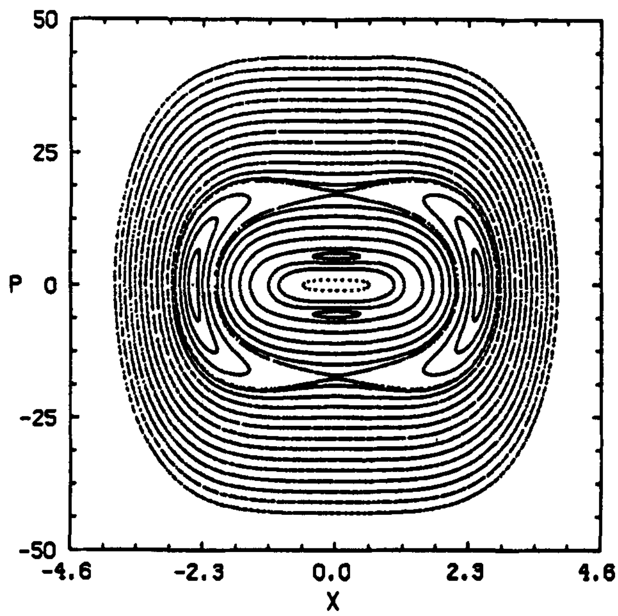


FIG. 7. Surface of section for the potential V_4.

of the long period oscillations. These oscillations do not arise because of incorrect estimation of the exponents due to insufficient sampling but because of a real change in the region in which the exponent is being calculated. The high-frequency fluctuations in the first five to ten million points are the normal oscillations seen in the Lyapunov spectrums and correspond to non-negligible values of $|\lambda_4 + \lambda_1|$, $|\lambda_2|$, and $|\lambda_3|$. As the number of points increases the magnitude of the noise decreases leading to a more accurate estimation of the spectrum. The curve in Fig. 11(a) is almost completely smooth over the last five million points. Figure 11(b) shows the maximal Lyapunov exponent, for the same chaotic region, using a different set of initial conditions for 100 million points. It rapidly approaches the expected average value of about 0.05. The long time scale oscillations are still obvious even after 100 million iterations.

In contrast Fig. 11(c) shows the maximal Lyapunov exponent for the Lorenz system with $\sigma = 16$, $R = 45.92$, and $b = 4$. After 2×10^5 iterations it has settled down to an asymptotic value of exactly 1.500. There are absolutely no long time scale oscillations present in the maximal Lyapunov exponent for this system. This occurs because the Lorenz attractor is a single structure and is not partitioned into a number of loosely connected pieces.

Figure 11(d) shows the maximal Lyapunov exponent for the potential V_1 and is a striking example of the movement between different cantori. The maximal Lyapunov exponent is not well estimated until about 1.5 million points

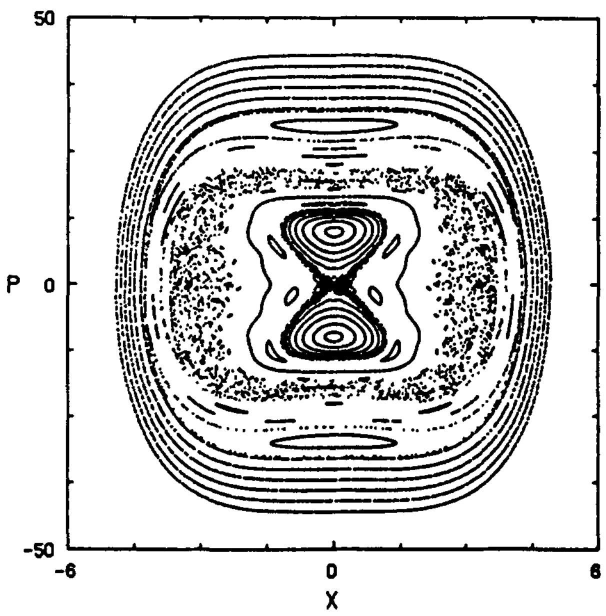


FIG. 6. Surface of section for the potential V_3.

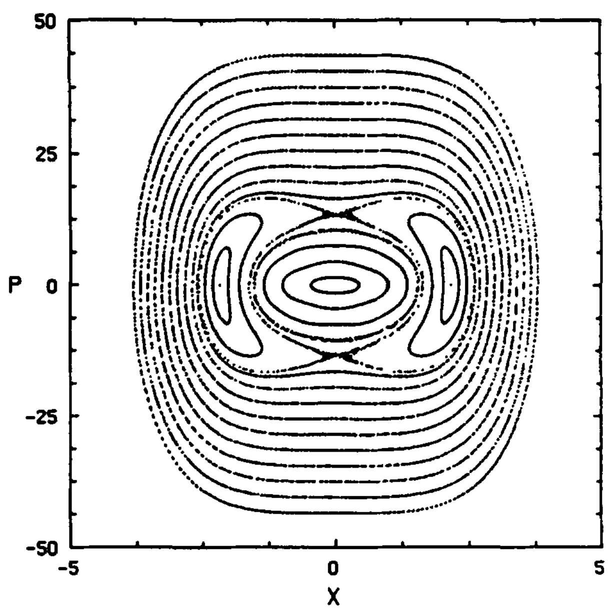


FIG. 8. Surface of section for the potential V_5.

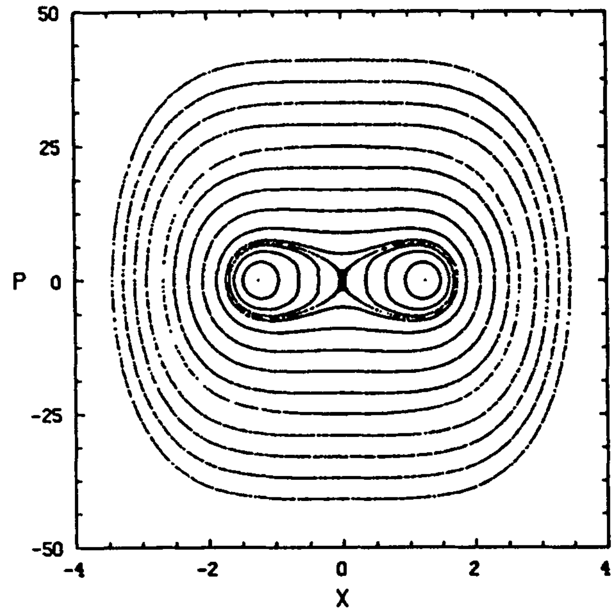


FIG. 9. Surface of section for the potential V_6.

have been used. The maximal Lyapunov exponent is then constant, as expected, for half a million points and then abruptly drops. It is then almost constant for another million points. This is followed by another abrupt increase in value, constancy for another million points, and another drop. This is followed by two more well-defined steps. Each of the steps corresponds to the chaotic orbit diffusing through a cantori from one region to an adjacent one. The constancy of the maximal Lyapunov exponent in each region after the rapid transition is precisely the behavior normally expected for Lyapunov exponents in attaining their asymptotic values. In this case the percentage change in each step is small. There is, as usual, high-frequency noise superimposed on the steplike average values of the maximal Lyapunov exponent.

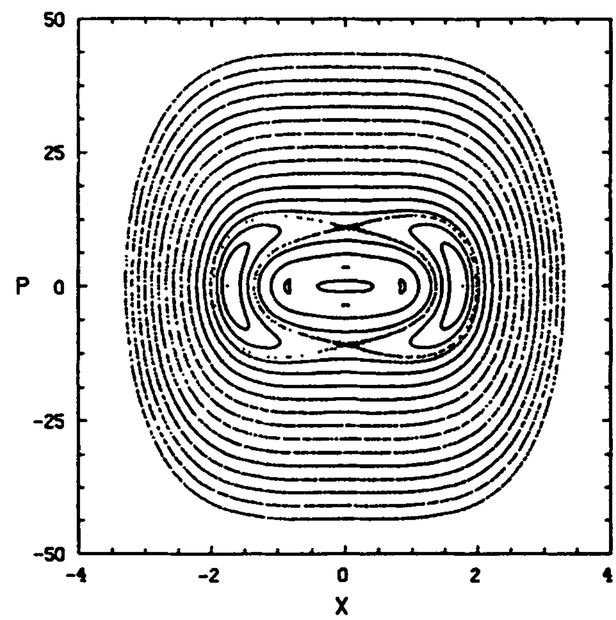


FIG. 10. Surface of section for the potential V_7.

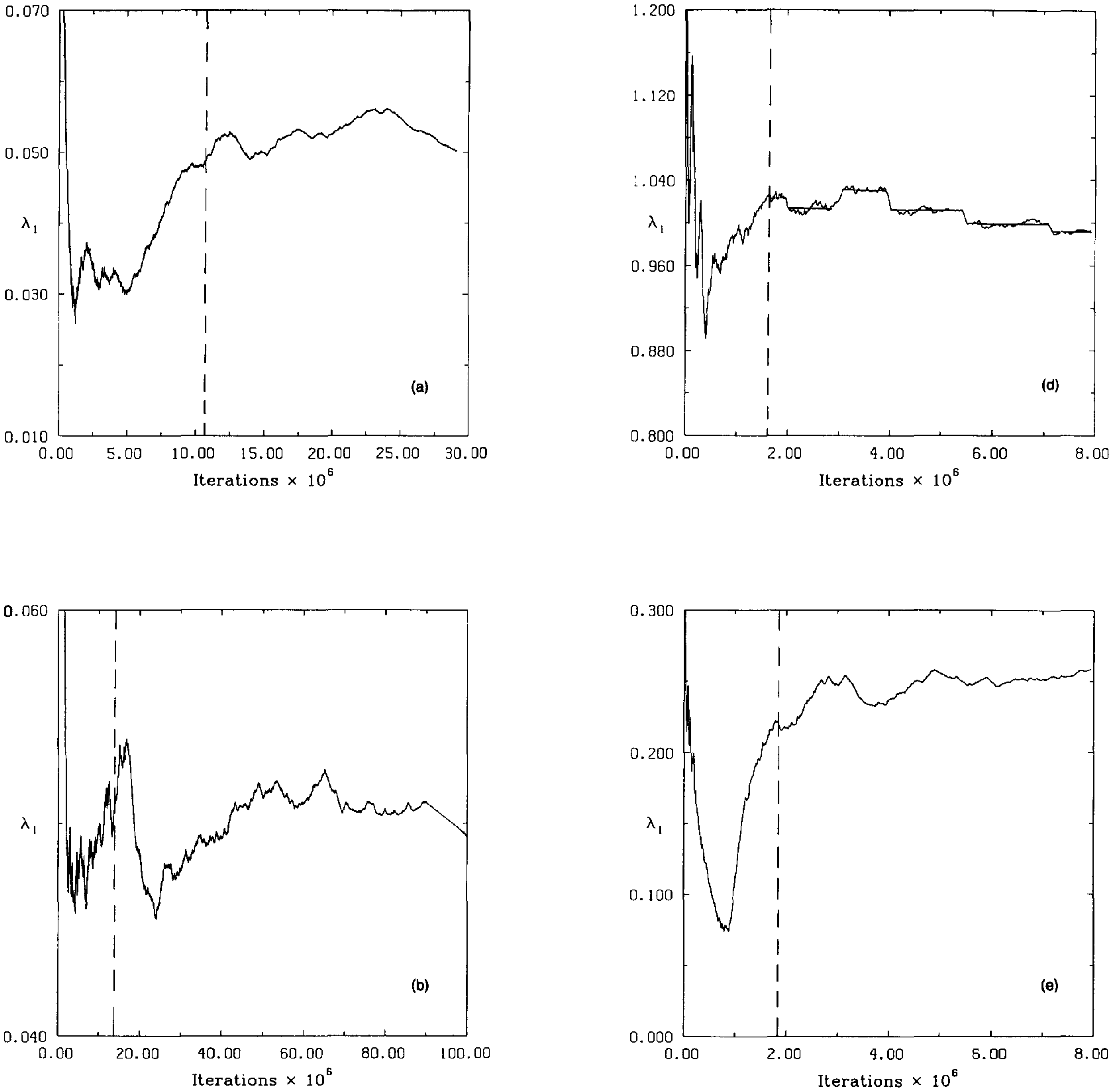
These oscillations decrease in magnitude as the number of points increases, as normal, and do not obscure the steplike nature of the profile.

Not all the chaotic orbits in quartic potentials behave as obviously as above. The maximal Lyapunov exponent for the potential U_2 is shown in Fig. 11(e). After the initial settling period it attains a value of about 0.252 and has no obvious long term oscillatory behavior. This indicates that there are either few cantori in the surface of section in Fig. 4 or the rate of diffusion through them is quite large. However, in general, cantori generated long time scale oscillations introduced non-numerical uncertainty into the Lyapunov spectrums for all the quartic polynomial potentials examined. This can only be overcome by calculating the spectrums for enormous lengths of time. The period and amplitude of these oscillations vary between orbits and systems. Typically the period of oscillation is of the order of thousands of orbital periods and the magnitude up to 20% of the average value. The resulting uncertainty will be given for all the maximal Lyapunov exponents calculated here.

III. MAXIMAL LYAPUNOV EXPONENTS AND SURFACES OF SECTION

The maximal Lyapunov exponent was calculated for the chaotic orbit filling the largest region in the surface of section, shown in Fig. 1, for the potential U_1. This two-dimensional region covers only a small proportion of the surface of section. The orbit is confined to a relatively thin band between two large regular regions. Figure 12 shows the actual chaotic orbit that generates this region in the surface of section. It looks quite regular. When a chaotic orbit is sandwiched between regular regions it is topologically very similar to the adjacent regular quasiperiodic orbits with a small chaotic perturbation superimposed on the motion. The wider the chaotic region the larger the magnitude of this perturbation. The maximal Lyapunov exponent essentially measures the size of this perturbation or the degree of chaotic behavior superimposed on the adjacent regular orbit. In this case the maximal Lyapunov exponent is approximately 0.05. This small value is consistent with the orbit appearing to be nearly regular and the chaotic region in the surface of section being quite narrow.

The large chaotic region in Fig. 2, for the potential U_2, dominates the surface of section, covering more than half of it. The corresponding maximal Lyapunov exponent is 0.252, indicating quite strong chaotic behavior. This is quantitatively consistent with its large coverage of the surface of section. The largest chaotic region in Fig. 5, for the potential V_2, covers a similar proportion of the surface of section and has a similar maximal Lyapunov exponent of 0.266, even though the surfaces of section are topologically very different. The potential V_1 possesses by far the largest chaotic region of all the potentials examined. It covers almost the entire surface of section, which is shown in Fig. 4. The corresponding maximal Lyapunov exponent is 0.996, indicating extremely strong chaotic behavior. This is by far the largest maximal Lyapunov exponent and is consistent with the orbit's extensive coverage of the surface of section.



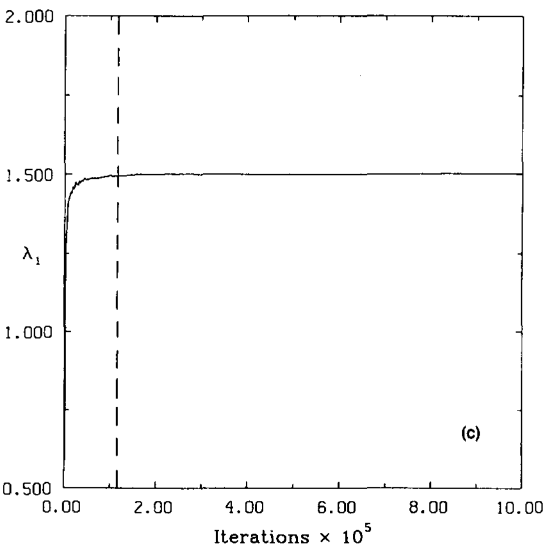


FIG. 11. (a) The maximal Lyapunov exponent for the potential U_1 is plotted against the number of iterations used in its calculation. The dotted line indicates the point at which the orbit has been sufficiently sampled so that $|\lambda_4 + \lambda_1|$, $|\lambda_2|$, and $|\lambda_3|$ are sufficiently small. (b) The maximal Lyapunov exponent for the same chaotic region as in (a) using a different set of initial conditions and for a larger number of iterations. (c) The maximal Lyapunov exponent for the Lorenz system with $\sigma = 16$, $R = 45.92$, and $b = 4$; (d) the maximal Lyapunov exponent for the potential V_1. The steplike behavior indicates diffusion from one cantori bounded region to another. (e) The maximal Lyapunov exponent for the potential U_2.

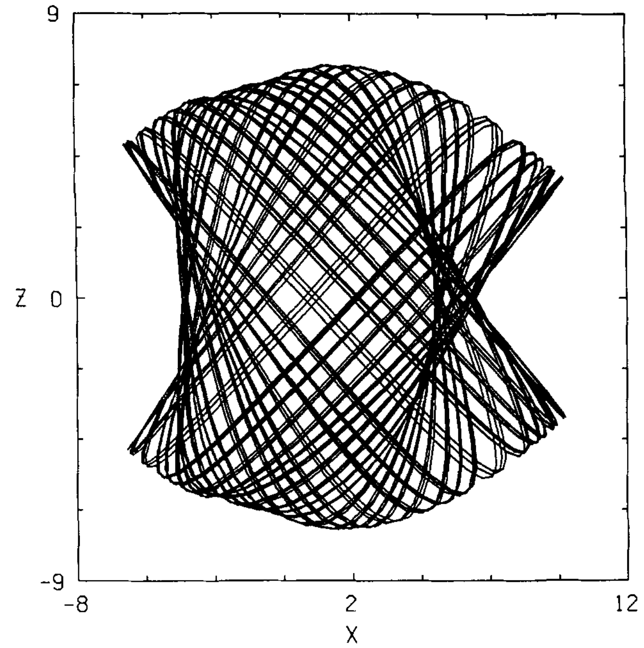


FIG. 12. The chaotic orbit corresponding to the largest chaotic band in the surface of section shown in Fig. 1.

The largest chaotic region in the surface of section of the potential V_3, shown in Fig. 6, is similar in size but different in shape to that of U_3 shown in Fig. 3. Both regions are relatively large but are confined between substantial regular regions. Their corresponding maximal Lyapunov exponents of 0.155 and 0.13 are comparable, considering their respective error bounds. The areas in the surface of section and the maximal Lyapunov exponents for this pair are markedly smaller than those of the previous pair of U_2 and V_2.

The largest chaotic orbit in U_3 has an enormous number of island chains embedded within it, far more than any other potential examined. Consequently this orbit contains a large number of cantori. The corresponding Lyapunov spectrum was particularly difficult to calculate and possessed many irregular oscillations causing much larger uncertainty than normal. The chaotic region in the surface of section, shown in Fig. 3, is not filled even approximately uniformly. There are clearly visible light and dark bands. These are examples of regions separated by cantori.

All the remaining potentials have maximal Lyapunov exponents smaller than 0.1. This is consistent with the observation that the chaotic regions in all the corresponding surface of section are very small and confined to narrow bands between the large dominant regular regions. All have smaller areas and smaller maximal Lyapunov exponents than any of the previously examined potentials.

The surfaces of section for the potentials V_5 and V_7 are shown in Figs. 8 and 10. The chaotic regions appear to be the same size. The corresponding maximal Lyapunov exponents are 0.082 and 0.081, respectively. The chaotic region centered on the origin in Fig. 9 is slightly smaller than those belonging to either of the previous pair. The maximal Lyapunov exponent 0.078 is also slightly smaller. The chaotic orbit surrounding the hyperbolic points in the surface of section in Fig. 7, for the potential V_4, is so small that it cannot be seen on this scale. It is the smallest chaotic orbit we examined and possessed the smallest Lyapunov exponent of 0.075.

For the potentials examined here there appears to be a direct correlation between the magnitude of the maximal Lyapunov exponent and the size of the corresponding chaotic region in the surface of section. The same ordering of the potentials is achieved if we order the potentials by either the size of their largest chaotic region or the magnitude of the corresponding maximal Lyapunov exponent. In this case the order was V_1, U_2 and V_2, U_3 and V_3, V_5 and V_7, V_6, V_4.

IV. CONCLUSION

For conservative two-dimensional systems the maximal Lyapunov exponent λ_1 provides a good qualitative and quantitative measure of chaos. Determining that $\lambda_1 > 0$ indicates that the corresponding orbit is chaotic. The magnitude of λ_1 then gives us a quantitative measure of just how chaotic it is. This allows us to compare different orbits and determine which orbit is more chaotic. There appears to be a direct correlation between the size of the chaotic region in the surface of section and the magnitude of the maximal Lyapunov exponent. Substantial chaotic regions dominating the entire surface of section had Lyapunov exponents in the range 0.25 to 1.0. Large chaotic areas confined between larger regular regions had maximal Lyapunov exponents in the range 0.1 to 0.25. Small bands of chaos confined to narrow regions between substantial regular regions typically had small maximal Lyapunov exponents of the order 0.01 to 0.10.

Chaotic orbits are divided into a number of adjacent regions possessing slightly different Lyapunov exponents by families of cantori. This causes very long time scale oscillations in the Lyapunov spectrums resulting in some uncertainty in the spectrums. This phenomenon is real and is not a numerical convergence problem. It does not appear to exist in a number of previously examined simpler dissipative systems such as the Lorenz system.

[1] V. I. Oseldec, Trans. Moscow Math. Soc. **19**, 197 (1968).
[2] J. D. Farmer, E. Ott, and J. A. Yorke, Physica D **7**, 153 (1983).
[3] G. Benettin, L. Galgani, and J.-M. Strelcyn, Phys. Rev. A **14**, 2338 (1976).
[4] I. Shimada and T. Nagashima, Prog. Theor. Phys. **61**, 1605 (1979).
[5] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, C. R. Acad. Sci. Ser. A **286**, 431 (1978).
[6] M. Sano and Y. Sawada, Phys. Rev. Lett. **55**, 1082 (1985).
[7] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica D **16**, 285 (1985).
[8] M. Hénon, Q. Appl. Math. **XXVII**, 291 (1969).
[9] O. E. Rössler, Phys. Lett. A **57**, 397 (1976).
[10] E. N. Lorenz, J. Atom. Sci. **20**, 130 (1983).
[11] F. Verhulst, Philos. Trans. R. Soc. London **290**, 435 (1979).
[12] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[13] A. Ramani, B. Dorizzi, and B. Grammaticos, Phys. Rev. Lett. **49**, 1539 (1982).
[14] J. D. Meiss and E. Ott, Physica D **20**, 387 (1986).
[15] R. S. Mackay, J. D. Meiss, and I. C. Percival, Physica D **13**, 55 (1984).

Large-N iterative solution of the Dirac equation

Satilmis Atag
Department of Physics, Faculty of Sciences, University of Ankara, Ankara, Turkey

(Received 11 March 1988; accepted for publication 9 November 1988)

The large-N iterative technique proposed recently by Chatterjee [J. Math. Phys. **27**, 2331 (1986)] is employed to solve the Dirac equation for a spherically symmetric potential. It is shown that the exact results are obtained in the case of the Coulomb potential.

I. INTRODUCTION

The large-N expansion method has been employed with remarkable success to deal with a variety of problems in different areas of physics.[1–4] In nonrelativistic quantum mechanics the method of large-N expansion represents an approach that gives rise to one of the most elegant analytic approximations for obtaining eigenvalues and eigenfunctions,[5,6] alternatively to ordinary perturbation theory. For spherically symmetric potentials $1/\kappa = 1/(N+2l)$ has been used as the expansion parameter in order to incorporate the finite-N corrections to the leading-order solution ($N\to\infty$) where N is the space dimension and l the angular momentum. In most cases the results appeared quite attractive.

In the relativistic case it was shown that the wave equations were solved exactly for the Coulomb potential in the N-space dimension.[7] Since the exact solutions to the relativistic wave equations cannot be obtained analytically for most potentials, one needs to have approximation schemes. Recently, the large-N iterative procedure has been proposed for the solution of the Klein–Gordon equation to obtain the energy eigenvalue of a spin-0 particle in a spherically symmetric potential $V(r)$.[8]

In this work we use a similar technique in the case of the Dirac equation. When the Coulomb potential is considered the method is shown to yield the exact series.[9]

II. LARGE-N ITERATION

Using the known factorization procedure in the N-space dimension,[10] the two pairs of first-order differential equations for the radial functions in polar coordinates are given as follows:

$$\begin{aligned}\frac{dF}{dr} &= \frac{K-\frac{1}{2}(N-3)}{r}F-(E-m-V(r))G,\\ \frac{dG}{dr} &= -\frac{K+\frac{1}{2}(N-3)}{r}G+(E+m-V(r))F.\end{aligned}\tag{1}$$

Here

$$K=-(l+\tfrac{1}{2}(N-1)),\quad j=l+\tfrac{1}{2},$$

and standard four-dimensional spinors have been used:

$$\Psi_{JM}=\frac{1}{r}\begin{bmatrix} iG\Phi^+\\ F\Phi^-\end{bmatrix},\quad j=l+\tfrac{1}{2},\quad j=l-\tfrac{1}{2}.\tag{2}$$

After some manipulation these equations can be transformed to one second-order differential equation

$$\frac{d^2g}{dr^2}-\frac{A'}{A}\frac{dg}{dr}+\left(-\frac{K(K+1)}{r^2}-\frac{A'}{A}\frac{K}{r}+AB\right)g=0,\tag{3}$$

with

$$A=E+m-V(r),\quad B=E-m-V(r),$$

$$A'=\frac{dA}{dr},\quad A''=\frac{d^2A}{dr^2}.$$

The first-order derivative term is eliminated by the transformation

$$U=\sqrt{A}\,g.\tag{4}$$

Then we obtain an eigenvalue equation with an energy-dependent effective potential

$$\frac{d^2U}{dr^2}+\left(\frac{A''}{2A}-\frac{3A'^2}{4A^2}-\frac{A'K}{Ar}-\frac{K(K+1)}{r^2}+AB\right)U=0.\tag{5}$$

Here, K may be expressed in terms of expansion parameter $k=N+2j$

$$K=-(k/2-1).\tag{6}$$

When k is large ($N\to\infty$), the leading-order approximation of the energy becomes

$$U\propto\delta(r-r_0),\tag{7}$$

$$E_\infty=V(r_0)+m(1+k^2/4m^2r_0^2)^{1/2},\tag{8}$$

where r_0 can be determined from minimization of the energy

$$r_0^3V'(r_0)(1+k^2/4m^2r_0^2)^{1/2}=k^2/4m.\tag{9}$$

The leading-order energy term (8) is the same as in the case of the Klein–Gordon equation except for the spin term in k.

Following the procedure in Ref. 8 we write Eq. (5) as

$$\begin{aligned}&\left(-\frac{1}{2m}\frac{d^2}{dr^2}+\frac{V''}{4mA_\infty}+\frac{3V'^2}{8mA_\infty^2}+\frac{V'(k-2)}{4mA_\infty r}\right.\\ &\left.\quad+\frac{(k-2)(k-4)}{8mr^2}+k^2f_1(r)\right)U^{(1)}(r)\\ &=\left(\frac{\widetilde{E}_1}{2m}\right)U^{(1)}(r),\end{aligned}\tag{10}$$

where

$$\widetilde{E}_1=(E^{(1)}-V(r_0))^2-(E_\infty-V(r_0))^2,\tag{11}$$

$$f_1(r)=\frac{(m^2-(E_\infty-V(r))^2)}{2mk^2},\tag{12}$$

 0022-2488/89/030696-04$02.50

and

$$A_\infty = E_\infty + m - V(r_0) . \quad (13)$$

It is convenient to shift the origin of coordinates to $r = r_0$ by defining

$$x = (\sqrt{k}/r_0)(r - r_0) . \quad (14)$$

Using (14) in Eq. (10) and expanding about $x = 0$ in powers of x gives an effective harmonic oscillator equation with a perturbation

$$\left(-\frac{1}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^{(1)2}x^2 + \epsilon_0^{(1)} + \widetilde{V}^{(1)}(x)\right)\phi^{(1)}(x) = \lambda^{(1)}\phi^{(1)}(x) , \quad (15)$$

where

$$\omega^{(1)} = (3/4m^2 + r_0^4 f_1''(r_0)/m)^{1/2} , \quad (16)$$

$$\lambda^{(1)} = (r_0^2/2mk)\widetilde{E}_1 , \quad (17)$$

$$\epsilon_0^{(1)} = \frac{k}{8m} + \frac{V_0' r_0}{4mA_\infty} - \frac{3}{4m} + \frac{1}{km} + r_0^2 f_1(r_0)k + \frac{r_0^2 V_0''}{4mkA_\infty} + \frac{3r_0^2 V_0'^2}{8mkA_\infty^2} - \frac{r_0 V_0'}{2mkA_\infty} , \quad (18)$$

$$\widetilde{V}^{(1)}(x) = (1/k^{1/2})(\epsilon_1^{(1)}x + \epsilon_3^{(1)}x^3) + (1/k)(\epsilon_2^{(1)}x^2 + \epsilon_4^{(1)}x^4) + (1/k^{3/2})(\delta_1^{(1)}x + \delta_3^{(1)}x^3 + \delta_5^{(1)}x^5) + (1/k^2)(\delta_2^{(1)}x^2 + \delta_4^{(1)}x^4 + \delta_6^{(1)}x^6) + \cdots , \quad (19)$$

$$V_0' = \left.\frac{dV}{dr}\right|_{r_0} .$$

The definitions of $\epsilon_j^{(1)}$ and $\delta_j^{(1)}$ are given in the Appendix.

The energy eigenvalue $E^{(1)}$ can be obtained using the standard formalism of the fourth-order perturbation theory in powers of $1/k$ (Ref. 6)

$$\begin{aligned}
E^{(1)} = V(r_0) + m\Bigg[&\frac{1}{m^2}(E_\infty - V(r_0))^2 + \frac{2}{m}\Bigg\{\frac{k}{r_0^2}\left(\epsilon_0^{(1)} + \left(n + \frac{1}{2}\right)\omega^{(1)}\right) \\
&+ \Bigg(\frac{(1+2n)}{r_0^2}\bar{\epsilon}_2^{(1)} + \frac{3(1+2n+2n^2)}{r_0^2}\bar{\epsilon}_4^{(1)} \\
&- \frac{1}{\omega^{(1)}r_0^2}\left[\bar{\epsilon}_1^{(1)2} + 6(1+2n)\bar{\epsilon}_1^{(1)}\bar{\epsilon}_3^{(1)} + (11+30n+30n^2)\bar{\epsilon}_3^{(1)2}\right]\Bigg) \\
&+ \frac{1}{k}\Bigg(\frac{1}{r_0^2}\left[(1+2n)\bar{\delta}_2^{(1)} + 3(1+2n+2n^2)\bar{\delta}_4^{(1)} + 5(3+8n+6n^2+4n^3)\bar{\delta}_6^{(1)}\right] \\
&- \frac{1}{\omega^{(1)}r_0^2}\big[(1+2n)\bar{\epsilon}_2^{(1)2} + 12(1+2n+2n^2)\bar{\epsilon}_2^{(1)}\bar{\epsilon}_4^{(1)} \\
&+ 2(21+59n+51n^2+34n^3)\bar{\epsilon}_4^{(1)2} + 2\bar{\epsilon}_1^{(1)}\bar{\delta}_1^{(1)} + 6(1+2n)\bar{\epsilon}_1^{(1)}\bar{\delta}_3^{(1)} \\
&+ 30(1+2n+2n^2)\bar{\epsilon}_1^{(1)}\bar{\delta}_5^{(1)} + 6(1+2n)\bar{\epsilon}_3^{(1)}\bar{\delta}_1^{(1)} \\
&+ 2(11+30n+30n^2)\bar{\epsilon}_3^{(1)}\bar{\delta}_3^{(1)} + 10(13+40n+42n^2+28n^3)\bar{\epsilon}_3^{(1)}\bar{\delta}_5^{(1)}\big] \\
&+ \frac{1}{\omega^{(1)2}r_0^2}\big[4\bar{\epsilon}_1^{(1)2}\bar{\epsilon}_2^{(1)} + 36(1+2n)\bar{\epsilon}_1^{(1)}\bar{\epsilon}_2^{(1)}\bar{\epsilon}_3^{(1)} \\
&+ 8(11+30n+30n^2)\bar{\epsilon}_2^{(1)}\bar{\epsilon}_3^{(1)2} + 24(1+2n)\bar{\epsilon}_1^{(1)2}\bar{\epsilon}_4^{(1)} \\
&+ 8(31+78n+78n^2)\bar{\epsilon}_1^{(1)}\bar{\epsilon}_3^{(1)}\bar{\epsilon}_4^{(1)} + 12(57+189n+225n^2+150n^3)\bar{\epsilon}_3^{(1)2}\bar{\epsilon}_4^{(1)}\big] \\
&- \frac{1}{\omega^{(1)3}r_0^2}\big[8\bar{\epsilon}_1^{(1)3}\bar{\epsilon}_3^{(1)} + 108(1+2n)\bar{\epsilon}_1^{(1)2}\bar{\epsilon}_3^{(1)2} \\
&+ 48(11+30n+30n^2)\bar{\epsilon}_1^{(1)}\bar{\epsilon}_3^{(1)3} + 30(31+109n+141n^2+94n^3)\bar{\epsilon}_3^{(1)4}\big]\Bigg) + \cdots \Bigg\}\Bigg]^{1/2} ,
\end{aligned} \quad (20)$$

where

$$\bar{\epsilon}_j^{(1)} = \epsilon_j^{(1)}/[2m\omega^{(1)}]^{1/2}, \qquad \bar{\delta}_j^{(1)} = \delta_j^{(1)}/[2m\omega^{(1)}]^{1/2}. \tag{21}$$

The next step is to get $E^{(2)}$ in an iterative manner by writing the eigenvalue equation in the following way:

$$\left(-\frac{1}{2m}\frac{d^2}{dr^2} + \frac{V''}{4mA_\infty} + \frac{3V'^2}{8mA_\infty^2} + \frac{V'(k-2)}{4mA_\infty r} + \frac{(k-2)(k-4)}{8mr^2} + k^2 f_2(r)\right)U^{(2)}(r) = \left(\frac{\widetilde{E}_2}{2m}\right)U^{(2)}(r) \tag{22}$$

with the definitions

$$f_2(r) = \frac{m^2 - (E_\infty - V(r))^2}{2mk^2} + \frac{(E^{(1)} - E_\infty)(V(r) - V(r_0))}{mk^2} \tag{23}$$

and

$$\widetilde{E}_2 = (E^{(2)} - V(r_0))^2 - (E_\infty - V(r))^2. \tag{24}$$

Repeating the above procedure yields

$$\left(-\frac{1}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^{(2)2}x^2 + \epsilon_0^{(2)} + \widetilde{V}^{(2)}(x)\right)\phi^{(2)}(x) = \lambda^{(2)}\phi^{(2)}(x), \tag{25}$$

where

$$\omega^{(2)} = (3/4m^2 + r_0^4 f_2''(r_0)/m)^{1/2}, \tag{26}$$

$$\lambda^{(2)} = (r_0^2/2mk)\widetilde{E}_2, \tag{27}$$

$$\epsilon_0^{(2)} = \frac{k}{8m} + \frac{r_0V_0'}{4mA_\infty} - \frac{3}{4m} + \frac{1}{km} + r_0^2 k f_2(r_0) + \frac{r_0^2 V_0''}{4mkA_\infty} + \frac{3r_0^2 V_0'^2}{8mkA_\infty^2} - \frac{r_0V_0'}{2mkA_\infty}, \tag{28}$$

$$\widetilde{V}^{(2)}(x) = (1/k^{1/2})(\epsilon_1^{(2)}x + \epsilon_3^{(2)}x^3) + (1/k)(\epsilon_2^{(2)}x^2 + \epsilon_4^{(2)}x^4) + (1/k^{3/2}) \times (\delta_1^{(2)}x + \delta_3^{(2)}x^3 + \delta_5^{(2)}x^5) + (1/k^2)(\delta_2^{(2)}x^2 + \delta_4^{(2)}x^4 + \delta_6^{(2)}x^6) + \cdots. \tag{29}$$

At this stage the fourth-order perturbation theory is applied to Eq. (25) once more. One can continue the iteration in a similar way for more improved energy solutions. Although the procedure can easily be applied to any spherically symmetric potentials, the results are given for the Coulomb potential $V(r) = -\beta/r$ in order to see how it works:

$$E^{(1)} = m\left\{1 - \frac{2\beta^2}{k^2}\left(1 + \frac{4}{k} + \frac{12}{k^2} + \cdots\right) - \frac{2\beta^4}{k^4}\left(1 + \frac{8}{k} + \frac{104}{k^2} + \cdots\right)\right\}, \tag{30}$$

$$E^{(2)} = m\left\{1 - \frac{2\beta^2}{k^2}\left(1 + \frac{4}{k} + \frac{12}{k^2} + \cdots\right) - \frac{2\beta^4}{k^4}\left(1 + \frac{8}{k} + \frac{40}{k^2} + \cdots\right)\right\}. \tag{31}$$

The agreement between the second iteration $E^{(2)}$ and the exact analytic $1/N$ series is perfect for the lowest-order relativistic correction to the ground state.[9]

III. CONCLUSION

A study has been made of the large-N iterative solution for the relativistic spin-$\frac{1}{2}$ particles based on the method developed for the solution of the Klein–Gordon equation. One may conclude that the large-N iterative energy solution of the Dirac equation leads to the exact agreement with the analytic series in the case of the Coulomb potential. The procedure can be applicable to the other spherically symmetric potentials of physical interest, such as an electron with anomalous magnetic moment interacting with an external Coulomb field, which is the possible explanation of the resonance during heavy ion collisions. Furthermore, the method may be useful in the charge-magnetic monopole solution of the Dirac equation when it is possible to write the equation in arbitrary dimension.

ACKNOWLEDGMENT

I wish to thank Mr. Harun Artuner for his kind help.

APPENDIX: DEFINITIONS OF ϵ_j AND δ_j

Here we give the definitions of ϵ_j and δ_j:

$$\epsilon_1^{(1)} = \left(\frac{3}{2m} + \frac{r_0^2V_0''}{4mA_\infty} + \frac{r_0^2V_0'^2}{4mA_\infty^2} - \frac{r_0V_0'}{4mA_\infty}\right), \tag{A1}$$

$$\epsilon_2^{(1)} = \left(-\frac{9}{4m} + \frac{r_0^3V_0'''}{8mA_\infty} + \frac{3r_0^3V_0''V_0'}{8mA_\infty^2} - \frac{r_0^2V_0''}{4mA_\infty} + \frac{r_0^3V_0'^3}{4mA_\infty^3} - \frac{r_0^2V_0'^2}{4mA_\infty^2} + \frac{r_0V_0'}{4mA_\infty}\right), \tag{A2}$$

$$\epsilon_3^{(1)} = \left(-\frac{1}{2m} + \frac{1}{6}r_0^5 f_1'''(r_0)\right), \tag{A3}$$

$$\epsilon_4^{(1)} = \left(\frac{5}{8m} + \frac{r_0^6 f_1''''(r_0)}{24}\right), \tag{A4}$$

$$\delta_1^{(1)} = \left(-\frac{2}{m} + \frac{r_0^3V_0'''}{4mA_\infty} + \frac{r_0^3V_0'V_0''}{mA_\infty^2} + \frac{3r_0^3V_0'^3}{4mA_\infty^3} - \frac{r_0^2V_0''}{2mA_\infty} - \frac{r_0^2V_0'^2}{2mA_\infty^2} + \frac{r_0V_0'}{2mA_\infty}\right), \tag{A5}$$

$$\delta_2^{(1)}=\left(\frac{3}{m}+\frac{r_0^4V_0''''}{8mA_\infty}+\frac{5r_0^4V_0'V_0'''}{8mA_\infty^2}+\frac{r_0^4V_0''^2}{2mA_\infty^2}+\frac{17r_0^4V_0'^2V_0''}{8mA_\infty^3}+\frac{9r_0^4V_0'^4}{8mA_\infty^4}-\frac{r_0^3V_0'''}{4mA_\infty}-\frac{3r_0^3V_0'V_0''}{4mA_\infty^2}+\frac{r_0^2V_0''}{2mA_\infty}-\frac{r_0^3V_0'^3}{2mA_\infty^3}+\frac{r_0^2V_0'^2}{2mA_\infty^2}-\frac{r_0V_0'}{2mA_\infty}\right), \quad \text{(A6)}$$

$$\delta_3^{(1)}=\left(\frac{3}{m}+\frac{r_0^4V_0''''}{24mA_\infty}+\frac{r_0^4V_0'V_0'''}{6mA_\infty^2}-\frac{r_0^3V_0'''}{8mA_\infty}+\frac{r_0^4V_0''^2}{8mA_\infty^2}+\frac{r_0^4V_0''V_0'^2}{2mA_\infty^3}-\frac{3r_0^3V_0'V_0''}{8mA_\infty^2}+\frac{r_0^2V_0''}{4mA_\infty}+\frac{r_0^4V_0'^4}{4mA_\infty^4}-\frac{r_0^3V_0'^3}{4mA_\infty^3}+\frac{r_0^2V_0'^2}{4mA_\infty^2}-\frac{r_0V_0'}{4mA_\infty}\right), \quad \text{(A7)}$$

$$\delta_4^{(1)}=\left(-\frac{15}{4m}+\frac{r_0^5V_0'''''}{96mA_\infty}+\frac{5r_0^5V_0'V_0''''}{96mA_\infty^2}-\frac{r_0^4V_0''''}{24mA_\infty}+\frac{5r_0^5V_0''V_0'''}{48mA_\infty^2}+\frac{5r_0^5V_0'''V_0'^2}{24mA_\infty^3}-\frac{r_0^4V_0'V_0'''}{6mA_\infty^2}+\frac{r_0^3V_0'''}{8mA_\infty}+\frac{5r_0^5V_0'V_0''^2}{16mA_\infty^3}-\frac{r_0^4V_0''^2}{8mA_\infty^2}+\frac{5r_0^5V_0''V_0'^3}{8mA_\infty^4}-\frac{r_0^4V_0'^2V_0''}{2mA_\infty^3}+\frac{3r_0^3V_0'V_0''}{8mA_\infty^2}-\frac{r_0^2V_0''}{4mA_\infty}+\frac{r_0^5V_0'^5}{4mA_\infty^5}-\frac{r_0^4V_0'^4}{4mA_\infty^4}+\frac{r_0^3V_0'^3}{4mA_\infty^3}-\frac{r_0^2V_0'^2}{4mA_\infty^2}+\frac{r_0V_0'}{4mA_\infty}\right), \quad \text{(A8)}$$

$$\delta_5^{(1)}=\left(-\frac{3}{4m}+\frac{r_0^7f_1'''''(r_0)}{120}\right), \quad \text{(A9)}$$

$$\delta_6^{(1)}=\left(\frac{7}{8m}+\frac{r_0^8f_1''''''(r_0)}{720}\right), \quad \text{(A10)}$$

$$\epsilon_1^{(2)}=\left(\frac{3}{2m}+\frac{r_0^2V_0''}{4mA_\infty}+\frac{r_0^2V_0'^2}{4mA_\infty^2}-\frac{r_0V_0'}{4mA_\infty}+\frac{r_0^3V_0'(E^{(1)}-E_\infty)}{mk}\right). \quad \text{(A11)}$$

Other coefficients have the previous values with the replacement $f_1 \to f_2$.

[1] E. Witten, Nucl. Phys. B **160**, 57 (1979).
[2] S. Coleman, SLAC preprint Pub-2484, March 1980.
[3] S. Ma, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 6.
[4] S. Hikami and E. Brezin, J. Phys. A **12**, 759 (1979).
[5] L. Mlodinow and M. Shatz, J. Math. Phys. **25**, 943 (1984).
[6] T. Imbo, A. Pagnamenta, and U. Sukhatme, Phys. Rev. D **29**, 1669 (1984).
[7] M. M. Neito, Am. J. Phys. **47**, 1067 (1979).
[8] A. Chatterjee, J. Math. Phys. **27**, 2331 (1986).
[9] J. L. Miramontes and C. Pajares, Nuovo Cimento B **84**, 10 (1984).
[10] A. Joseph, Rev. Mod. Phys. **39**, 829 (1967); C. A. Coulson and A. Joseph, *ibid.* **39**, 838 (1967).

On the compatibility of relativistic wave equations in Riemann–Cartan spaces

H. A. Buchdahl[a)]
Institut für Theoretische Physik der Universität, D3400-Göttingen, Bunsenstrasse 9, West Germany

(Received 12 April 1988; accepted for publication 26 October 1988)

The so-called compatibility problem for relativistic wave equations in Riemann spaces V_4 has its counterpart in Riemann–Cartan spaces U_4. What is to be understood by incompatibility is elucidated. It is shown that in a U_4, the minimally coupled massive spin-S Dirac equations imply constraints for all values of $S > \frac{1}{2}$. In a U_4 (though not in a V_4), one has a constraint already when $S = 1$. It is shown that the equations can be compatible (when $S \geqslant 1$) only if torsion is absent, i.e., if the U_4 is in fact a V_4. If one contemplates the equations on the level of first quantization, the minimal coupling prescription is inappropriate, and minimal Hermitian coupling is prescribed in its place. However, the previous conclusion, that for $S \geqslant 1$ the equations are always incompatible in the presence of torsion, remains valid.

I. INTRODUCTION

In an earlier sequence of papers[1–6] (hereafter referred to as V0, V1, V2, V3, V4, and V5, respectively), I investigated the compatibility of pairs of relativistic first-order spin-S $(= :\frac{1}{2}n)$ nonzero rest mass equations in Riemannian space-times V_4 of signature -2. Since a corresponding investigation is now to be undertaken in a Riemann–Cartan space-time U_4, it is appropriate to set out clearly the original problem, and in particular to clarify how the term "compatibility" is to be understood. In broad terms, one is concerned with the question of whether the curvature of the V_4 will, in general, imply constraints upon the field spinors ξ,η; that is, constraints which are absent in flat space. It is necessary to be more explicit, as follows. The spinors ξ and η must satisfy two coupled first-order equations [Eqs. V3(1.1)]. These imply, by elimination, not only that ξ and η (each of which is required to be symmetric in its dotted and undotted indices) must separately satisfy second-order wave equations, but also that, in general, at least one of them—let it be ξ—satisfies a certain algebraic equation, homogeneous and linear in its components: the "(equation of) constraint." Now in flat space, the manifold of solutions corresponds to the choice of all possible sets of Cauchy data, any particular solution (valid in a neighborhood of some chosen spacelike hypersurface $\mathscr{H}$) being fixed by freely specified values, on $\mathscr{H}$, of the components of ξ and of the derivatives of these with respect to x^4. The same state of affairs would obtain in a (nonflat) V_4 were it not for the presence of constraints, for the latter imply restrictions on the freedom of choice of the Cauchy data, in as far as the values on $\mathscr{H}$ of the components of ξ can no longer be freely prescribed once the values of their x^4 derivatives have been chosen. The requirement that in a general V_4 one should have the same freedom in the choice of the Cauchy data as one would have if the V_4 were flat is thus inconsistent with the form of the given pair of equations for ξ and η. It is just this conclusion which is often translated into the ill-chosen, though convenient, phrase "in a general V_4 the equations [i.e., V3(1.1)] are incompatible," for this suggests that there is some internal contradiction between them. There is not; the contradiction is between the equations and an additional prescribed condition concerning the Cauchy data.

Two concomitant questions now arise. The first is this: can the V_4 be so chosen that the constraints become nugatory? In V2 and V3, it was shown that for arbitrary S and arbitrary representation (see Sec. I of V4), compatibility requires the V_4 to be an S_4, i.e., to have constant Riemannian curvature. One has weaker conditions in only two cases: (i) when $S = \frac{3}{2}$ and the representation is $\mathscr{D}_1$, the V_4 needs to be merely an Einstein space, or (ii) for any value of S, when the representation is $\mathscr{D}_0$, the V_4 must be conformally flat. These results almost inevitably lead to the second question: can the field equations be so modified by the inclusion of additional terms which vanish in flat space that the compatibility of the resulting nonminimally coupled equations is assured? In V2 and V3, the construction of such equations is achieved for the special representation $\mathscr{D}_0$ and $\mathscr{D}_1$, but for arbitrary representations I have hitherto found the problem intractable. Finally there is the additional problem of finding Lagrangians which generate the compatible (nonminimally coupled) field equations, granted that the subsidiary conditions are imposed after variation. This is solved in V4 in two special cases, namely for $S = \frac{3}{2}$ and $S = 2$ when the representation is $\mathscr{D}_1$.

One generalization of general relativity theory which has been widely investigated is U_4 theory. (For a comprehensive survey of the latter, see Ref. 7.) The geometrical framework underlying U_4 theory is a Riemann–Cartan space-time U_4, that is, a four-dimensional linearly connected space on which a covariant constant metric tensor of signature -2 is defined. This definition differs from that of a Riemannian space-time only in that the skew-symmetric part $\Gamma_{[ij]}{}^k = :S_{ij}{}^k$ of the linear connection is not required to vanish. The meaning of minimal coupling remains as before: partial derivatives in the flat-space equations are replaced by covariant derivatives. This prescription entails that the curvature tensor does not appear explicitly in the (first-order)

[a)] Permanent address: Department of Physics and Theoretical Physics, Faculty of Science, Australian National University, GPO Box 4, Canberra, A.C.T. 2601, Australia.

 0022-2488/89/030700-06$02.50

equations at all, and the torsion tensor $S_{ij}{}^{k}$ occurs in them only through the $\Gamma_{ij}{}^{k}$ present in covariant derivatives.

Questions analogous to those examined in V0–V4 of necessity arise also in U_4 theory. The compatibility problem as such has been considered by Barth and Christensen,[8] who quote a number of other references. However, all these, except Ref. 8 itself, lack generality in as far as only special, low values of S are contemplated, or else they deal only with the case of zero rest mass. On the other hand, in Ref. 8, eight distinct criteria are adopted towards the development of massive spin-S equations, the compatibility of which is then examined; but they are taken, *ab initio*, to be second-order equations. Moreover, the boson equations differ generically from the fermion equations. (Equations which differ only in the valences of the field spinors that enter into them are not regarded here as generally distinct.) By contrast, the present paper (which presupposes knowledge of the notation of V0–V4) makes a beginning with an investigation in the context of U_4 theory of problems outlined above in the Riemannian context, the equations to be examined being appropriate modifications of the (first-order) flat-space massive spin-S Dirac equations. The boson and fermion equations are, moreover, not generically distinct.

A rough outline of the work that now follows is this: in Sec. II, a number of definitions of, and relations between, various tensors in U_4 are given. Although they are, generally speaking, to be found—widely scattered—in the literature, this is done partly as a matter of convenience and partly to fix the notation. Section III contains an analogous collection of definitions of spinorial quantities, together with a set of relations involving them, which generalizes to a U_4 a corresponding set in a V_4, given previously.[9] In Sec. IV, the existence of constraints implicit in the minimally coupled massive spin-S Dirac equations is deduced. Whereas in a V_4 there are no constraints for $S<\frac{3}{2}$, there is in a U_4 already a constraint when $S=1$. In Secs. V and VI it is then shown that, for all values of $S>\frac{1}{2}$, the equations under consideration cannot be compatible unless torsion is absent, i.e., when the U_4 is a V_4. (Such a U_4 will hereafter be called "trivial.") If one contemplates the equations on the level of first quantization, the minimal coupling prescription is inappropriate. In Secs. VII and VIII, minimal Hermitian coupling is therefore adopted in its place; but, as before, compatibility of the equations requires the absence of torsion.

II. THE U_4: TENSOR RELATIONS

A Riemann–Cartan space-time is characterized by a metric-compatible asymmetric linear connection $\Gamma_{kl}{}^{m}$, i.e., it is a linearly connected space endowed with a (symmetric) metric tensor g_{ij}, such that

$$g_{ij;k}=0\,, \tag{2.1}$$

granted that, with

$$\Gamma_{[kl]}{}^{m}=:S_{kl}{}^{m}\,, \tag{2.2}$$

the "torsion" $S_{kl}{}^{m}$ is, in general, nonzero. Indices following a semicolon denote covariant derivation with respect to $\Gamma_{kl}{}^{m}$, e.g.,

$$t_k{}^l{}_{;m}=t_k{}^l{}_{,m}-\Gamma_{km}{}^{n}t_n{}^l+\Gamma_{nm}{}^{l}t_k{}^n\,, \tag{2.3}$$

as usual. If $\underset{0}{\Gamma}_{kl}{}^{m}$ are the Christoffel symbols which belong to g_{ij}, i.e.,

$$\underset{0}{\Gamma}_{kl}{}^{m}:=g^{mn}(g_{n(k,l)}-\tfrac{1}{2}g_{kl,n})\,, \tag{2.4}$$

one finds from (2.1) that

$$\Gamma_{kl}{}^{m}=\underset{0}{\Gamma}_{kl}{}^{m}+L_{kl}{}^{m}\,, \tag{2.5a}$$

where

$$L_{kl}{}^{m}:=S_{kl}{}^{m}-S^{m}{}_{kl}-S_{lk}{}^{m}\,, \tag{2.5b}$$

is the "contorsion tensor." Note in passing that

$$L_{(k|l|m)}=0\,, \tag{2.6a}$$

$$L_{[kl]}{}^{m}=S_{kl}{}^{m}\,, \tag{2.6b}$$

$$L_{k[lm]}=S_{lmk}\,. \tag{2.6c}$$

As suggested by the notation of (2.5a), a subscript 0 under a kernel symbol shall indicate that the quantity in question belongs to the V_4 whose metric tensor is g_{ij}.

By straightforward calculation one finds that, if t_i is a vector,

$$2t_{i;[jk]}=R^{m}{}_{ijk}t_m-2S_{jk}{}^{m}t_{i;m}\,, \tag{2.7}$$

where

$$R^{m}{}_{ijk}:=\Gamma_{ik}{}^{m}{}_{,j}-\Gamma_{ij}{}^{m}{}_{,k}+\Gamma_{ik}{}^{n}\Gamma_{nj}{}^{m}-\Gamma_{ij}{}^{n}\Gamma_{nk}{}^{m}\,, \tag{2.8}$$

is the Riemann tensor of the U_4. More generally, if X is any tensor (or spinor), $X_{;[kl]}$ consists of a sum of terms familiar from Riemannian geometry, together with a further, additive term $-S_{kl}{}^{m}X_{;m}$. I write

$$X_{;[kl]}=(X_{;[kl]})^{*}-S_{kl}{}^{m}X_{;m}\,. \tag{2.9}$$

(Note that the curvature tensor contained in the first term on the right is, of course, that of the U_4.) The Ricci tensor is

$$R_{ij}:=R^{m}{}_{ijm}\,, \tag{2.10}$$

and it is, in general, nonsymmetric,

$$R_{[ij]}=2S_{[i;j]}-S_{ij}{}^{m}{}_{;m}+2S_{ij}{}^{m}S_m\,, \tag{2.11}$$

where

$$S_i:=S_{ij}{}^{j}\,. \tag{2.12}$$

A U_4 with zero $R_{[ij]}$ will be called "Ricci-symmetric." The scalar curvature is $R:=g^{ij}R_{ij}$ as usual, while the trace-free part of R_{ij} is

$$E_{ij}:=R_{ij}-\tfrac{1}{4}g_{ij}R\,. \tag{2.13}$$

R_{mijk} is of course skew in j and k, but it is also skew in m and i:

$$R_{mi(jk)}=0\,, \tag{2.14a}$$

$$R_{(mi)jk}=0\,. \tag{2.14b}$$

In place of the cyclic identity $\underset{0}{R}{}^{m}{}_{[ijk]}=0$ one has here,

$$R^{m}{}_{[ijk]}=-2S_{[ij}{}^{m}{}_{;k]}+4S_{[ij}{}^{n}S_{k]n}{}^{m}=:z^{m}{}_{ijk}\,. \tag{2.15}$$

Finally, while

$$p_{klmn}:=R_{klmn}-R_{mnkl} \tag{2.16a}$$

vanishes in a V_4, one has here,

$$p_{klmn} = \tfrac{3}{2}(z_{klmn} + z_{lknm} + z_{mknl} + z_{nklm}) . \tag{2.16b}$$

Finally one has the differential identity,

$$R^{m}{}_{i[jk;l]} = -2S_{[jk}{}^{n}R^{m}{}_{|i|l]n} . \tag{2.17}$$

From this one obtains further relations by contraction; e.g., contraction over l,m, followed by transvection with g^{ij}, leads to the identity,

$$R^{j}{}_{k;j} - \tfrac{1}{2}R_{;k} = -2S_{kab}R^{ab} + S^{abc}R_{abck} . \tag{2.18}$$

Occasionally it is useful to express $R^{m}{}_{ijk}$ in terms of $\underset{0}{R}{}^{m}{}_{ijk}$,

$$R^{m}{}_{ijk} = \underset{0}{R}{}^{m}{}_{ijk} + 2L_{i[k}{}^{m}{}_{|j]} + 2L_{i[k}{}^{n}L_{|n|j]}{}^{m} , \tag{2.19}$$

indices following a bar indicating covariant derivatives with respect to $\underset{0}{\Gamma}{}_{ij}{}^{k}$.

In Secs. VI and VII of Ref. 9 (also referred to as S) there occurs a number of relations which involve the conformal curvature tensor. There is a formal analog of this here, namely,

$$C_{klmn} := R_{klmn} + \tfrac{1}{2}(g_{km}R_{ln} - g_{kn}R_{lm} - g_{lm}R_{kn} + g_{ln}R_{km}) - \tfrac{1}{6}(g_{km}g_{ln} - g_{kn}g_{lm})R . \tag{2.20}$$

Like the conformal curvature tensor, it is trace-free, but its geometrical significance is obscure.

The torsion may not be general: (i) when there is a vector $\check{S}_m$ such that

$$S_{ijk} = e_{ijkm}\check{S}^{m} , \tag{2.21}$$

one speaks of "axial torsion"; and (ii) when there is a vector S_i such that

$$S_{ij}{}^{k} = \tfrac{2}{3}\delta^{k}{}_{[j}S_{i]} , \tag{2.22}$$

one speaks of "vector torsion." Even more specialized is the case of vector torsion when S_i is a gradient, i.e., there is a scalar S such that

$$S_i = S_{,i} \tag{2.23}$$

and then I shall speak of "scalar torsion." (This occurs, for instance, in Hojman *et al.*[10]) Scalar torsion implies symmetry of the Ricci tensor. If one requires R_{klmn} to have the form

$$R_{klmn} = \lambda(g_{kn}g_{ml} - g_{km}g_{nl}) \tag{2.24}$$

one finds, by using (2.17), that the torsion must be scalar, with

$$S_k = \tfrac{3}{4}(\ln\lambda)_{,k} . \tag{2.25}$$

One can also contemplate the possibility of $\check{S}_m$ being a gradient. This case appears to be less interesting, though here again R_{kl} is symmetric.

III. THE U_4: SPINOR RELATIONS

The spinor calculus associated with a U_4 follows the calculus of Infeld and van der Waerden[11] quite closely. A two-dimensional complex vector space S_2—spin space—is attached to each point of the U_4. The elements of S_2 are "(two-) spinors." Transformations of S_2 are elements of GL(2,C), and in general they are carried out independently of coordinate transformations in U_4. (Under the latter alone spinors are therefore invariants.)

A linear spinor connexion $\Gamma^{\alpha}{}_{\beta k}$ serves to define covariant differentiation of spinors; e.g., if ξ_μ is a spin vector,

$$\xi_{\mu;k} := \xi_{\mu,k} - \Gamma^{\alpha}{}_{\mu k}\xi_{\alpha} . \tag{3.1}$$

It is fixed by the prescription that the Hermitian connecting symbol ("Pauli matrices") $\sigma_{k\dot{\mu}\nu}$ be covariant constant,

$$\sigma_{k\dot{\mu}\nu;l} = \sigma_{k\dot{\mu}\nu,l} - \Gamma_{kl}{}^{m}\sigma_{m\dot{\mu}\nu} - \Gamma^{\dot{\alpha}}{}_{\dot{\mu}l}\sigma_{k\dot{\alpha}\nu} - \Gamma^{\beta}{}_{\nu l}\sigma_{k\dot{\mu}\beta} \overset{*}{=} 0 . \tag{3.2}$$

(The symbol $\overset{*}{=}$ is to be read as "is required to be." Its use will prove beneficial later on.)

The skew-symmetric metric spinor $\gamma_{\mu\nu}$ is also required to be covariant constant:

$$\gamma_{\mu\nu;k} = \gamma_{\mu\nu,k} - \Gamma^{\lambda}{}_{\lambda k}\gamma_{\mu\nu} \overset{*}{=} 0 , \tag{3.3}$$

whence it follows by transvection with $\gamma^{\mu\nu}$ that $\Gamma^{\lambda}{}_{\lambda k}$ is a gradient:

$$\Gamma^{\lambda}{}_{\lambda k} = \tfrac{1}{2}(\ln\det\gamma_{\alpha\beta})_{,k} =: \tfrac{1}{2}\lambda_{,k} .$$

Transvecting now (3.2) with $\sigma^{k\dot{\mu}\lambda}$ and taking (3.3) into account, there comes an explicit expression for the linear connection:

$$\Gamma^{\mu}{}_{\nu l} = \tfrac{1}{2}(\sigma^{m\dot{\rho}\mu}\sigma_{m\dot{\rho}\nu,l} - \Gamma_{kl}{}^{m}\sigma^{k\dot{\rho}\mu}\sigma_{m\dot{\rho}\nu} - \tfrac{1}{2}\delta^{\mu}{}_{\nu}\bar{\lambda}_{,l}) , \tag{3.4}$$

a bar denoting complex conjugation. The counterpart to (2.5a) is

$$\Gamma^{\mu}{}_{\nu k} - \underset{0}{\Gamma}{}^{\mu}{}_{\nu k} =: \Lambda^{\mu}{}_{\nu k} = -\tfrac{1}{2}S^{mn\mu}{}_{\nu}L_{mkn} . \tag{3.5}$$

The analog of (2.7) is the relation

$$2\xi_{\mu;[kl]} = P^{\lambda}{}_{\mu kl}\xi_{\lambda} - 2S_{kl}{}^{m}\xi_{\mu;m} , \tag{3.6}$$

where

$$P^{\lambda}{}_{\mu kl} := 2\Gamma^{\lambda}{}_{\mu[l,k]} + 2\Gamma^{\alpha}{}_{\mu[l}\Gamma^{\lambda}{}_{\alpha k]} \tag{3.7}$$

is the spin curvature tensor. More generally, the relation (2.9) continues to be valid when X is any tensor spinor. $(X_{;[kl]})^{*}$ thus stands for the usual sum of terms of which $X_{;[kl]}$ consists in the calculus of Infeld and van der Waerden,[10] except in as far as here $\Gamma^{\rho}{}_{\sigma p}$ as given by (3.5) enters into the definition (3.7) of $P^{\lambda}{}_{\mu kl}$.

Taking X to be $\gamma_{\mu\nu}$, (2.9) merely shows that $P^{\lambda}{}_{\mu kl}$ is trace-free:

$$P^{\lambda}{}_{\lambda kl} = 0 . \tag{3.8}$$

Less trivially, take X to be $\sigma_{k\dot{\mu}\nu}$. Then at once,

$$R^{p}{}_{kmn}\sigma_{p\dot{\mu}\nu} + P^{\dot{\lambda}}{}_{\dot{\mu}mn}\sigma_{k\dot{\lambda}\nu} + P^{\rho}{}_{\nu mn}\sigma_{k\dot{\mu}\rho} = 0 , \tag{3.9}$$

bearing in mind that $\sigma_{k\dot{\mu}\nu;l}$ itself vanishes. By transvection of (3.9) with $\sigma^{k\dot{\mu}\epsilon}$, it follows that

$$P^{\mu}{}_{\nu kl} = \tfrac{1}{2}S^{mn\mu}{}_{\nu}R_{mnkl} , \tag{3.10}$$

and conversely,

$$\eta^{ab}{}_{kl}R_{abmn} = -2S_{kl}{}^{\mu}{}_{\nu}P^{\nu}{}_{\mu mn} , \tag{3.11}$$

which happens to be the same as S(2.37). [One has to be on one's guard against uncritically taking over as they stand those relations of Ref. 9 which involve curvature quantities. For example, S(2.35) would not be correct here, on account of (2.16a).]

It is sometimes appropriate to write $\nabla_k X$ in place of $X_{;k}$. Then, writing $\nabla^{\dot\mu}{}_\nu := \sigma^{k\dot\mu}{}_\nu \nabla_k$, if X is any tensor spinor, one has

$$\nabla^{\dot\lambda}{}_\mu \nabla_{\dot\lambda}{}^\nu X = -\tfrac{1}{2}\delta_\mu{}^\nu \Box X - S^{kl\nu}{}_\mu X_{;kl}, \tag{3.12}$$

and the second term on the right may be transformed by means of (3.6). The notation

$$\hat{D}_{kl} := \nabla_{[k}\nabla_{l]}, \tag{3.13a}$$

$$\hat{D}_{\mu\nu} := S^{kl}{}_{\mu\nu}\hat{D}_{kl}, \tag{3.13b}$$

is sometimes helpful [cf. V3(2.4)]. As regards the commutator of $\nabla^{\dot\mu\nu}$ and $\nabla^{\dot\rho\sigma}$ one has [cf. V3(3.6)]

$$[\nabla^{\dot\mu\nu},\nabla^{\dot\rho\sigma}] = -(\gamma^{\dot\mu\dot\rho}\hat{D}^{\nu\sigma} + \gamma^{\nu\sigma}\hat{D}^{\dot\mu\dot\rho}). \tag{3.14}$$

Just as S(2.35) had to be replaced by (3.10), so the various identities of Secs. 3, 5, and 6 of Ref. 9 are not valid here as they stand, and corresponding identities need to be found. Thus, pursuing exactly the method of Sec. 3a of Ref. 9, one has, in place of S(3.1),

$$e^{klab}e^{mncd}R_{abcd} = -4R^{mnkl} + 16g^{[k[n}E^{m]l]}; \tag{3.15}$$

while in place of S(3.2)

$$e^{mncd}R_{cd}{}^{kl} - e^{klcd}R_{cd}{}^{mn} = -4e^{kls[m}E_s{}^{n]}. \tag{3.16}$$

From this a "new" identity may be obtained:

$$e^{mncd}p^{kl}{}_{cd} - e^{klcd}p^{mn}{}_{cd} = -4e^{kla[m}(E^{n]}{}_a - E_a{}^{n]}). \tag{3.17}$$

Next, the relation analogous to S(3.3) may be obtained by the method used previously, but the work involved is much more tedious. The final result is

$$\eta^{ab}{}_{kl}\bar\eta^{cd}{}_{mn}R_{abcd} = 4\eta_{kla[m}E_{n]}{}^a + \bar\eta_{mn}{}^{cd}p_{klcd}. \tag{3.18}$$

As regards (3.4), the analogous relation is

$$\eta^{kabc}R_{labc} + R_l{}^k = -\tfrac{1}{2}ie^{kabc}z_{labc}. \tag{3.19}$$

All but the first of the relations of Sec. V of Ref. 9 require modification in going over to a U_4. It will suffice to quote, without proof, merely the more important of the modified relations:

$$\mathrm{Im}(S^{ka\mu}{}_\nu P^\nu{}_{\mu la}) = \tfrac{1}{4}e^{kabc}(z_{labc} + p_{labc}), \tag{3.20}$$

$$R_{kl} = 2S_k{}^{m\nu}{}_\mu P^\mu{}_{\nu lm} - \tfrac{1}{2}ie_k{}^{abc}(z_{labc} + p_{labc}), \tag{3.21}$$

$$S^{kl\mu}{}_\alpha P^{\nu\alpha}{}_{kl} = -\tfrac{1}{4}\gamma^{\mu\nu}(R + \tfrac{1}{2}ie^{abcd}z_{abcd}) - S^{kl\mu\nu}R_{[kl]}, \tag{3.22}$$

$$S^{ab\dot\rho\dot\sigma}P^{\mu\nu}{}_{ab} = \sigma^{a\dot\rho\mu}\sigma^{b\dot\sigma\nu}E_{ab} + \tfrac{1}{4}S^{ab\mu\nu}S^{cd\dot\rho\dot\sigma}p_{abcd}, \tag{3.23}$$

$$ie^{kabc}S^l{}_{c\sigma}{}^\rho P^\sigma{}_{\rho ab} = (R^{kl} - \tfrac{1}{2}g^{kl}R) - \tfrac{1}{2}ie^{kabc}z^l{}_{abc}, \tag{3.24}$$

$$P^\mu{}_{\nu[jk;l]} = -2S_{[jk}{}^n P^\mu{}_{\nu l]n}, \tag{3.25}$$

$$\begin{aligned} S^{kl}{}_{(\rho\sigma}P^\mu{}_{\lambda)kl} &= \tfrac{1}{2}S^{kl\mu}{}_\lambda S^{mn\rho}{}_\sigma C_{klmn} - \tfrac{1}{2}(\delta^\mu{}_{(\rho}S^{ab}{}_{\sigma)\lambda} \\ &\quad + \tfrac{1}{3}\gamma_{\lambda(\rho}S^{ab}{}_{\sigma)}{}^\mu)R_{[ab]} \\ &\quad - \tfrac{1}{12}i\gamma_{\lambda(\rho}\delta^\mu{}_{\sigma)}e^{abcd}z_{abcd}. \end{aligned} \tag{3.26}$$

IV. MINIMALLY COUPLED EQUATIONS: CONSTRAINTS

Adopting minimal coupling for the time being, the equations of interest are, by prescription, formally identical with V3(1.1)—only the significance of the covariant derivative is altered:

$$\nabla^{\dot\mu_{s+1}}{}_{\nu_t}\xi^{\dot\mu_1\cdots\dot\mu_s\nu_1\cdots\nu_t} = \kappa\eta^{\dot\mu_1\cdots\dot\mu_{s+1}\nu_1\cdots\nu_{t-1}}, \tag{4.1a}$$

$$\nabla_{\dot\mu_{s+1}}{}^{\nu_t}\eta^{\dot\mu_1\cdots\dot\mu_{s+1}\nu_1\cdots\nu_{t-1}} = \kappa\xi^{\dot\mu_1\cdots\dot\mu_s\nu_1\cdots\nu_t}. \tag{4.1b}$$

Using the abbreviated notation introduced in Sec. 2 of V3, they become

$$\nabla^{\dot\mu_{s+1}}{}_{\nu_t}\xi^{\dot M_s N_t} = \kappa\eta^{\dot M_{s+1}N_{t-1}}, \tag{4.2a}$$

$$\nabla_{\dot\mu_{s+1}}{}^{\nu_t}\eta^{\dot M_{s+1}N_{t-1}} = \kappa\xi^{\dot M_s N_t}. \tag{4.2b}$$

Proceeding exactly as in Sec. II of V2, one immediately finds the equations of constraint

$$\hat{D}_{\nu_{t-1}\nu_t}\xi^{\dot M_s N_t} \stackrel{*}{=} 0, \tag{4.3a}$$

$$\hat{D}_{\dot\mu_s\dot\mu_{s+1}}\eta^{\dot M_{s+1}N_{t-1}} \stackrel{*}{=} 0. \tag{4.3b}$$

V. THE CASE OF SPIN 1

A. The scalar constraint

In contrast with the situation in a V_4, one encounters a constraint already when $S = 1$. For the sake of orientation, it is worthwhile to consider this special case first. If the field spinors are $\xi^{\mu\nu}$ and $\eta^{\dot\rho\nu}$, only (4.3a) survives here. Thus

$$\hat{D}_{\alpha\beta}\xi^{\alpha\beta} \stackrel{*}{=} 0. \tag{5.1}$$

The two terms involving the spin curvature which appear on the left are mutually equal, so that (5.1) becomes explicitly

$$K := S^{kl}{}_{\alpha\beta}(P^\alpha{}_{\gamma kl}\xi^{\beta\gamma} + S_{kl}{}^m\xi^{\alpha\beta}{}_{;m}) \stackrel{*}{=} 0. \tag{5.2}$$

Thus one has here the single scalar constraint $K = 0$. Using (3.22) it may be written as

$$K = (-R_{[kl]} + S_{kl}{}^m\nabla_m)S^{kl}{}_{\alpha\beta}\xi^{\alpha\beta} \stackrel{*}{=} 0. \tag{5.3}$$

It is convenient to define the operator

$$\hat{K}_{kl} := -R_{[kl]} + S_{kl}{}^m\nabla_m. \tag{5.4}$$

B. Implications of the constraint

When $S = 1$, the $s = 0$, $t = 2$ equations (4.1) are

$$\nabla^{\dot\lambda}{}_\nu\xi^{\mu\nu} = \kappa\eta^{\dot\lambda\nu}, \tag{5.5a}$$

$$\nabla_{\dot\lambda}{}^\nu\eta^{\dot\lambda\mu} = \kappa\xi^{\mu\nu}. \tag{5.5b}$$

Eliminating $\eta^{\dot\lambda\mu}$ between these, one obtains the second-order equation

$$(\Box + 2\kappa^2)\xi^{\mu\nu} + 2\hat{D}^\nu{}_\rho\xi^{\mu\rho} = 0. \tag{5.6}$$

On account of the prescribed symmetry of $\xi^{\mu\nu}$, this splits up into the pair of equations

$$(\Box + 2\kappa^2)\xi^{\mu\nu} + 2\hat{D}^{(\mu}{}_\rho\xi^{\nu)\rho} = 0, \tag{5.7}$$

$$\hat{D}_{\mu\nu}\xi^{\mu\nu} = 0, \tag{5.8}$$

the second of which is, of course, the constraint (5.1). The skew-symmetrized second derivatives are to be thought of as removed by means of Ricci's identity (2.9). Written out in more detail with the aid of (3.22), (5.7) reads,

$$\Box\xi^{\mu\nu} + 2S_{kl}{}^{m}S^{kl(\mu}{}_{\rho}\xi^{\nu)\rho}{}_{;m} + \{S^{kl(\mu}{}_{\alpha}(P^{\nu)}{}_{\beta kl} + \delta^{\nu)}{}_{\beta}R_{[kl]}) + \delta^{\mu}{}_{\alpha}\delta^{\nu}{}_{\beta}[2\kappa^2 - \tfrac{1}{4}(R + iz)]\}\xi^{\alpha\beta} = 0\,, \tag{5.9}$$

with $z := \frac{1}{2}e^{abcd}z_{abcd}$. By inspection, the generic form of this equation is

$$g^{kl}\xi^{\mu\nu}{}_{,kl} + A^{m\mu\nu}{}_{\alpha\beta}\xi^{\alpha\beta}{}_{,m} + B^{\mu\nu}{}_{\alpha\beta}\xi^{\alpha\beta} = 0\,, \tag{5.10}$$

while that of (5.8), that is, of (5.3), is

$$F^{m}{}_{\mu\nu}\xi^{\mu\nu}{}_{,m} + H_{\mu\nu}\xi^{\mu\nu} = 0\,. \tag{5.11}$$

A, B, F, and H are known functions of the coordinates. For example, using the abbreviation

$$S^{m\mu\nu} := S_{kl}{}^{m}S^{kl\mu\nu}\,, \tag{5.12}$$

a quantity which recurs frequently, one has

$$F^{m}{}_{\mu\nu} = S^{m}{}_{\mu\nu}\,, \tag{5.13}$$

but A, B, and H are more complicated.

Now, if the index w only goes over the range 1, 2, 3, one has from (5.11),

$$(F^{w}{}_{\mu\nu}\xi^{\mu\nu}{}_{,w} + H_{\mu\nu}\xi^{\mu\nu}) + F^{4}{}_{\mu\nu}\xi^{\mu\nu}{}_{,4} = 0\,. \tag{5.14}$$

In the absence of this equation, the Cauchy data for the wave equation (5.10)—a hyperbolic second-order equation—consist of the values of $\xi^{\mu\nu}$ and of $\xi^{\mu\nu}{}_{,4}$ on $\mathscr{H}$, and these can be arbitrarily prescribed. This freedom of choice is violated by the condition (5.14), for any particular choice (on $\mathscr{H}$) of the $\xi^{\mu\nu}$ already fixes the derivatives $\xi^{\mu\nu}{}_{,4}$ on $\mathscr{H}$. It follows that $F^{4}{}_{\mu\nu}$ must vanish. Equation (5.14) then implies that the $\xi^{\mu\nu}$ cannot be chosen arbitrarily on $\mathscr{H}$, which is not acceptable. In short, compatibility requires that $H_{\mu\nu}$ and $F^{m}{}_{\mu\nu}$ vanish. However, the vanishing of $F^{m}{}_{\mu\nu}$ implies that $S_{kl}{}^{m}$ must vanish, bearing in mind that this tensor is real; and then $H_{\mu\nu}$ also vanishes. One thus arrives at the conclusion that in a nontrivial U_4, the minimally coupled massive spin-1 Dirac equations are not compatible.

VI. IMPLICATIONS OF THE SPIN-S CONSTRAINTS

In the case of spin-S fields, the constraints are given by (4.3). It will be supposed that $t \geqslant 2$. (Should this condition not be satisfied, one simply goes over to the conjugate equations.) The equation corresponding to (5.6) is

$$(\Box + 2\kappa^2)\xi^{\dot{M}_sN_t} + 2\hat{D}^{\nu_t}{}_{\lambda}\xi^{\dot{M}_sN_{t-1}\lambda} = 0\,. \tag{6.1}$$

Because of the required symmetry of $\xi^{\dot{M}_sN_t}$ in its undotted indices, this splits up into a wave equation for $\xi^{\dot{M}_sN_t}$ and an equation of constraint:

$$(\Box + 2\kappa^2)\xi^{\dot{M}_sN_t} + 2\hat{D}^{(\nu_t}{}_{\lambda}\xi^{\dot{M}_sN_{t-1})\lambda} = 0\,, \tag{6.2}$$

$$\hat{D}_{\nu_t\nu_{t-1}}\xi^{\dot{M}_sN_t} = 0\,. \tag{6.3}$$

The generic form of (6.2) is

$$g^{kl}\xi^{\dot{M}_sN_t}{}_{,kl} + A^{m\dot{M}_sN_t}{}_{\dot{K}_s\Lambda_t}\xi^{\dot{K}_s\Lambda_t}{}_{,m} + B^{\dot{M}_sN_t}{}_{\dot{K}_s\Lambda_t}\xi^{\dot{K}_s\Lambda_t} = 0\,, \tag{6.4}$$

while that of (6.13) is

$$F^{m}{}_{\nu_t\nu_{t-1}}\xi^{\dot{M}_sN_t}{}_{,m} + H^{\dot{M}_sN_t}{}_{\dot{K}_s\Lambda_t}\xi^{\dot{K}_s\Lambda_t} = 0\,. \tag{6.5}$$

[With indices suppressed, the functions A, B, H in Eqs. (5.10) and (5.11) are special cases of those in (6.4) and (6.5); $F^{m}{}_{\alpha\beta}$ in Eq. (5.11) on the one hand and in Eq. (6.5) on the other are, however, the same functions.]

In essence the argument following Eq. (5.13) can now simply be repeated: Eq. (6.5) represents a limitation upon the prescribed freedom of choice of the Cauchy data for the wave equation (6.4), unless $S_{kl}{}^{m} = 0$. In other words, *in a nontrivial Riemann–Cartan space, the minimally coupled massive spin-S ($S > \frac{1}{2}$) Dirac equations are not compatible.*

VII. MINIMAL HERMITIAN COUPLING

The minimal-coupling prescription contains some arbitrariness already in the general theory of relativity. There, one usually appeals to the principle of equivalence; yet it is difficult to see how constraints of the kind [V2(2.3)] can be made nugatory except by the inclusion of further, nonminimally coupled, terms in the equations [V2(1.1)]. In a U_4, there is a greater degree of arbitrariness on account of the presence of an additional tensor field, i.e., the torsion. Minimal coupling, as understood hitherto, is the prescription that the field equations result from the formal replacement, in the flat-space equations, of partial by covariant derivatives with respect to the linear connection of the U_4. As a result of this process, neither the curvature tensor nor the torsion tensor will make their appearance explicitly in the equations, the torsion tensor being present only through its occurrence in the linear connection where it enters into the covariant derivative. There does not seem to be any compelling reason why the appearance of the torsion tensor in the equations should be limited in exactly this way; one is indeed merely reflecting the spirit of U_4 theory.

Now there is in fact one modification of the previous prescription which is, in a sense, inevitable; namely, when one contemplates the field equations on the level of first quantization.[12] In that case, initially again with flat space and Cartesian coordinates, one will think of $-i\partial_k$ in the derivative $-i\xi_{,k}$ (indices of ξ suppressed) as being $\hbar^{-1}$ times the operator representing the x^k component of linear momentum. While $-i\partial_k$ is Hermitian, its counterpart, $-i\nabla_k$ in a U_4, is not. On the contrary, as will be shown in a moment, it is

$$-i\nabla'_k := -i(\nabla_k - S_k) \tag{7.1}$$

which is Hermitian. Accordingly, the "simplest" acceptable prescription, one which comes closest to the previous minimal-coupling prescription, is what I shall call the prescription of *minimal Hermitian coupling*: in a U_4, the equations satisfied by the field spinors are those which are obtained from the flat-space equations by the replacement in these of ∂_k by ∇'_k. If t is any tensor spinor associated with the U_4 as such, e.g., g_{kl}, $\sigma_{k\dot{\mu}\nu}$, R_{klmn}, its covariant derivative is, of course, $t_{;k} := \nabla_k t$, as before.

To investigate the question of Hermiticity, it suffices to consider the example of the integral

$$J := i\int \xi_{\dot{\mu}}\sigma^{k\dot{\mu}\nu}(\xi_{\nu;k} + \alpha S_k\xi_\nu)w\,d^4x\,, \tag{7.2}$$

where $w := |\det g_{ij}|^{1/2}$ and α is a real constant. Here the term

$$i\int \sigma^{k\dot{\mu}\nu}\xi_{\dot{\mu}}\xi_{\nu,k}\,w\,d^4x$$

may be transformed by an integration by parts. Rejecting a surface term, it becomes

$$-i\int \xi_\nu(\sigma^{k\dot{\mu}\nu}\xi_{\dot{\mu}}w)_{,k}\,d^4X\,.$$

The partial derivative may be replaced by the covariant derivative ∇_k provided compensating terms are supplied. In so doing, two of the terms which appear differ only to the extent that corresponding to a factor $\Gamma_{km}{}^m$ in the one there is a factor $\Gamma_{mk}{}^m$ in the other and the difference between them is $2S_k$, not zero as in a V_4. In this way, one finds that

$$J = -i\int \xi_\nu\sigma^{k\dot{\mu}\nu}[\xi_{\dot{\mu};k} - (\alpha+2)\xi_{\dot{\mu}}S_k]\,w\,d^4x\,. \quad (7.3)$$

This is the complex conjugate of the expression on the right-hand side of Eq. (7.2) if and only if $\alpha = -1$, a conclusion which justifies Eq. (7.1).

VIII. THE CONSTRAINTS REVISITED

If ζ is any field spinor (indices suppressed), write $\zeta_{:k} := \nabla'_k\zeta$.

Then

$$\xi^{\mu\nu}{}_{:kl} = \xi^{\mu\nu}{}_{;kl} - 2S_{(k}\xi^{\mu\nu}{}_{;l)} - (S_{k;l} - S_kS_l)\xi^{\mu\nu}\,. \quad (8.1)$$

Therefore, if a prime indicates Hermitian coupling,

$$\Box' = \Box - 2S^k\nabla_k - (S_{k;}{}^k - S_kS^k)\,, \quad (8.2a)$$

$$\hat{D}'_{kl} = \hat{D}_{kl} + S_{[k;l]}\,. \quad (8.2b)$$

For example, in place of Eqs. (5.3) and (5.4), one now has

$$K' := \hat{K}'_{kl}S^{kl}{}_{\alpha\beta}\xi^{\alpha\beta}$$

$$:= (S_{[k;l]} - R_{[kl]} + S_{kl}{}^m\nabla_m)S^{kl}{}_{\alpha\beta}\xi^{\alpha\beta} \overset{*}{=} 0\,. \quad (8.3)$$

According to Eq. (8.2), $\Box'\xi$ differs from $\Box\xi$ only by terms free of second derivatives of ξ, while $D'_{kl}\xi$ differs from $D_{kl}\xi$ only by terms altogether free of derivatives of ξ. It follows that the generic form of Eqs. (6.4) and (6.5) is preserved, with $A,\ldots,H$ (indices suppressed) becoming $'A,\ldots,'H$, say; but, in fact, $'F = F$. Therefore the general conclusion reached at the end of Sec. VI continues to hold when minimal coupling is replaced by minimal Hermitian coupling.

ACKNOWLEDGMENTS

It is a pleasure to express my sincere thanks to Hubert Goenner for providing the motivation for this work, for many instructive conversations and much helpful advice, and for reading a preliminary version of the manuscript.

I am grateful to the Deutsche Forschungsgemeinschaft for financial support which made my stay in Göttingen possible.

[1] H. A. Buchdahl, Nuovo Cimento **10**, 96 (1958).

[2] H. A. Buchdahl, Nuovo Cimento **25**, 486 (1962).

[3] H. A. Buchdahl, J. Phys. A: Math. Gen. **15**, 1 (1982).

[4] H. A. Buchdahl, J. Phys. A: Math. Gen. **15**, 1057 (1982).

[5] H. A. Buchdahl, Class. Quantum Grav. **1**, 189 (1984).

[6] H. A. Buchdahl, Class. Quantum Grav. **4**, 1055 (1987).

[7] F. W. Hehl, P. von der Heyde, and G. D. Kerlick, Rev. Mod. Phys. **48**, 393 (1976).

[8] N. H. Barth and S. M. Christensen, J. Phys. A: Math Gen. **16**, 543 (1983).

[9] H. A. Buchdahl, J. Austral. Math. Soc. **11**, 369 (1962), hereafter referred to as S. The definitions and notation of this reference are used here.

[10] S. Hojman, M. Rosenbaum, M. P. Ryan, and L. C. Shepley, Phys. Rev. D **17**, 3141 (1978).

[11] L. Infeld and B. L. van der Waerden, Sitz. Preuss. Akad. Wiss. **9**, 380 (1930).

[12] F. Müller-Hoissen (unpublished).

A comparison of solution generating techniques for the self-dual Yang–Mills equations

Jacek Tafel
Institute of Theoretical Physics, University of Warsaw, 00-681 Warsaw, Hoża 59, Poland

(Received 17 June 1988; accepted for publication 5 October 1988)

It is shown that all self-dual Yang–Mills fields generated by the Atiyah–Ward *Ansätze* $\mathscr{A}_l$ can be obtained from the *Ansatz* $\mathscr{A}_0$ by the "dressing" method of Zakharov and Shabat [Funct. Anal. Appl. **13**, 166 (1979)].

I. INTRODUCTION

Self-dual solutions of the Yang–Mills equations are important from the mathematical and physical point of view. Among these solutions are instantons,[1] which play an important role in path integration, and static solutions, which after transformation to the Minkowski space can be interpreted as magnetic monopoles[2,3] (in the Bogomolny–Prasad–Sommerfield limit). The self-dual Yang–Mills equations represent a valuable example of a "completely integrable" system of nonlinear equations. Surprisingly they are also useful in differential geometry and topology. In particular, the existence of nonequivalent differential structures on $\mathbf{R}^4$ was proved by means of the moduli space of instantons.[4]

Self-dual fields can be obtained by different methods, in particular by (i) the Atiyah–Ward *Ansätze*,[5–7] (ii) the Zakharov–Shabat transformation,[8,9] (iii) the Bäcklund transformations related to Yang's formulation of the self-duality equations,[6,10,11] (iv) the Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction of instantons,[12] (v) Nahm's modification of the ADHM method for monopoles,[13] and (vi) others, which usually exploit strong symmetry assumptions (see, e.g., Ref. 14). The first three approaches are local. Methods (iv) and (v) refer to specific boundary conditions and it is known that they yield solutions available through the Atiyah–Ward *Ansätze*.[5,15] In this paper we focus on the first two techniques. Our main result is an observation that the space of local $GL(2,C)$ solutions generated by the Atiyah–Ward *Ansätze* $\mathscr{A}_l$ is invariant under the Zakharov–Shabat transformations and can be reproduced, using these transformations, from the *Ansätze* $\mathscr{A}_0$. In virtue of this result it is not strange that SU(2) magnetic monopoles (satisfying the Bogomolny equations) can be obtained by both methods.[9,15]

We consider pure Yang–Mills theory with the gauge group $GL(2,C)$ in the complexified Minkowski space CM. The metric tensor of CM is $g = g_{\mu\nu}\, dx^\mu \otimes dx^\nu$, where $\mu,\nu = 0,1,2,3$ and $g_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$. The Yang–Mills field is represented by a $gl(2,C)$-valued one-form (pullback of a connection)

$$A = A_\mu\, dx^\mu .$$

The field strengths are given by

$$F = dA + A \wedge A = \tfrac{1}{2} F_{\mu\nu}\, dx^\mu \wedge dx^\nu .$$

We say that A is self-dual iff

$$^*F = \pm iF, \tag{1}$$

where $^*F$ is the form dual to F. In virtue of the Bianchi identity Eq. (1) implies the Yang–Mills equation $D\,{}^*F = 0$.

In the Weyl spinor notation condition (1) reads

$$F_{A\dot{C}B}{}^{\dot{C}} = 0 \tag{2}$$

or

$$F^{C}{}_{A C \dot{B}} = 0 . \tag{3}$$

Equation (3) can be obtained from (2) by a change of orientation in space-time. For this reason we will restrict ourselves to Eq. (2) in what follows.

All considerations are local in CM and all functions are assumed to be analytic with respect to x^μ. A reduction to the SU(2) gauge theory in the four-dimensional Euclidean space can be achieved by restricting CM to a submanifold defined by $\mathrm{Im}\, x^0 = \mathrm{Re}\, x^k = 0$ and imposing reality conditions on gauge fields and corresponding wave functions.

II. THE ATIYAH–WARD AND THE ZAKHAROV–SHABAT METHODS

It can be easily verified that Eqs. (2) are integrability conditions for the following system of linear equations for a $GL(2,C)$-valued function $\psi(\lambda,x)$ (Ref. 16):

$$\lambda^A(\partial_{A\dot{B}} + A_{A\dot{B}})\psi = 0 , \tag{4}$$

where $\lambda^A = (1,\lambda)$ and $\lambda \in \overline{C} = C \cup \infty$. Given a self-dual field A one can find two solutions $\psi_\pm$ of (4), which are holomorphic functions of λ on

$$U_+ = \{\lambda \in \overline{C}:\ |\lambda| < R_+\}$$

and

$$U_- = \{\lambda \in \overline{C}:\ |\lambda| > R_-\} ,$$

respectively, where the radii $R_\pm$ satisfy $R_+ > 1 > R_-$. The intersection $U = U_+ \cap U_-$ forms a neighborhood of the unit circle Γ in $\overline{C}$. On U we can define the matrix function

$$G = \psi_+^{-1}\psi_- , \tag{5}$$

which, in virtue of (4), satisfies

$$\lambda^A \partial_{A\dot{B}} G = 0 . \tag{6}$$

Equation (6) says that G is a function of $\lambda_A x^{A\dot{B}}$ and λ only.

Conversely, given a function G satisfying (6), holomorphic on a neighborhood U of the unit circle Γ and decompos-

 0022-2488/89/030706-05$02.50

able according to (5), we can connect with it a self-dual Yang–Mills field.[16] This fact follows from the equality

$$(\lambda^A \partial_{A\dot{B}} \psi_+)\psi_+^{-1} = (\lambda^A \partial_{A\dot{B}} \psi_-)\psi_-^{-1} \quad \text{on } U \tag{7}$$

obtained by derivating (5). A direct consequence of (7) is that the lhs (or rhs) of (7) is extendable to a holomorphic function on C with at most a first-order pole at ∞. Hence there exist x-dependent coefficients $A_{A\dot{B}}$ such that (4) is satisfied with ψ equal to ψ_+ or ψ_-. They define a self-dual field since the integrability conditions of Eqs. (4) are necessarily satisfied. The most difficult part of this scheme is the Riemann–Hilbert splitting problem (5). It can be easily solved if G satisfies the Atiyah–Ward *Ansatz* $\mathscr{A}_l$ (we admit $\det G \neq 1$)

$$G = \begin{bmatrix} \lambda^{-l} e^{-f}, & P, \\ 0 & \lambda^{l} e^{q} \end{bmatrix}, \tag{8}$$

where l is a non-negative integer and f, p, and q are complex functions. Such *Ansätze* (with $f = q = 0$) were proposed to describe instantons[5] but their most spectacular success is connected with multimonopoles.[7,15]

In the Zakharov–Shabat transformation method[17] we start from a given solution A and the corresponding wave function ψ. We search for new fields A' and ψ' such that

$$\psi' = \chi\psi, \tag{9}$$

where χ is a meromorphic function of λ. In the simplest case each of the functions χ and χ^{-1} has only one first-order pole (at μ and ν, respectively). Then χ takes one of the following forms (modulo gauge transformations, which correspond to the multiplication of χ by a λ-independent nonsingular matrix function):

$$\chi = 1 + [(\mu - \nu)/(\lambda - \mu)]P, \tag{10}$$

$$\chi = 1 - P + [1/(\lambda - \mu)]P, \quad \nu = \infty, \tag{11}$$

$$\chi = 1 + (\lambda - \nu - 1)P, \quad \mu = \infty, \tag{12}$$

$$\chi = 1 + [1/(\lambda - \mu)]S, \quad \nu = \mu, \tag{13}$$

$$\chi = 1 + \lambda S, \quad \nu = \mu = \infty, \tag{14}$$

where $P^2 = P$, $S^2 = 0$, and P and S do not depend on λ. Each of the above expressions can be used to generate new self-dual solutions of the Yang–Mills equations. The case (10) is generic in the sense that all the remaining cases can be obtained from it by taking suitable limits, e.g., (12) is a limit, for $\mu \to \infty$, of (10) multiplied by $[1 + (\nu - \mu - 1)P]$ and (13) follows from (10) in the limit $\nu \to \mu$, $(\mu - \nu)P \to S$. For this reason we will restrict ourselves to the transformation (10) in what follows. This transformation can be considered as a building block for more complicated transformations (with many poles in χ).

Inserting (9) and (10) into (4) (with primed quantities) yields new potentials

$$A'_{1\dot{B}} = A_{1\dot{B}} - \partial_{2\dot{B}}((\mu - \nu)P) - [A_{2\dot{B}}, (\mu - \nu)P],$$
$$A'_{2\dot{B}} = A_{2\dot{B}}, \tag{15}$$

and a set of solvable conditions on μ, ν, and P. Concerning μ and ν they are

$$\mu^A \partial_{A\dot{B}} \mu = 0, \tag{16}$$

$$\nu^A \partial_{A\dot{B}} \nu = 0, \tag{17}$$

where $\mu^A = (1, \mu)$ and $\mu^A = (1, \nu)$. The remaining conditions yield (if $0 \neq P \neq 1$)

$$P = \rho K N^T, \quad \rho = (N^T K)^{-1}, \tag{18}$$

where K and N are C^2-valued functions (columns) satisfying

$$\nu^A(\partial_{A\dot{B}} + A_{A\dot{B}})K = 0, \tag{19}$$

$$\mu^A(\partial_{A\dot{B}} N^T - N^T A_{A\dot{B}}) = 0, \tag{20}$$

and such that $N^T K \neq 0$.

Analytic solutions of (16) and (17) are given implicitly by

$$\mu - h_1(\mu_A x^{A\dot{B}}) = 0, \quad \nu - h_2(\nu_A x^{A\dot{B}}) = 0, \tag{21}$$

where h_1 and h_2 are functions of two complex variables. If the wave function ψ is well defined for $\lambda = \mu, \nu$ then

$$N^T = n^T \psi(\mu)^{-1}, \quad K = \psi(\nu) k, \tag{22}$$

where $\psi(\alpha) = \psi|_{\lambda = \alpha}$ for any α and

$$\mu^A \partial_{A\dot{B}} n = 0, \quad \nu^A \partial_{A\dot{B}} k = 0. \tag{23}$$

It follows from (23) that the vector functions n and k depend only on $\mu_A x^{A\dot{B}}$ and $\nu_A x^{A\dot{B}}$, respectively. Note that the projection P remains unchanged if N and K undergo the following transformation:

$$N \to f_1(\mu_A x^{A\dot{B}})N, \quad K \to f_2(\nu_A x^{A\dot{B}})K,$$

which preserves Eqs. (19) and (20).

The wave function $\chi\psi$ has additional singularities at μ and ν in comparison to ψ. If μ, ν, n, and k are constants and $n^T k \neq 0$ then these singularities can be removed by multiplying $\chi\psi$ to the right by

$$Q = 1 - [(\mu - \nu)/(\lambda - \nu)](n^T k)^{-1} k n^T. \tag{24}$$

It is interesting that χ and A' can be derived in this case from the assumption that $\chi\psi Q$, with Q given by (24), has no singularities at μ and ν. This property can be used to characterize the Zakharov–Shabat transformation in the case of σ models (then μ, ν, k, and n are necessarily constant).

We can use different solutions of (4) to obtain N and K according to the formula (22). The only restriction is that one of them should be defined for $\lambda = \mu$ and the second for $\lambda = \nu$. Such a situation may occur when we apply the Zakharov–Shabat transformation to the functions $\psi_\pm$ described in the beginning of this section.

III. THE ZAKHAROV–SHABAT TRANSFORMATION OF THE ATIYAH–WARD *ANSÄTZE*

A problem arises: what are domains $U'_\pm$ and functions $\psi'_\pm$ corresponding to the self-dual field A' obtained from A by the Zakharov–Shabat transformation. Since the $\chi\psi_\pm$ satisfy Eq. (4) (with A') the functions $\psi'_\pm$ should have the form

$$\psi'_\pm = \chi\psi_\pm Q_\pm, \tag{25}$$

where

$$\lambda^A \partial_{A\dot{B}} Q_\pm = 0. \tag{26}$$

The new patching function G' is given by

$$G' = Q_+^{-1} G Q_-. \tag{27}$$

The assumption that $Q_\pm$ have the form (24) is in general incompatible with (26) (e.g., when $\mu \neq \text{const}$). Roughly speaking we should rather replace μ, ν, n, k in (24) by their λ-dependent extensions (holomorphic in U_+ or U_-) annihilated by $\lambda^A \partial_{A\dot B}$ and tending to μ, ν, n, k when $\lambda \to \mu$ or $\lambda \to \nu$. Below we construct these extensions in a neighborhood of an arbitrary (but fixed) point $p_0 \in \text{CM}$. We assume that

$$x^{A\dot B} = 0 \quad \text{at } p_0, \tag{28}$$

a condition that can be easily satisfied by a shift of coordinates.

We define the following transformations of CM depending on λ:

$$\begin{aligned} T_\lambda^+ (x^{1\dot B}, x^{2\dot B}) &= (0, \lambda_A x^{A\dot B}), \\ T_\lambda^- (x^{1\dot B}, x^{2\dot B}) &= (-\lambda^{-1}\lambda_A x^{A\dot B}, 0). \end{aligned} \tag{29}$$

If μ satisfies (16) then the pullback

$$\mu_\pm = (T_\lambda^+)^* \mu \tag{30}$$

satisfies (6) and is a holomorphic function on $U_\pm$ for small values of $x^{A\dot B}$. Moreover [compare with (21)], if μ takes values in $U_\pm$ then

$$\mu_\pm(\mu) = \mu \tag{31}$$

since both sides of (21) satisfy Eq. (16) and coincide with each other when $x^{1\dot B} = 0$ (for μ_+) or $x^{2\dot B} = 0$ (for μ_-). A more precise characteristic of $\mu_\pm$ is given by

$$\lambda - \mu_\pm = (\lambda - \mu)\exp f_\pm, \tag{32}$$

where $f_\pm$ are holomorphic functions for $\lambda \in U_\pm$. In a similar way we introduce $\nu_\pm = (T_\lambda^\pm)^* \nu$ having the property

$$\lambda - \nu_\pm = (\lambda - \nu)\exp f'_\pm. \tag{33}$$

Factorizations (32) and (33) are valid independent of values of μ and ν.

Given N and K satisfying (19) and (20) one can define

$$n_\pm^T = \begin{cases} N^T \psi_\pm(\mu), & \text{if } \mu_0 \in U_\pm, \\ N^T \psi_\mp(\mu), & \text{if } \mu_0 \notin U_\pm, \end{cases} \tag{34}$$

$$k_\pm = \begin{cases} \psi_\pm(\nu)^{-1} K, & \text{if } \nu_0 \in U_\pm, \\ \psi_\mp(\nu)^{-1} K, & \text{if } \nu_0 \notin U_\pm, \end{cases} \tag{35}$$

where μ_0 and ν_0 denote values of μ and ν, respectively, at p_0 (locally μ_0 and ν_0 determine positions of images of μ and ν with respect to boundaries of U_+ and U_-). It follows from (34), (35), and (5) that

$$n_-^T = \begin{cases} n_+^T G(\mu), & \text{if } \mu_0 \notin U, \\ n_+^T, & \text{if } \mu_0 \in U, \end{cases} \tag{36}$$

$$k_- = \begin{cases} G(\nu)^{-1} k_+, & \text{if } \nu_0 \in U, \\ k_+, & \text{if } \nu_0 \notin U. \end{cases} \tag{37}$$

The pullbacks $N_\pm = (T_\lambda^\pm)^* n_\pm$ and $K_\pm = (T_\lambda^\pm)^* k_\pm$, regarded as functions of λ, are holomorphic on $U_\pm$. They are annihilated by $\lambda^A \partial_{A\dot B}$ and satisfy

$$N_\pm(\mu) = n_\pm, \quad \text{if } \mu_0 \in U_\pm, \tag{38}$$

$$K_\pm(\nu) = k_\pm, \quad \text{if } \nu_0 \in U_\pm. \tag{39}$$

If $\psi_\pm$ are normalized in such a way that

$$\psi_\pm = 1, \quad \text{for } x^{A\dot B} = 0, \tag{40}$$

then the function $N_\pm^T K_\pm$ coincides with $N^T K$ for $x^{A\dot B} = 0$, and hence it is different from zero for small values of $x^{A\dot B}$. We assume (40) and we define [analogous to (24)]

$$\begin{aligned} Q_\pm &= 1 - [(\mu_\pm - \nu_\pm)/(\lambda - \nu_\pm)] P_\pm, \\ P_\pm &= (N_\pm^T K_\pm)^{-1} K_\pm N_\pm^T. \end{aligned} \tag{41}$$

The functions $Q_\pm$ and $Q_\pm^{-1}$ are meromorphic functions on $U_\pm$ with first-order poles at zeros of $\lambda - \mu_\pm$ and $\lambda - \nu_\pm$, respectively. The properties (32), (33), (38), and (39) guarantee that $\psi'_\pm$ given by (25) and $\psi'^{-1}_\pm$ are holomorphic functions on $U'_\pm \subset U_\pm$. If none of the values μ_0, ν_0 lies on a boundary of U_+ or U_- then $U'_\pm = U_\pm$. If it is not the case then, in general, one of the functions $\psi_\pm(\mu)$, $\psi_\pm(\nu)$ is defined only in a part of any neighborhood of the point p_0. We can avoid this problem by taking $U'_\pm$ smaller than $U_\pm$. We summarize the above results in the following proposition.

Proposition 1: The Zakharov–Shabat transformation maps solutions $\psi_\pm$ of the regular Riemann–Hilbert (5) into solutions $\psi'_\pm$ of the regular Riemann–Hilbert problem with $G' = Q_+^{-1} G Q_-$, where $\psi'_\pm$ and $Q_\pm$ are given by (25) and (41).

The functions $\psi'_\pm$ are not uniquely defined. They can be multiplied to the right by GL$(2,C)$-valued functions annihilated by $\lambda^A \partial_{A\dot B}$ and holomorphic on $U'_\pm$, respectively. By using this freedom we can show (under some restrictions on the point p_0) that if G satisfies the Atiyah–Ward *Ansatz* $\mathscr{A}_l$ then G' is also equivalent to a triangular patching function.

Proposition 2: The Zakharov–Shabat transformation preserves the space of the self-dual solutions (defined in a neighborhood of almost every point in CM) generated by the Atiyah–Ward *Ansätze*.

Proof: Let $\psi_\pm$ be solutions of (5) for G given by (8). The normalization (40) cannot be imposed in this case since it is incompatible with (8). We look for $A_\pm$ in the form

$$Q_\pm = \begin{bmatrix} a_\pm & b_\pm \\ 0 & c_\pm \end{bmatrix}, \tag{42}$$

which guarantees the triangularity of G'. Let S denote a (zero measure) set of points p such that $n_\pm^1 = 0$ or $k_\pm^2 = 0$ at p, but $n_\pm^1 \neq 0$ or $k_\pm^2 \neq 0$, respectively, in any neighborhood of p. The superscripts 1,2 denote here particular components of $n_\pm$ and $k_\pm$. Note that the vanishing of n_+^1 (k_+^2) is equivalent to the vanishing of n_-^1 (k_-^2) in virtue of (36) and (37). We will assume that $p_0 \notin S$. If $\mu_0 \in U_\pm$ and $Q_\pm$ are given by (42) then regularity of $\psi'_\pm$ at $\lambda = \mu$ requires either

$$n_\pm^1 \neq 0, \quad c_\pm \neq 0, \quad (b_\pm / c_\pm)(\mu) = -n_\pm^2 / n_\pm^1,$$
$$a_\pm(\mu) = 0 \quad \text{(a first-order zero)}$$

or

$$n_\pm^1 = 0, \quad a_\pm \neq 0, \quad c_\pm(\mu) = 0 \quad \text{(a first-order zero)}.$$

Similar considerations at $\lambda = \nu$ (if $\nu_0 \in U_\pm$) yield

$$k_\pm^2 \neq 0, \quad a_\pm(\nu) \neq \infty, \quad (b_\pm / c_\pm)(\nu) = -k_\pm^1 / n_\pm^2,$$
$$c_\pm = \infty \quad \text{(a first-order pole)}$$

or

$$k_\pm = 0, \quad c_\pm(\nu) \neq \infty ,$$

$$a_\pm(\nu) = \infty \quad \text{(a first-order pole)} .$$

The above conditions can be fulfilled with the help of the functions $\mu_\pm$, $\nu_\pm$, $N_\pm$, and $K_\pm$. In particular, we can set

$$a_+ = (\lambda - \mu_+)^\alpha (\lambda - \nu_+)^{\beta - 1} ,$$

$$a_- = (1 - \lambda^{-1}\mu_-)^\alpha (1 - \lambda^{-1}\nu_-)^{\beta-1} ,$$

$$c_+ = (\lambda - \mu_+)^{1-\alpha}(\lambda - \nu_+)^{-\beta} ,$$

$$c_- = (1 - \lambda^{-1}\mu_-)^{1-\alpha}(1 - \lambda^{-1}\nu_-)^{-\beta} ,$$

$$b_+ = c_+(\nu_+ - \mu_+)^{-1}[\alpha(\lambda - \nu_+)N^2_+ (N^1_+)^{-1} + \beta(\lambda - \mu_+)K^1_+ (K^2_+)^{-1}] ,$$

$$b_- = c_-(\nu_- - \mu_-)^{-1}[\alpha\mu_-(1 - \lambda^{-1}\nu_-)N^2_- (N^1_-)^{-1} + \beta\nu_-(1 - \lambda^{-1}\mu_-)K^1_- (K^2_-)^{-1}] ,$$

where

$$\alpha = \operatorname{sgn}|n^1_\pm|, \quad \beta = \operatorname{sgn}|k^2_\pm| \tag{43}$$

(thus $\alpha,\beta = 0,1$) and we adopt the convention that $\alpha(N_\pm)^{-1} = 0$ if $\alpha = 0$ and $\beta(K^2_\pm)^{-1} = 0$ if $\beta = 0$. The above expressions are acceptable for any values of μ_0 and ν_0. By substituting them into (42) and (27) G' is yielded in the form (8) with new functions f, p, q, and with l replaced by $l + \Delta l$, where

$$\Delta l = \alpha + \beta - 1 . \tag{44}$$

□

It follows from (44) that $\Delta l = 0, \pm 1$. A question arises whether we can always decrease or increase the index l starting from a solution corresponding to $\mathscr{A}_l$. It is not difficult to find solutions of Eqs. (19) and (20) yielding appropriate α and β [we first choose n_+ and k_+ and then we define N and K according to (34) and (35)], but it may happen that

$$N^T K = 0 \tag{45}$$

in some neighborhood. It turns out that (45) implies strong restrictions on the initial gauge field A. To obtain them we subtract Eqs. (19) and (20) contracted with N and K, respectively. In the gauge $A_{2\dot B} = 0$ (existence of such a gauge is guaranteed by $F_{2\dot B 2\dot C} = 0$) the resulting equation is $N^T \partial_{2\dot B} K = 0$. It follows from this equation that K is proportional to a x^{2B}-independent vector K'. Using a gauge transformation preserving $A_{2\dot B}$ we can obtain

$$K^T = (K^1, 0), \quad N^T = (0, N^2) . \tag{46}$$

By substituting (46) into (19) it is shown that the components $A_{1\dot B}$ have the upper triangular form in this gauge. Such solutions (and only such) correspond to the *Ansatz* $\mathscr{A}_0$ (however, they can be also constructed from $\mathscr{A}_l$ with $l > 0$). Since the Zakharov–Shabat transformations with $\Delta l = -1$ can be applied to all other solutions, it follows that the index l can be reduced, step by step, to the value $l = 0$. In virtue of the invertibility of the Zakharov–Shabat transformations we obtain the following proposition.

Proposition 3: Every self-dual solution generated by the Atiyah–Ward *Ansatz* $\mathscr{A}_l$ can be obtained from a solution corresponding to $\mathscr{A}_0$ by means of the l-step Zakharov–Shabat transformation.

Proposition 3 is valid in a neighborhood of almost every point in CM. Propositions 1–3 can be generalized to the cases (11)–(14) by taking appropriate limits.

IV. CONCLUDING REMARKS

We have found (Proposition 1) an action of the Zakharov–Shabat transformation (with one pole) on the regular Riemann–Hilbert problem related to the self-duality equations through the Ward construction. This result allows us to repeat this transformation (with different parameters) with no restriction on the position of poles, e.g., the inverse transformation is well defined.

We have shown (Proposition 2) that the Zakharov–Shabat transformation converts the Atiyah–Ward *Ansatz* $\mathscr{A}_l$ into the *Ansatz* $\mathscr{A}_{l'}$. This remains true for a superposition of such transformations and their limits. The index l plays the role of the monopole charge for monopole solutions. It changes by $0, \pm 1$ for the simple transformations considered here. Multiple transformations can generate any self-dual field corresponding to the *Ansatz* $\mathscr{A}_l$ from a particular solution corresponding to $\mathscr{A}_0$ (Proposition 3).

Conformal transformations of CM induce transformations of self-dual fields, which can be applied alternatively with the Zakharov–Shabat transformations. Effectively we obtain in this way multiple Zakharov–Shabat transformations followed by an overall conformal transformation, provided we admit limits leading to (11)–(14).

Our results are purely local and sometimes (Propositions 2 and 3) zero-measure subsets of CM are excluded from the considerations. It seems that transforming G into the triangular form, if possible, may lead to singularities in G, which do not correspond to singularities of the gauge field (note that the surface S is not distinguished by Proposition 1).

ACKNOWLEDGMENTS

I am indebted to Werner Nahm for encouragement and helpful discussions during my stay at the Physikalisches Institut in Bonn, where part of this work was carried out. I also thank Lochlainn O'Raifeartaigh for his kind interest in this paper.

This work was partially supported by the Polish Research Programme CPBP 01.03.

[1]A. A. Belavin, A. M. Polyakov, A. S. Schwarz, and Y. S. Tuypkin, Phys. Lett. B **59**, 85 (1975).

[2]G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).

[3]A. M. Polyakov, JETP Lett. **20**, 194 (1974).

[4]S. K. Donaldson, Bull. Am. Math. Soc. **8**, 81 (1983).

[5]M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).

[6]E. Corrigan, D. B. Fairlie, P. Goddard, and R. G. Yates, Commun. Math. Phys. **58**, 223 (1978).

[7]R. S. Ward, Commun. Math. Phys. **80**, 563 (1981).

[8]A. A. Belavin and V. E. Zakharov, Phys. Lett. B **73**, 53 (1978).

[9]P. Forgács, Z. Horváth, and L. Palla, Phys. Lett. B **109**, 200 (1982).

[10]K. Pohlmeyer, Commun. Math. Phys. **72**, 37 (1980).

[11]M. K. Prasad, A. Sinha, and L. L. Chau Wang, Phys. Rev. Lett. **43**, 750 (1979).
[12]M. F. Atiyah, N. J. Hitchin, V. G. Drinfeld, and Yu. I. Manin, Phys. Lett. A **65**, 185 (1978).
[13]W. Nahm, in *Monopoles in Quantum Field Theory*, edited by N. Craigie *et al.* (World Scientific, Singapore, 1982).
[14]P. Forgács, Z. Horváth, and L. Palla, Phys. Rev. Lett. **45**, 505 (1980).
[15]E. Corrigan and P. Goddard, Commun. Math. Phys. **80**, 575 (1981).
[16]R. S. Ward, Phys. Lett A **61**, 81 (1977).
[17]V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. **13**, 166 (1979).

Three generation superstring models with maximal discrete symmetries

G. Lazarides
Physics Division, School of Technology, University of Thessaloniki, Thessaloniki, Greece
Q. Shafi
Bartol Research Institute, University of Delaware, Newark, Delaware 19716

(Received 24 June 1988; accepted for publication 19 October 1988)

Discrete symmetries typically arise when the ten-dimensional heterotic $E_8 \times E_8$ superstring theory is compactified on Calabi–Yau (CY) manifolds and they are expected to play an essential role in phenomenological applications. Yau has shown how a CY manifold that leads to three generation models can be constructed. The discrete symmetries associated with several of these models are found. In particular, the models that have the largest (maximal) discrete symmetries are identified.

I. INTRODUCTION

In the standard Calabi–Yau (CY) compactification[1] of the heterotic $E_8 \times E_8$ superstring theory in ten dimensions[2] the spin connection of the CY manifold K is embedded in the "known" E_8. This gives rise to a four-dimensional model that, among other things, possesses E_6 gauge symmetry and $N = 1$ supersymmetry. Some time ago Yau[3] constructed a simply connected CY manifold K_0 with Euler character $\chi = -18$. He also displayed a freely acting Z_3 symmetry on K_0 so that modding out K_0 by Z_3 ($K = K_0/Z_3$) leads to four-dimensional superstring models with three chiral fermion families.[4]

Discrete symmetries typically arise in the above compactification scheme.[5] They are expected to play a crucial role in phenomenological applications of the superstring theory. For instance, the presence of a suitable discrete matter parity is usually needed in superstring models to suppress rapid [i.e., unsuppressed by inverse powers of a superheavy ($\gg$TeV) scale] proton decay.[4,6] Discrete symmetries, under certain circumstances, can effectively act as "continuous" symmetries.[7] If the continuous symmetry happens to be a Peccei–Quinn U(1) symmetry, the strong CP problem encountered in the standard model can be neatly resolved through the axion mechanism. These symmetries are also expected to play an essential role in answering questions related to fermion masses, flat directions, flavor changing neutral currents, etc.

In view of the important and wide role that discrete symmetries are expected to play in physical applications, we carry out in this paper a systematic search for the three-generation Yau models that are accompanied with the largest possible (maximal) discrete symmetries. We identify two physically distinct three-generation models whose group of "honest" discrete symmetries is larger than has hitherto been discussed in the literature. As a by-product of our investigation the discrete symmetries associated with several other potentially interesting models are also found.

II. THE YAU MANIFOLD K_0

Our starting point is the simply connected CY manifold K_0 [$\chi(K_0) = -18$] constructed by Yau[3] and defined in $CP^3 \times CP^3$ through the following algebraic equations[3,4]:

$$\sum_{i=0}^{3} x_i^3 + a_0 x_1 x_2 x_3 + a_1 x_2 x_3 x_0 + a_2 x_3 x_0 x_1 + a_3 x_0 x_1 x_2 = 0,$$
$$\sum_{i=0}^{3} y_i^3 + b_0 y_1 y_2 y_3 + b_1 y_2 y_3 y_0 + b_2 y_3 y_0 y_1 + b_3 y_0 y_1 y_2 = 0, \tag{1}$$
$$x_0 y_0 + \sum_{\substack{i,j=0 \\ (i,j)\neq(0,0)}}^{3} c_{ij} x_i y_j = 0.$$

Here, x's and y's are the homogeneous coordinates of the two CP^3's and a's, b's, and c's are the 23 parameters that remain after using PGL(4) transformations in the two CP^3's and the fact that overall rescalings of the three algebraic equations are irrelevant. Provided that the transversality condition holds, the intersection of the three algebraic equations in (1) gives rise to the CY manifold K_0. Before stating this condition it is convenient to reexpress (1) as follows:

$$\sum_{i=0}^{3} x_i^3 + \sum_{(i_0 i_1 i_2 i_3) = CP(0123)} a_{i_0} x_{i_1} x_{i_2} x_{i_3} = 0,$$
$$\sum_{i=0}^{3} y_i^3 + \sum_{(i_0 i_1 i_2 i_3) = CP(0123)} b_{i_0} y_{i_1} y_{i_2} y_{i_3} = 0, \tag{2}$$
$$\sum_{i,j=0}^{3} c_{ij} x_i y_j = 0.$$

Here CP(0123) are all the cyclic permutations of (0123) and $c_{00} = 1$. Then the transversality condition reads

$$d\Omega = \frac{1}{2}\left(3x_i^2 + \sum_{\substack{(i_0 i_1 i_2 i_3) = CP(0123) \\ i_0 \neq i}} \frac{a_{i_0} x_{i_1} x_{i_2} x_{i_3}}{x_i}\right)\left[\left(3y_j^2 + \sum_{\substack{(j_0 j_1 j_2 j_3) = CP(0123) \\ j_0 \neq j}} \frac{b_{j_0} y_{j_1} y_{j_2} y_{j_3}}{y_j}\right) c_{kl} x_k - \left(3y_l^2 + \sum_{\substack{(j_0 j_1 j_2 j_3) = CP(0123) \\ j_0 \neq l}} \frac{b_{j_0} y_{j_1} y_{j_2} y_{j_3}}{y_l}\right) c_{kj} x_k\right] dx_i \wedge dy_j \wedge dy_l$$

 0022-2488/89/030711-08$02.50

$$+\frac{1}{2}\left(3y_j^2+\sum_{\substack{(j_0j_1j_2j_3)=\mathrm{CP}(0123)\\ j_0\neq j}}\frac{b_{j_0}y_{j_1}y_{j_2}y_{j_3}}{y_j}\right)\left[\left(3x_i^2+\sum_{\substack{(i_0i_1i_2i_3)=\mathrm{CP}(0123)\\ i_0\neq i}}\frac{a_{i_0}x_{i_1}x_{i_2}x_{i_3}}{x_i}\right)c_{kl}\,y_l\right.$$

$$\left.-\left(3x_k^2+\sum_{\substack{(i_0i_1i_2i_3)=\mathrm{CP}(0123)\\ i_0\neq k}}\frac{a_{i_0}x_{i_1}x_{i_2}x_{i_3}}{x_k}\right)c_{il}\,y_l\right]dx_i\wedge dy_j\wedge dx_k\neq 0, \tag{3}$$

everywhere on K_0. Here repeated indices (i,j,k,l) are summed from 0 to 3.

III. RESTRICTIONS FROM THE TRANSVERSALITY CONDITION

Now we can prove the following crucial proposition.

Proposition 1: If the matrix $c=(c_{ij})$ has at most two nonzero columns or at most two nonzero rows, the transversality condition cannot be satisfied. In other words, if at most two x's or at most two y's appear in $c_{ij}x_i\,y_j=0$, transversality cannot hold.

Proof: Without loss of generality, we will assume that only two x's are involved in $c_{ij}x_i\,y_j=0$, namely, x_0 (which is always there, since $c_{00}=1$) and, say, x_1. This means that $c_{ij}=0$, for $i=2,3$, $j=0,1,2,3$. We can choose $x_0=x_1=0$. The last equation in (2) is, then, automatically satisfied and the first becomes $x_2^3+x_3^3=0$, which has three nonzero solutions. But, for $x_0=x_1=0$, $c_{kl}x_k=0$, $\forall l$, and the first of the two terms in the rhs of (3) vanishes. To achieve the vanishing of the second term in the rhs of (3), too, we can choose y's so that $c_{0l}\,y_l=0$ and $c_{1l}\,y_l=0$. If such a choice is possible in K_0, then there exist points in K_0 where $d\Omega=0$ and the transversality condition does not hold. If $(c_{10},c_{11},c_{12},c_{13})\neq\lambda\,(c_{00},c_{01},c_{02},c_{03})$, $\forall\lambda\in C$, we can eliminate y_0 among $c_{0l}\,y_l=0$ and $c_{1l}\,y_l=0$ and obtain an equation $\Sigma_{i=1}^3\alpha_i\,y_i=0$ with $(\alpha_1,\alpha_2,\alpha_3)\neq(0,0,0)$. Assuming, for example, $\alpha_1\neq 0$, we then get $y_1=\beta y_2+\gamma y_3$. By substituting this into $c_{0l}\,y_l=0$ and solving w.r.t. y_0 (remember $c_{00}=1$) we obtain $y_0=\delta y_2+\epsilon y_3$. The second equation in (2), then, becomes $ay_2^3+by_3^3+cy_2^2\,y_3+dy_2\,y_3^2=0$. For $a,b\neq 0$, this equation can be written as $a(\,y_2/y_3)^3+c(\,y_2/y_3)^2+d(\,y_2/y_3)+b=0$. This cubic equation has three nonzero solutions for y_2/y_3 and, so, there are at least three choices of y's on K_0 with $d\Omega=0$. (We have found $3\times 3=9$ "singular" points on K_0.) For $a=0$, we can choose $y_3=0$ and any y_2 for $ay_2^3+by_3^3+cy_2^2\,y_3+dy_2\,y_3^2=0$ to be satisfied. Thus there is at least one choice of y's with $d\Omega=0$ and $3\times 1=3$ "singular" points on K_0. The same is true for $b=0$. To complete the proof, consider the case $(c_{10},c_{11},c_{12},c_{13})=\lambda(c_{00},c_{01},c_{02},c_{03})$, for some $\lambda\in C$. Then, there is only one independent equation $c_{0l}\,y_l=0$, which we can solve w.r.t. y_0 to get $y_0=\alpha_1\,y_1+\alpha_2\,y_2+\alpha_3\,y_3$. Substitution of this into the second equation in (2) gives a cubic in y_1,y_2,y_3, i.e., in a CP^2. There are many possible choices of singular points in this case. For example, by choosing $y_1=0$, this cubic reduces to $ay_2^3+by_3^3+cy_2^2\,y_3+dy_2\,y_3^2=0$ and we can follow the previous discussion.

Thus we have learned that for transversality to have a chance, we need at least three x's *and* at least three y's to appear in $c_{ij}x_i\,y_j=0$. Without loss of generality, we can choose x_0,x_1,x_2 and y_0,y_1,y_2 to be involved in this bilinear equation (x_0,y_0 are, of course, always involved). This can be "minimally" achieved if $c_{00}=1$, $c_{11}\neq 0$, $c_{22}\neq 0$, and all other c's as well as all a's and b's are equal to zero (the other possible "minimal" choice $c_{12}\neq 0$, $c_{21}\neq 0$ reduces to this if $y_1\leftrightarrow y_2$). In this case, Eq. (2) becomes

$$\sum_{i=0}^3 x_i^3=0,\quad \sum_{j=0}^3 y_j^3=0,\quad \sum_{k=0}^2 c_kx_k\,y_k=0, \tag{4}$$

with $c_0\equiv c_{00}=1$, $c_1\equiv c_{11}\neq 0$, $c_2\equiv c_{22}\neq 0$.

Now we will show our second important theorem.

Proposition 2: The transversality condition does hold for the set of equations (4) with $1+c_1^3\neq 0$, $1+c_2^3\neq 0$, $c_1^3+c_2^3\neq 0$, and $1+c_1^3+c_2^3\neq 0$.

Proof: For singular points on K_0, we find, from Eq. (3), that

$$c_lx_l\,y_j^2=c_jx_j\,y_l^2,\quad c_l\,y_lx_j^2=c_j\,y_jx_l^2,\quad \forall l,j. \tag{5}$$

(There is no summation over repeated indices.) From these equations and for $l\neq 3$, since we divide by c_l, we obtain

$$\begin{aligned}x_j\cdot c_lx_l\,y_j^2&=x_jc_jx_j\,y_l^2\\&=(c_j/c_l)y_l\cdot c_l\,y_lx_j^2\\&=(c_j/c_l)y_l\cdot c_j\,y_jx_l^2\\&=(c_j^2/c_l)y_l\,y_jx_l^2,\quad \forall j \text{ and } \forall l\neq 3.\end{aligned} \tag{6}$$

For $y_j\neq 0$ [there must be at least two such y's as can be seen from Eq. (4)], we then have

$$\begin{aligned}&c_lx_lx_j\,y_j=(c_j^2/c_l)y_lx_l^2\\&\Rightarrow x_lc_jx_j\,y_j\\&\quad=(c_j^3/c_l^2)y_lx_l^2,\quad \forall j, \text{ and } \forall l\neq 3.\end{aligned} \tag{7}$$

For $y_j=0$,

$$x_lc_jx_j\,y_j=0,\quad \forall l\neq 3. \tag{8}$$

Thus

$$0=x_l\sum_{j=0}^2 c_jx_j\,y_j=\left[\sum_{\substack{j=0\\ y_j\neq 0}}^2 c_j^3\right]\cdot\frac{y_lx_l^2}{c_l^2},\quad \forall l\neq 3. \tag{9}$$

The sum in the rhs of Eq. (9) is over all j's between 0 and 2 for which $y_j\neq 0$. There is at least one such j (since there are at least two nonzero y's) and this sum is never empty. By assuming that $1+c_1^3\neq 0$, $1+c_2^3\neq 0$, $c_1^3+c_2^3\neq 0$, and $1+c_1^3+c_2^3\neq 0$, we can ensure that this sum is nonzero. Then, Eq. (9) implies that either $y_l=0$ or $x_l=0$, $\forall l\neq 3$. Since at least one of the y_l's $(l=0,1,2)$ *and* one of the x_l's $(l=0,1,2)$ must be nonzero, we are led to the following two possibilities: either one or two of the x_l's $(l=0,1,2)$ are zero. In the first case (the second is analogous) we can take, for example, $x_0=0$. Then $y_1=y_2=0$ and $y_0\neq 0$. From $c_lx_l\,y_j^2=c_jx_j\,y_l^2$, for $j=0$ and $l=1,2$, we then obtain $x_1=x_2=0$, which is not possible, since $x_0=0$, too. This shows that there are no points in K_0 where $d\Omega=0$.

For $1 + c_1^3 + c_2^3 = 0$, there are singular points. For example, $x_0 = y_0 = 1, x_1 = y_1 = c_1, x_2 = y_2 = c_2, x_3 = y_3 = 0$. This obviously satisfies Eqs. (4) and (5). For $1 + c_1^3 = 0$, $x_0 = y_0 = 1$, $x_1 = y_1 = c_1$, $x_2 = y_2 = x_3 = y_3 = 0$ is a singular point of K_0. Similarly for $1 + c_2^3 = 0$ and $c_1^3 + c_2^3 = 0$ there are singular points: $x_0 = y_0 = 1$, $x_2 = y_2 = c_2$, $x_1 = y_1 = x_3 = y_3 = 0$ and $x_1 = y_1 = c_1$, $x_2 = y_2 = c_2$, $x_0 = y_0 = x_3 = y_3 = 0$, respectively.

The discrete symmetries of (2) will include, in general, permutation symmetries involving the x's and the y's. Let P_4 denote the permutation group of four objects consisting of 24 elements. We will denote by $p(\alpha), q(\beta),\ldots$ permutations of order $\alpha, \beta,\ldots$ (p is of order α if $p^\alpha = 1$ and $p^n \neq 1$ for $n = 1,2,\ldots,\alpha - 1$). For $\alpha = 2$ we will use the further notation $\alpha = 2'$ or $2''$ for permutations of one pair or simultaneous permutations of two separate pairs, respectively. Any permutation of the x's and y's is then denoted as $(p(\alpha), q(\beta))$. We next explore the restrictions imposed on the allowed $(p(\alpha), q(\beta))$ by the requirement that the transversality condition holds.

Lemma 1: $(1, p(\alpha))$ [and $(p(\alpha), 1)$] with $\alpha = 2'', 3, 4$ are not allowed.

Proof: For $\alpha = 2''$, there will be two pairs of equal columns in the matrix $c = (c_{ij})$. Without loss of generality we can assume $c_{i0} = c_{i1}$, $c_{i2} = c_{i3}$, $i = 0,1,2,3$. For $\alpha = 3$, there will be three equal columns in c, say, $c_{i0} = c_{i1} = c_{i2}$, $i = 0,1,2,3$ and, for $\alpha = 4$, all columns are equal. We can then choose, in all these three cases, $y_0 = -y_1, y_2 = y_3 = 0$. This satisfies the second equation in (2) (terms in the second sum are zero since two of the y's vanish) and also $c_{ij}\, y_j = 0$, $i = 0,1,2,3$, which implies that the third equation in (2) is satisfied, too. The equations $c_{ij} x_i = 0, j = 0,1,2,3$, reduce to at most two linearly independent equations in all three cases and thus they leave at least two undetermined x's (the other x's can be expressed as linear combinations of the undetermined x's). The first equation in (2) then reduces to a homogeneous cubic equation with at least two variables and certainly admits nonzero solutions. Thus there are points in $CP^3 \times CP^3$ that satisfy the equations in (2) and $c_{ij}\, y_j = 0$, $c_{ji} x_j = 0, i = 0,1,2,3$. From (3), $d\Omega = 0$ for these points and so transversality fails.

Lemma 2: $(1, p(2'))$ [and $(p(2'), 1)$] are not allowed.

Proof: Without loss of generality we can assume that $p(2')$ interchanges y_0 and y_1. Then, the first two columns of c will be equal, i.e., $c_{i0} = c_{i1}$, $i = 0,1,2,3$. We can take $y_0 = -y_1$, $y_2 = y_3 = 0$. The last two equations in (2) are then automatically satisfied. Moreover, $c_{ij}\, y_j = 0$, $i = 0,1,2,3$, and the terms proportional to it in the rhs of Eq. (3) vanish. In order to have $d\Omega = 0$, it suffices to have

$$\left(y_i^2 + \frac{1}{3} \sum_{\substack{(i_0 i_1 i_2 i_3) = \mathrm{CP}(0123) \\ i_0 \neq i}} \frac{b_{i_0} y_{i_1} y_{i_2} y_{i_3}}{y_i} \right) c_{kj} x_k \tag{10}$$

symmetric in i, j. The term in parentheses in (10) is equal to $y_0^2, y_0^2, -(b_3/3) y_0^2, -(b_2/3) y_0^2$, for $i = 0,1,2,3$, respectively. Keeping in mind that $c_{k0} x_k = c_{k1} x_k$, one can easily show that symmetry of (10) is guaranteed if

$$c_{k2} x_k = -(b_3/3) c_{k0} x_k, \quad c_{k3} x_k = -(b_2/3) c_{k0} x_k. \tag{11}$$

These equations leave at least two undetermined x's and the first equation in (2) admits nonzero solutions.

Lemma 3: $(p(\alpha), q(\beta))$ with $\alpha \neq \beta$ ($\alpha, \beta = 1,2,3,4$) are not allowed.

Proof: This follows from Lemmas 1 and 2, for α or $\beta = 1$. Elements of the type $(p(2), q(3))$ and $(p(2), q(4))$ are also forbidden since their squares are of the type $(1, r(3))$ and $(1, r(2))$, respectively. The cube of $(p(3), q(4))$ is of the type $(1, r(4))$ and so elements of this type are also not allowed.

Lemma 4: $(p(2''), q(2'))$ [and $(p(2'), q(2''))$] are not allowed.

Proof: For definiteness we will take $p(2'')$ to be a simultaneous interchange of x_0, x_1 and x_2, x_3 ($x_0 \leftrightarrow x_1, x_2 \leftrightarrow x_3$) and $q(2')$ a permutation of y_0, y_1. The matrix c then takes the form

$$\left(\begin{array}{cc|cc} \alpha & \beta & \gamma & \delta \\ \beta & \alpha & \gamma & \delta \\ \hline \epsilon & \zeta & \eta & \theta \\ \zeta & \epsilon & \eta & \theta \end{array} \right). \tag{12}$$

The cubic equations in (2) become

$$\sum_{i=0}^{3} x_i^3 + a x_1 x_2 x_3 + a x_0 x_2 x_3 + a' x_0 x_1 x_3 + a' x_0 x_1 x_2 = 0,$$

$$\sum_{i=0}^{3} y_i^3 + b y_1 y_2 y_3 + b y_0 y_2 y_3 + b' y_0 y_1 y_3 + b'' y_0 y_1 y_2 = 0. \tag{13}$$

We now choose $x_0 = -x_1$, $x_2 = -x_3$. The first equation in (13) is then automatically satisfied. We further take $(\alpha - \beta) x_0 + (\epsilon - \zeta) x_2 = 0$. There is at least one nonzero choice of x's that satisfies these equations and, for these x's, $c_{ij} x_i = 0$, $j = 0,1,2,3$. So, the third equation in (2) is also satisfied and the first term in the rhs of (3) vanishes. To complete the argument we must now show that

$$\left(3x_i^2 + \sum_{\substack{(i_0 i_1 i_2 i_3) = \mathrm{CP}(0123) \\ i_0 \neq i}} \frac{a_{i_0} x_{i_1} x_{i_2} x_{i_3}}{x_i} \right) c_{jl}\, y_l \tag{14}$$

is symmetric in i, j for a nonzero choice of y's that satisfies the second equation in (13). The term in parentheses in (14) takes the same value for $i = 0,1$ (namely, $3x_0^2 - a x_2^2 \equiv \mu$) and also the same value for $i = 2,3$ (namely, $3x_2^2 - a' x_0^2 \equiv \nu$). By taking $y_0 = y_1$,

$$c_{0l}\, y_l = c_{1l}\, y_l = (\alpha + \beta) y_0 + \gamma y_2 + \delta y_3,$$

$$c_{2l}\, y_l = c_{3l}\, y_l = (\epsilon + \zeta) y_0 + \eta y_2 + \theta y_3.$$

Symmetry of (14) can then be achieved by taking $\mu[(\epsilon + \zeta) y_0 + \eta y_2 + \theta y_3] = \nu[(\alpha + \beta) y_0 + \gamma y_2 + \delta y_3]$. This, together with $y_0 = y_1$, leaves at least two undetermined y's and the second equation in (13) becomes a homogeneous cubic with at least two variables and admits nonzero solutions.

By combining all of the above lemmas we arrive at the following proposition.

Proposition 3: The allowed nontrivial permutation symmetries of (1) are of the type $(p(\alpha), q(\alpha))$ with $\alpha = 2', 2'', 3, 4$.

IV. MAXIMALLY SYMMETRIC MODELS

We are now ready to analyze the various possibilities in detail.

(i) No permutation symmetry: We assume here that the group of permutation symmetries of Eqs. (1) is trivial, i.e., consists of (1,1) only. By taking into account the previous analysis, we can conclude that this can be achieved with the minimal number of parameters by taking $a_i = b_i = 0$ $(i = 0,1,2,3)$, $c_{00} = 1$, $c_{11} \equiv c_1 \neq 0$, $c_{22} \equiv c_2 \neq 0$ $(c_1 \neq c_2$, $c_1, c_2 \neq 1$, $1 + c_1^3 \neq 0$, $1 + c_2^3 \neq 0$, $c_1^3 + c_2^3 \neq 0$, $1 + c_1^3 + c_2^3 \neq 0)$, and all other c's equal to zero. The symmetries of K_0 are of three types.

(a) $x_0 \to a x_0$, $y_0 \to a^{-1} y_0$; $x_1 \to b x_1$,

$y_1 \to b^{-1} y_1$; $x_2 \to c x_2$,

$y_2 \to c^{-1} y_2$; $x_3 \to d x_3$, $y_3 \to d y_3$,

$a,b,c,d \in \{1,\alpha,\alpha^2\}$, $\alpha = \exp(2\pi i/3)$.

(Here we take x_3, y_3 transforming in the same way by using the fact that an overall multiplification of x's by λ and y's by λ^{-1} is irrelevant.) So we have four commuting Z_3's. Three of them form the elements

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right]$$

(acting on the column vector $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$), $A = \operatorname{diag}(a,b,c,1)$ and A' obtained from A by the replacement $\alpha \to \alpha^2$. The fourth $\widetilde{Z}_3$ consists of

$$\left[\begin{array}{c|c} B & \\ \hline & B \end{array}\right],$$

$B = \operatorname{diag}(1,1,1,d)$.

(b) the swapping operator $x \leftrightarrow y$, represented by

$$\left[\begin{array}{c|c} & 1 \\ \hline 1 & \end{array}\right].$$

This operator commutes with $\widetilde{Z}_3$ (but not with the other Z_3's). Thus the group of discrete symmetries of K_0 is $\mathbb{W} \times \widetilde{Z}_3$, where $\mathbb{W}$ consists of

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right] \quad \text{and} \quad \left[\begin{array}{c|c} & A \\ \hline A' & \end{array}\right]$$

with $A = \operatorname{diag}(a,b,c,1)$, $a,b,c \in \{1,\alpha,\alpha^2\}$, and $\widetilde{Z}_3$ is generated by

$$\left[\begin{array}{c|c} B & \\ \hline & B \end{array}\right],$$

$B = \operatorname{diag}(1,1,1,\alpha)$. It has $3 \times 3 \times 3 \times 2 \times 3 = 162$ elements.

It is easily checked that further introduction of any other a, b, and c terms from Eqs. (1) breaks this symmetry further to a subgroup. Therefore $|\text{sym } K_0| \leqslant 54 \cdot 3$, in this case (sym K_0 is the group of discrete symmetries of K_0 and $|\ |$ denotes the number of elements). After dividing K_0 by an appropriate freely acting Z_3 (if it exists) we obtain K with $|\text{sym } K| \leqslant 54$.

(ii) The permutation symmetry contains only elements of the type $(p(2'), q(2'))$: We assume that there is at least one element $(p(2'),q(2'))$ that we can take to be the simultaneous interchange of x_1,x_2, and y_1, y_2 without loss of generality. Then no other element of this type is allowed since any $(p(2'),r(2'))$ with $r \neq q$ multiplied with $(p(2'),q(2'))$ gives $(1,rq)$ with $rq \neq 1$, which is not allowed. Also, any interchange of two x's, one of which is x_1 or x_2 combined with the interchange of x_1,x_2, gives an element of order 3 (cyclic permutation of three objects). Similarly for the y's. So we are left only with the possibility of simultaneous interchange of x_0,x_3 and y_0, y_3. But this element multiplied with the original one gives an element of the type $(p(2''),q(2''))$. So we conclude that in this case the permutation symmetry is simply a Z_2 generated by $(p(2'),q(2'))$. Now, if $c_{ij} = 0$, for $i,j = 1,2$, $(p(2'),q(2'))$ reduces to $(p(2'),1)$ and $(1,q(2'))$, which are not allowed. Thus at least one of these c's must be nonzero. Without loss of generality we take $c_{11} = c \neq 0$ and, by $(p(2'),q(2'))$, $c_{22} = c$. We need one more c, which we can take to be $c_{00} = 1$ $(c \neq 1)$ without loss of generality. All other c's as well as all a's and b's are zero. Transversality holds for $1 + c^3 \neq 0$ and $1 + 2c^3 \neq 0$.

Now the sym $K_0 = \mathbb{V} \times \widetilde{Z}_3$, where $\widetilde{Z}_3$ as before and $\mathbb{V}$ consists of

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right], \quad \left[\begin{array}{c|c} & A \\ \hline A' & \end{array}\right],$$

with

$$A = \begin{pmatrix} a & & & \\ & b & & \\ & & c & \\ & & & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & & & \\ & 0 & b & \\ & c & 0 & \\ & & & 1 \end{pmatrix},$$

$a,b,c \in \{1,\alpha,\alpha^2\}$.

Thus $|\text{sym } K_0| = 108 \cdot 3$, for this particular choice with the minimal number of parameters. One can now divide K_0 with the freely acting Z_3 generated by

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right],$$

$A = \operatorname{diag}(\alpha^2,\alpha,\alpha,1)$ or $\operatorname{diag}(1,\alpha,\alpha^2,1)$ (Ref. 8) to obtain two "maximally symmetric" K's with $|\text{sym } K| = 108$ (of course, the fact that they are maximally symmetric is to be proved). Division with any other freely acting Z_3 [for example, the one generated by $A = \operatorname{diag}(\alpha,\alpha^2,\alpha,1)$] reduces the sym K further since these Z_3's break the Z_2 permutation symmetry. (Notice that these Z_3's must be considered separately, in this case, since they are not conjugate to the first two ones.) Further introduction of any other a, b, or c terms breaks the symmetry of K_0 further, i.e., $|\text{sym } K_0| < 108 \cdot 3$. Division with an appropriate freely acting Z_3 (if it exists) gives $|\text{sym } K| < 108$.

(iii) The permutation symmetry contains only elements of the type $(p(\alpha),q(\alpha))$ with $\alpha = 2',2''$: We assume there exists at least one element $(p(2''),q(2''))$ that, without loss of generality, we take to be the simultaneous interchange

$x_0\leftrightarrow x_1$, $x_2\leftrightarrow x_3$, $y_0\leftrightarrow y_1$, $y_2\leftrightarrow y_3$. This is not compatible with elements of the type $(p(2''),r(2''))$ with $r\neq q$, because their product gives $(1,rq)$ with $rq\neq 1$, which is not allowed. Similarly for $(r(2''),q(2''))$. However, $(\hat{p}(2''),\hat{q}(2''))$ with $\hat{p}\neq p$ and $\hat{q}\neq q$ can exist. Without loss of generality, we can take this element to be the simultaneous interchanges $x_0\leftrightarrow x_3$, $x_1\leftrightarrow x_2$, $y_0\leftrightarrow y_3$, $y_1\leftrightarrow y_2$. This commutes with $(p(2''),q(2''))$, and their product is the simultaneous interchange $x_0\leftrightarrow x_2$, $x_1\leftrightarrow x_3$, $y_0\leftrightarrow y_2$, $y_1\leftrightarrow y_3$ [we will call this element $(\tilde{p}(2''),\tilde{q}(2''))$]. No other element of this type is allowed because, for every such element, its first component will coincide with $p(2'')$, $\hat{p}(2'')$, or $\tilde{p}(2'')$ and its second component will be different from the corresponding $q(2'')$, $\hat{q}(2'')$, or $\tilde{q}(2'')$. We could then produce an element $(1,r(2''))$ with $r(2'')\neq 1$, which is not allowed. Also, elements of the type $(\pi(2'),\rho(2'))$ are excluded since $\pi(2')$ can be multiplied with at least one of the $p(2'')$, $\hat{p}(2'')$ to produce a permutation of order 4. Thus the permutation symmetry, in this case, is maximally $Z_2\times Z_2$ with generators $(p(2''),q(2''))$ and $(\hat{p}(2''),\hat{q}(2''))$. The form of c is then

$$\left(\begin{array}{cc|cc} \alpha & \beta & \gamma & \delta \\ \beta & \alpha & \delta & \gamma \\ \hline \gamma & \delta & \alpha & \beta \\ \delta & \gamma & \beta & \alpha \end{array}\right). \tag{15}$$

All four blocks in (15) must be nonvanishing since if, say, $\gamma=\delta=0$, there exists a $(\pi(2'),\rho(2'))$ symmetry (namely $x_0\leftrightarrow x_1, y_0\leftrightarrow y_1$) that is not allowed. Thus the choice with the minimal number of c's is $(c\neq 1,0)$

$$c=\begin{pmatrix} 1 & 0 & 0 & c \\ 0 & 1 & c & 0 \\ 0 & c & 1 & 0 \\ c & 0 & 0 & 1 \end{pmatrix}. \tag{16}$$

(All other possibilities are equivalent to this.) We also take all a's and b's equal to zero. Apart from the $Z_2\times Z_2$ permutation symmetry we have swapping and a Z_3 generated by $x_i\to\alpha x_i$, $y_i\to\alpha^2 y_i$, $i=1,2$. So we have $2\cdot2\cdot2\cdot3=24$ elements and $|\text{sym}\,K_0|=24$. Thus, $|\text{sym}\,K|\leqslant 8$, if K exists at all. Further introduction of other a, b, or c terms violates the only candidate freely acting Z_3 and there is no K.

But we have not finished with case (iii) yet. If we assume that $(p(2''),q(2''))$ is the only element of this type, we can introduce two elements $(\hat{p}(2'),\hat{q}(2'))$ and $(\tilde{p}(2'),\tilde{q}(2'))$, say, $x_0\leftrightarrow x_1$, $y_0\leftrightarrow y_1$ and $x_2\leftrightarrow x_3$, $y_2\leftrightarrow y_3$, respectively. These two elements commute and their product gives $(p(2''),q(2''))$. Thus there exists one more choice of maximal permutation symmetry in this case. This symmetry again is $Z_2\times Z_2$. The matrix c takes the form

$$\left(\begin{array}{cc|cc} \alpha & \beta & \epsilon & \epsilon \\ \beta & \alpha & \epsilon & \epsilon \\ \hline \zeta & \zeta & \gamma & \delta \\ \zeta & \zeta & \delta & \gamma \end{array}\right). \tag{17}$$

The two diagonal blocks in (17) must be nonzero since, if $\alpha=\beta=0$, the unacceptable permutation of x_0,x_1 is restored. We must minimally then take $c_{00}=c_{11}=1$, $c_{22}=c_{33}=c\neq 0,1$. All other c's, a's, and b's are taken to be zero. In this case we have three Z_3's ($x_1\to\alpha x_1$, $y_1\to\alpha^2 y_1$; $x_2\to\alpha x_2$, $y_2\to\alpha^2 y_2$; $x_3\to\alpha x_3$, $y_3\to\alpha^2 y_3$) and $\text{sym}\,K_0=3\cdot3\cdot3\cdot2\cdot2\cdot2=3\cdot72$. Transversality holds for $1+c^3\neq 0$, $1+2c^3\neq 0$, $2+c^3\neq 0$. Now we can divide by one of the freely acting Z_3's generated by

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right],$$

$A=\text{diag}(1,1,\alpha,\alpha^2)$ or $\text{diag}(1,\alpha^2,\alpha,\alpha)$ to obtain K's with $|\text{sym}\,K|=72$. The other freely acting Z_3's, like $A=\text{diag}(1,\alpha,1,\alpha^2)$, give $|\text{sym}\,K|<72$. Introduction of any other c's, a's, or b's reduces the symmetry of K_0 further, i.e., $|\text{sym}\,K_0|<3\cdot72$ and $|\text{sym}\,K|<72$ if K exists.

Finally, in case (iii) we can have a nonmaximal permutation symmetry generated just by $(p(2''),q(2''))$. This is simply a Z_2 symmetry. The matrix c takes a block form. Now, at least three blocks of this matrix must be nonzero since otherwise we have the higher permutation symmetry of the previous case [Eq. (17)]. The choice with minimal number of parameters is (other possibilities are equivalent)

$$c=\left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & c' & c & 0 \\ c' & 0 & 0 & c \end{array}\right), \quad a_i=b_i=0, \quad i=0,1,2,3 \quad (c'\neq 1,c;\ c\neq 1). \tag{18}$$

For this choice, we have a single Z_3 ($x_i\to\alpha x_i$, $y_i\to\alpha^2 y_i$, $i=1,2$) and no swapping. Thus $|\text{sym}\,K_0|=3\cdot2$ and $|\text{sym}\,K|\leqslant 2$ if K exists. We can restore swapping by taking $c_{03}=c_{12}=c'$. Then $|\text{sym}\,K_0|=3\cdot2\cdot2$ and $|\text{sym}\,K|\leqslant 4$ if it exists. Introduction of any other c's, a's, or b's breaks necessarily the only Z_3 symmetry and is, therefore, unacceptable.

(iv) The permutation symmetry contains only elements of the type $(p(\alpha),q(\alpha))$ with $\alpha=2',2'',3$: We assume there exists at least one element $(p(3),q(3))$ which, without loss of generality, we take to be $x_0\to x_1\to x_2\to x_0$, $y_0\to y_1\to y_2\to y_0$. We will denote this cyclic permutation by $((012),(012))$. [In general, $((ij\cdots k),(lm\cdots n))$, where all $i,j,\ldots,k$ and all $l,m,\ldots,n$ are different, will denote the cyclic permutation $x_i\to x_j\to\cdots\to x_k\to x_i$, $y_l\to y_m\to\cdots\to y_n\to y_l$. Of course, $((ij\cdots k),(lm\cdots n))=((j\cdots ki),(lm\cdots n))=\cdots$. We also have products like $((ij\cdots k)(lm\cdots n)\cdots,(pq\cdots r)(st\cdots v)\cdots)$, for example, $((01)(23),(01)(23))$ = the element $(p(2''),q(2''))$ in the beginning of (iii). Also, note that $(12)(01)=(012)$, $(23)(12)(01)=(0123),\ldots$.] The c matrix takes the form

$$c=\left(\begin{array}{ccc|c} \alpha & \beta & \gamma & \delta \\ \gamma & \alpha & \beta & \delta \\ \beta & \gamma & \alpha & \delta \\ \hline \epsilon & \epsilon & \epsilon & \zeta \end{array}\right). \tag{19}$$

The a and b terms take the form

$$\begin{aligned} &a x_0 x_1 x_2 + a'(x_1x_2x_3 + x_2x_3x_0 + x_3x_0x_1), \\ &b y_0 y_1 y_2 + b'(y_1y_2y_3 + y_2y_3y_0 + y_3y_0y_1). \end{aligned} \tag{20}$$

Now we want to see which permutations of order 3 can coexist with the element $(p(3),q(3))$. One of them is, of course, its inverse $((021),(021))$. The only other element which acts only on x_0,x_1,x_2 and y_0,y_1,y_2 is $((012),(021))$ and its inverse. But $((012),(021))\cdot((012),(012))=((021),1)$. This element is, of course, not allowed. Let us now consider elements of order 3 that act on y_3 but not on x_3. Without loss of generality, we can take cyclic permutations of x_0,x_1,x_2 and y_0,y_1,y_3. There are two such elements $((021),(013))$ and $((021),(031))$ as well as their inverses. But $((021),(013))\cdot((012),(012))=(1,(02)(13))$ and $((021),(031))\cdot((012),(012))=(1,(032))$, which is unacceptable. [Some useful formulas are $(012)=(01)(02)=(12)(10)=(20)(21)$; more generally $(ijk)=(ij)(ik)=(jki)=(jk)(ji)=(kij)=(ki)(kj)$, i,j,k all different.] The last possibility is elements of order 3 that act on both x_3 and y_3. There are two choices in this case: (1) the two extra x's correspond to the two extra y's; without loss of generality, we can take cyclic permutations of x_0,x_1,x_3 and y_0,y_1,y_3, and (2) the two extra x's do not correspond to the two extra y's; we can generically take cyclic permutations of x_0,x_1,x_3 and y_0,y_2,y_3. In case (1) the possible elements are $((013),(013))$,$((013),(031))$ and their inverses. In case (2) we have $((013),(023))$,$((013),(032))$ and their inverses. But $((012),(012))\cdot((013),(031))=((12)(03),(123))$ and $((012),(012))\cdot((013),(023))=((12)(03),(013))$. Also $((013),(013))$ implies $\alpha=\zeta$, $\beta=\gamma=\delta=\epsilon$ in (19) and $a=a'$, $b=b'$ in (20), so $((0123),(0123))$ is a symmetry too. Here $((013),(032))$ implies $\alpha=\beta=\delta=\epsilon$, $\gamma=\zeta$, $a=a'$, $b=b'$, and $((0123),(2013))$ is a symmetry too. Thus we see that we either produce unacceptable elements of the type $(p(2''),q(3))$ or elements of order 4. The overall conclusion is that there is no other permutation of order 3 compatible with our original $(p(3),q(3))$ except its inverse.

Elements of the type $(p(2''),q(2''))$ are not allowed since any such element can be written in the x space as $(01)(23)$ by a cylic renaming of x_0,x_1,x_2. [Note that even independent cyclic renamings of x_0,x_1,x_2 and y_0,y_1,y_2 (for example, $x_0\to x_1$, $x_1\to x_2$, $x_2\to x_0$, $y_0\to y_2$, $y_1\to y_0$, $y_2\to y_1$) do not affect the definition of $(p(3),q(3))$ in the beginning of (iv).] Then

$$(01)(23)(012)=(01)(23)(01)(02)$$
$$=(23)(20)=(230)=(023),$$

which is unacceptable (since other elements of order 3 are not allowed).

Any element of the type $(p(2'),q(2'))$ that acts on x_3 can have $p(2')=(03)$ (by cyclic renaming of x_0,x_1,x_2). But $(012)(03)=(0123)$, which is unacceptable since it is of order 4. The same is true for y_3.

Thus we are left with the possibility of elements of the type $(p(2'),q(2'))$ that act on x_0,x_1,x_2; y_0,y_1,y_2. Any such element can be written as $((01),(01))$ by cyclic renaming of x_0,x_1,x_2; y_0,y_1,y_2 (independently). Then, since

$$(01)(012)=(01)(01)(02)=(02)$$

and

$$(012)(01)=(120)(01)=(12)(10)(01)=(12),$$

we automatically have two more elements of this type: $((02),(02))$ and $((12),(12))$. No other element of this type is allowed since, necessarily, its components will be different, i.e., $(r(2'),s(2'))$, $r\neq s$, and combined with $(r(2'),r(2'))$ gives an element of the type $(1,t(3))$, which is unacceptable.

The maximal allowed permutation in case (iv) is thus P_3 consisting of (p,p), $p=1$, $(012),(021),(01),(02),(12)$. This symmetry can be achieved with the minimal number of a's, b's, and c's by taking $\alpha=1$ in (19) (the choice β or $\gamma=1$ is equivalent by cyclic renaming), i.e., $c_{00}=c_{11}=c_{22}=1$ and all other parameters zero. We then have four Z_3's ($x_i\to\alpha x_i$, $y_i\to\alpha^2 y_i$, $i=0,1,2$, and $x_3\to\alpha x_3$, $y_3\to\alpha y_3$ [see case (i)]) and the swapping operation. Thus $|\mathrm{sym}\,K_0|=6\cdot3\cdot3\cdot3\cdot3\cdot2=108\cdot3\cdot3$. It turns out that there are three independent freely acting Z_3's in this case,[8] generated by

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right],$$

$A=\mathrm{diag}(1,\alpha,\alpha^2,1)$, $\mathrm{diag}(\alpha^2,\alpha,\alpha,1)$, and

$$\left(\begin{array}{c|c} P & \\ \hline & 1 \end{array}\right)$$

with

$$P=\begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix}. \tag{21}$$

In the first two cases the permutation symmetry P_3 breaks to Z_2 generated by $((12),(12))$. Thus $|\mathrm{sym}\,K|=108$. sym $K\times$(freely acting Z_3) is respected even if $c_{11}=c_{22}=c\neq1$, $c_{00}=1$, but all other parameters must remain zero. So, we obtain the maximally symmetric case in (ii). For

$$A=\left(\begin{array}{c|c} P & \\ \hline & 1 \end{array}\right),$$

P_3 remains unbroken and $P_3/$(freely acting Z_3) $=Z_2$ generated by, say, $((12),(12))$, but two of the four Z_3's break (only

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right],$$

$A=\mathrm{diag}(\alpha,\alpha,\alpha,1)$ and $x_3\to\alpha x_3$, $y_3\to\alpha y_3$ remain). Thus $|\mathrm{sym}\,K|=2\cdot3\cdot3\cdot2=36$.

Introduction of extra a, b, and c terms that respect the permutation symmetry P_3 breaks at least two Z_3's (and $|\mathrm{sym}\,K_0|\leqslant108$, $|\mathrm{sym}\,K|\leqslant36$) except $c_{33}=c\neq1$ or $ax_0x_1x_2$ together with $ay_0y_1y_2$ that breaks only one Z_3. For $c_{33}=c\neq1$, we can repeat the previous analysis and $|\mathrm{sym}\,K|=36$ or 12. For $ax_0x_1x_2$ and $ay_0y_1y_2$, only three Z_3's survive in sym K_0, namely

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right],$$

$A=\mathrm{diag}(\alpha,\alpha,\alpha,1)$ or $\mathrm{diag}(1,\alpha,\alpha^2,1)$ and $x_3\to\alpha x_3$, $y_3\to\alpha y_3$. Thus $|\mathrm{sym}\,K_0|=6\cdot3\cdot3\cdot3\cdot2=108\cdot3$. The only possible freely acting Z_3's are $A=\mathrm{diag}(1,\alpha,\alpha^2,1)$ and

$$A=\left(\begin{array}{c|c} P & \\ \hline & 1 \end{array}\right).$$

But $A = \mathrm{diag}(1,\alpha,\alpha^2,1)$ breaks P_3 to Z_2 and

$$A = \left(\begin{array}{c|c} P & \\ \hline & 1 \end{array}\right)$$

breaks $A = \mathrm{diag}(1,\alpha,\alpha^2,1)$. Thus $|\mathrm{sym}\, K| = 36$.

Finally, we must consider the case of nonmaximal permutation symmetry generated by $(p(3),q(3))$ only. This is simply a Z_3. In this case we must at least take β or $\gamma = c \neq 0,1$ in (19) in addition to $\alpha = 1$ and break two Z_3's. Thus $|\mathrm{sym}\, K_0| \leqslant 54$ and $|\mathrm{sym}\, K| \leqslant 18$.

(v) The permutation symmetry contains elements of the type $(p(4),q(4))$: We assume there exists at least one element of type $(p(4),q(4))$ that, without loss of generality, we take to be $((0123),(0123))$. The c matrix takes the form

$$c = \begin{pmatrix} \alpha & \beta & \gamma & \delta \\ \delta & \alpha & \beta & \gamma \\ \gamma & \delta & \alpha & \beta \\ \beta & \gamma & \delta & \alpha \end{pmatrix}. \tag{22}$$

The a and b terms become

$$\begin{aligned} &a(x_0x_1x_2 + x_1x_2x_3 + x_2x_3x_0 + x_3x_0x_1), \\ &b(y_0y_1y_2 + y_1y_2y_3 + y_2y_3y_0 + y_3y_0y_1). \end{aligned} \tag{23}$$

Now we want to see which permutations of order 4 can coexist with this one. There are $3! = 6$ cyclic permutations of four objects. We will denote them as $p = (0123)$, $q = (0231)$, $r = (0312)$, $p^{-1} = (0321)$, $q^{-1} = (0132)$, $r^{-1} = (0213)$. Then, our original element is (p,p) and its inverse (p^{-1},p^{-1}) is automatically included. Any other compatible element (s,t) must have both s and t different from p and p^{-1}, since, if, say, $s = p$, $t \neq p$, $(p^{-1},p^{-1})(p,t) = (1,p^{-1}t)$ with $p^{-1}t \neq 1$. But we have the freedom of independent cyclic renamings of x_0,x_1,x_2,x_3 and y_0,y_1,y_2,y_3 without affecting (p,p). Under such cyclic renamings:

$$\begin{aligned} q = (0231) \to (1302) &= (0213) = r^{-1} \to (1320) \\ &= (0132) = q^{-1} \to (1203) \\ &= (0312) = r \to (1023) \\ &= (0231) = q. \end{aligned}$$

Any extra (s,t) can be brought to the form (q,q) [its inverse (q^{-1},q^{-1}) is, of course, included]. More elements (s,t) compatible with $(p,p),(q,q)$ (and their inverses) must have $s,t \neq p,q,p^{-1},q^{-1}$. We can have (r,r) and its inverse or (r,r^{-1}) and its inverse [(r,r) and (r,r^{-1}) are not compatible with each other since their product is $(r^2,1)$]. But

$$(p,p)(q,q)(p^{-1},p^{-1}) = (r,r)$$

and this element together with its inverse is automatically included. From the above discussion it is obvious that no other element of order 4 can coexist with these. Thus the maximal set of elements of order 4 consists of $(p,p),(q,q),(r,r)$ and their inverses, i.e., all "diagonal" (acting identically on x's and y's) permutations of order 4.

Now note that $q^{-1}p^{-1} = (123)$, $rp^{-1} = (023)$, $qp^{-1} = (013)$, and $qr^{-1}(012)$. By taking various products of the diagonal permutations of order 4 we can produce all diagonal permutations of order 3 [there are eight such permutations (s,s) with $s = (123),(023),(013),(012)$, and their inverses]. No other permutation (s,t) with $s \neq t$ can exist since

$$(s^{-1},s^{-1})(s,t) = (1,s^{-1}t)$$

with $s^{-1}t \neq 1$. Also $p^2 = (02)(13)$, $q^2 = (03)(12)$, $r^2 = (01)(23)$. So all three diagonal permutations of type $(s(2''),s(2''))$ with $s = (02)(13),(03)(12),(01)(23)$ can be produced, too. No other element of type $(s(2''),t(2''))$ with $s \neq t$ can exist since

$$(s(2''),s(2''))(s(2''),t(2'')) = (1,st),$$

with $st \neq 1$. Finally note that

$$\begin{aligned} p = (0123) &= (012)(03) = (1230) = (123)(01) \\ &= (2301) = (230)(21) = (023)(12) \\ &= (3012) = (301)(32) = (013)(23), \end{aligned}$$

$$\begin{aligned} q = (0231) &= (2310) = (231)(20) = (123)(02) \\ &= (1023) = (102)(13) = (021)(13). \end{aligned}$$

The six diagonal permutations of type $(s(2'),s(2'))$ with $s = (01),(02),(03),(12),(13),(23)$ can thus be obtained. No other element of the type $(s(2'),t(2'))$ with $s \neq t$ is allowed since $(s,s)(s,t) = (1,st)$ with $st \neq 1$.

The maximal permutation symmetry in this case is the 24-element group of all "diagonal" permutations. In (22) we must then have $\beta = \gamma = \delta$. This maximal permutation symmetry of K_0 can be achieved with the minimal number of parameters by taking $c_{00} = c_{11} = c_{22} = c_{33} = 1$ and all other parameters zero. In this case we have three Z_3's ($x_i \to \alpha x_i$, $y_i \to \alpha^2 y_i$, $i = 1,2,3$) and the Z_2 swapping. Thus $|\mathrm{sym}\, K_0| = 24\cdot 3\cdot 3\cdot 3\cdot 2$. There are two independent freely acting Z_3's generated by

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right],$$

$A = \mathrm{diag}(1,\alpha^2,\alpha,\alpha)$ or

$$A = \left(\begin{array}{c|c} 1 & \\ \hline & P \end{array}\right)$$

with

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

For $A = \mathrm{diag}(1,\alpha^2,\alpha,\alpha)$, the group of permutations breaks to $Z_2 \times Z_2$ (diagonal permutations in 0,1 and 2,3) and $|\mathrm{sym}\, K| = 2\cdot 2\cdot 3\cdot 3\cdot 2 = 72$. For

$$A = \left(\begin{array}{c|c} 1 & \\ \hline & P \end{array}\right),$$

the three Z_3's reduce to one generated by

$$\left[\begin{array}{c|c} A & \\ \hline & A' \end{array}\right],$$

$A = \mathrm{diag}(1,\alpha,\alpha,\alpha)$ and, consequently, $|\mathrm{sym}\, K_0| \leqslant 24\cdot 3\cdot 2$. This implies $|\mathrm{sym}\, K| \leqslant 48$. Introduction of more a's, b's, and c's breaks the three Z_3's (remember $\beta = \gamma = \delta$) and $|\mathrm{sym}\, K_0| \leqslant 48$. Thus $|\mathrm{sym}\, K| \leqslant 16$.

Lastly, we must consider the cases of nonmaximal permutation symmetry of K_0. This means that we must take at least one of the parameters β,γ,δ in (22) different from zero (together with $\alpha \neq 0$). But then the three Z_3's are broken and $|\mathrm{sym}\, K_0| \leqslant 48$. This implies that $|\mathrm{sym}\, K| \leqslant 16$ in these cases.

The overall conclusion is that the two K's constructed in (ii) with $|\mathrm{sym}\, K| = 108$ are really the maximally symmetric spaces.

In conclusion, we have identified specific complex structures of the three generation Calabi–Yau manifold that lead to four-dimensional superstring models with maximal discrete symmetries. Armed with this information the next step is to investigate the physical implications that are a consequence of these symmetries. We hope to discuss this in the future.

ACKNOWLEDGMENT

Q.S. received partial financial support from the Department of Energy under Contract No. DE-AC02-78ER05007.

[1]P. Candelas, G. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. B **258**, 46 (1985); E. Witten, *ibid.* B **258**, 75 (1985).

[2]M. Green and J. Schwarz, Phys. Lett. B **149**, 117 (1984); D. Gross, J. Harvey, E. Martinec, and R. Rohm, Phys. Rev. Lett. **55**, 502 (1985).

[3]S. T. Yau, in *Symposium on Anomalies, Geometry, and Topology*, edited by W. A. Bardeen and A. R. White (World Scientific, Singapore, 1985), p. 395.

[4]For phenomenological applications see B. R. Greene, K. Kirklin, P. Miron, and G. G. Ross, Nucl. Phys. B **278**, 667 (1986); Phys. Lett. B **180**, 69 (1986); Nucl. Phys. B **292**, 606 (1987).

[5]See, particularly, E. Witten, Ref. 1.

[6]M. C. Bento, L. Hall, and G. G. Ross, Nucl. Phys. B **292**, 400 (1987); G. Lazarides and Q. Shafi, Nucl. Phys. B **308**, 451 (1988).

[7]G. Lazarides, C. Panagiotakopoulos, and Q. Shafi, Phys. Rev. Lett. **56**, 432 (1986).

[8]See G. Lazarides and Q. Shafi, Phys. Lett. B **206**, 32 (1988), for a discussion of how the various freely acting Z_3 symmetries can be found.

Finite-dimensional irreducible representations of the SU(3/1) superalgebra

Bani Mitra[a)]
S. G. T. B. Khalsa College, University of Delhi, Delhi-110007, India

(Received 15 July 1988; accepted for publication 9 November 1988)

The SU(3/1) superalgebra is constructed using Schwinger's harmonic oscillator technique. Finite-dimensional irreducible representations of the superalgebra are obtained choosing a suitable weight Ψ_m, which will be called the maximal weight, using the commutation relations of the superalgebra only and without any knowledge of the matrix elements of the generators of the superalgebra. Atypical representations of the superalgebra are obtained and a comparison of these representations is made with finite-dimensional irreducible representations of SU(4).

I. INTRODUCTION

Lie superalgebras have become increasingly important in theoretical physics in the last decade and a half.[1] We only mention their application in superunification,[2] nuclear physics,[3] and in supergravity.[4]

The classical Lie superalgebras have been classified by Kac,[5] and independently by Scheunert *et al.*[6,7] Finite-dimensional irreducible representations (FDIR's) of Lie superalgebras were studied by Kac[8] and these were divided into two groups, typical representations and atypical representations. Kac constructed a character formula for typical representations for the classical Lie superalgebras and thus solved the weight space problem for representations of these superalgebras. Atypical representations are harder to deal with as the weights decouple at some particular stage.[9]

However, representations of very few superalgebras have been analyzed in detail,[10,11] and it is of interest to find explicitly the decomposition of the irreducible representation (IR) of the superalgebra into at least a direct sum of IR's of its even parts, in order to visualize the total representation and in order to gain an understanding of the action of its odd generators.

We examine the FDIR's of the SU(3/1) superalgebra as an example, with a view to answering this question. We begin by noting the strong analogy between representations of the SU(2/1) (see Ref. 10) superalgebra and the SU(3) Lie algebra.[12,13] We also note that the IR's of SU(2/1) become atypical, when the odd lowering operators (defined differently from Ref. 9) act on a certain weight Ψ_m, which we call the maximal weight (to distinguish it from the highest weight Ψ_w), to give zero, thus making the atypical representation of SU(2/1) analogous to an elementary representation of SU(3). We thus define the maximal weight of SU(3/1) in a suitable way and note that indeed the FDIR becomes atypical if the action of one of the odd lowering operators on a maximal state of SU(3) is zero. For the independent operators,[14] the decoupling takes place at the first stage Ψ_m itself. For the nonindependent operators, the decoupling takes place at a later stage and is coupled with the elementarity of the representation (see Sec. III for clarification) as we shall see. In SU(2/1), both odd lowering operators are independent.

This premise gains further ground from the knowledge of the fact that any member of the osp(1/2p) sequence contains no finite-dimensional atypical representations.[15] If one examines the osp(1/2) superalgebra,[10] we note that the action of the odd lowering operators on the maximal state can be zero, only if the representation is trivial (this has been elaborated in Sec. II). The same result holds for the osp(1/4) superalgebra, and we conjecture that this holds in general.

Our task would have greater validity if a familiarity with SU(3) tensorial techniques would have enabled us to write down the actual matrix elements of the generators of the SU(3/1) superalgebra on an element of the IR space. Nevertheless, fairly general statements on the nature of the FDIR space can be made with this simple, "workman's" technique, especially on the nature of the decoupled representation, the reason being that the so-called "spin" associated with the lowering operator under consideration for the maximal state is zero. In particular, we also show that the adjoint representation of SU(3/1) is atypical.

This paper has been given a pedagogical approach and the notation used is that of Ref. 13, which, though just adequate for the present paper, needs to be modified if a generalization to the SU(n/1) case is contemplated. This will be done in a forthcoming paper.

The outline of this paper is as follows. In Sec. II, we review the FDIR's of the SU(3) Lie algebra partly to establish the notation for the SU(3/1) case, but basically to display the strong analogy between the FDIR's of SU(3) and those of SU(2/1).

In Sec. III, we construct the SU(3/1) superalgebra using Schwinger's harmonic oscillator technique[16,17] and obtain the FDIR space of SU(3/1). We also show how the FDIR space becomes atypical using the above-mentioned conditions. In Sec. IV, we make a comparison of the FDIR space of SU(3/1) with that of SU(4) [the FDIR space of SU(4) being discussed in the Appendix] and also study the structure of atypical representations of SU(3/1) vis-à-vis their SU(4) counterparts. In Sec. V, we end by summarizing our conclusions and with some comments on our philosophy and outlook.

[a)] Present address: EP Division, CERN, CH-1211, Geneva-23, Switzerland.

 0022-2488/89/030719-13$02.50

II. COMPARATIVE STUDY OF THE FDIR SPACE OF SU(3) AND SU(2/1)

In this section we develop the FDIR space of SU(3) and dwell on the analogy between the two algebras SU(3) and SU(2/1) while constructing the FDIR space of the Lie superalgebra (LSA).

A. FDIR space of SU(3)

The SU(3) Lie algebra (LA) is made up of the eight generators F_i ($i = 1,...,8$) (see Ref. 3) with commutation relations

$$[F_i,F_j] = if_{ijk}F_k\,, \tag{2.1}$$

where the f_{ijk} are the structure constants of SU(3) that are totally antisymmetric in i, j, and k. If one chooses the linear combinations of these F_i,

$$T_\pm = F_1 \pm iF_2\,, \tag{2.2a}$$

$$V_\pm = F_4 \pm iF_5\,, \tag{2.2b}$$

$$U_\pm = F_6 \pm iF_7\,, \tag{2.2c}$$

$$T_3 = F_3\,, \tag{2.2d}$$

$$Y = (2/\sqrt{3})F_8\,, \tag{2.2e}$$

one obtains the following LA for the generators $T_\pm$, $V_\pm$, $U_\pm$, T_3, and Y:

$$[T_3,T_\pm] = \pm T_\pm\,, \quad [Y,T_\pm] = 0\,, \tag{2.3a}$$

$$[T_3,V_\pm] = \pm \tfrac{1}{2} V_\pm\,, \quad [Y,V_\pm] = \pm V_\pm\,, \tag{2.3b}$$

$$[T_3,U_\pm] = \mp \tfrac{1}{2} U_\pm\,, \quad [Y,U_\pm] = \pm U_\pm\,, \tag{2.3c}$$

$$[T_+,T_-] = 2T_3, \quad [U_+,U_-] = \tfrac{3}{2}Y - T_3,$$

$$[V_+,V_-] = \tfrac{3}{2}Y + T_3\,, \tag{2.3d}$$

$$[T_\pm,V_\pm] = [T_\pm,U_\mp] = [U_\pm,V_\pm] = 0\,, \tag{2.3e}$$

$$[T_\pm,V_\mp] = \mp U_\mp\,, \quad [T_\pm,U_\pm] = \pm V_\pm\,, \tag{2.3f}$$

$$[U_\pm,V_\mp] = \pm T_\mp\,, \quad [T_3,Y] = 0\,. \tag{2.3g}$$

A representation state is taken to be an eigenstate of the completely commuting set of operators T_3, Y, and T^2, and is labeled as $|t,t_3,y\rangle$ where

$$T_3|t,t_3,y\rangle = t_3|t,t_3,y\rangle, \tag{2.4a}$$

$$T^2|t,t_3,y\rangle = t(t+1)|t,t_3,y\rangle, \quad -t \leqslant t_3 \leqslant t\,, \tag{2.4b}$$

$$Y|t,t_3,y\rangle = y|t,t_3,y\rangle. \tag{2.4c}$$

In order to obtain the IR space, we label the maximal state Ψ_m as the one for which the eigenvalue of T_3 is the maximum and cannot be raised further. As T_+, V_+, and U_- raise t_3, we stipulate that

$$T_+\Psi_m = V_+\Psi_m = U_-\Psi_m = 0. \tag{2.5}$$

The term maximal is used for Ψ_m in order to distinguish it from the conventional usage of the term "highest weight." The maximal weight of a representation Ψ_m is not, in general, the highest weight Ψ_w that is defined:

$$T_+\Psi_m = V_+\Psi_w = U_+\Psi_w = 0.$$

The FDIR space, denoted by the labels (p,q), where p and q are positive integers, is obtained as follows. The boundary of the IR is obtained by the repeated application of V_- on Ψ_m p times:

$$V_-^{p+1}\Psi_m = 0\,.$$

After this we apply T_- on the state $V_-^p\,\Psi_m$ q times, until we reach a point

$$T_-^{q+1}V_-^p\,\Psi_m = 0\,.$$

The boundary of the IR space is completed by applying U_+ on $T_-^q\,V_-^p\,\Psi_m$ p times, V_+ on $U_+^p\,T_-^q\,V_-^p\,\Psi_m$ q times, T_+ on $V_+^q\,U_+^p\,T_-^q\,V_-^p\,\Psi_m$ p times, and finally U_- on $T_+^p\,V_+^q\,U_+^p\,T_-^q\,V_-^p\,\Psi_m$ q times. The boundary is a six-sided figure, symmetric under reflections about the y axis, as well as under rotations by 120° (Fig. 1). Also there is only one state at each site on the boundary (specified by eigenvalues of the commuting operators T_3 and Y), two states at each site on the inner layer, three states at each site on the next layer, until a triangular layer is reached beyond which the multiplicity ceases to increase. For "triangular" or elementary IR's characterized by either p or q equal to zero, each site is occupied only once.

The representation space is best understood by the harmonic oscillator realization of SU(3).[16,17] If one introduces creation and annihilation operators for two sets of three-dimensional harmonic oscillators, a_i^+, b_i^+, a_i, b_i ($i = 1,2,3$), one obtains

$$[a_i,a_j^+] = \delta_{ij}, \quad [b_i,b_j^+] = \delta_{ij}\,. \tag{2.6}$$

The following realization of the operators $T_\pm$, $V_\pm$, $U_\pm$, T_3, and Y is obtained:

$$T_+ = a_1^+a_2 - b_2^+b_1, \quad T_- = a_2^+a_1 - b_1^+b_2\,, \tag{2.7a}$$

$$V_+ = a_1^+a_3 - b_3^+b_1, \quad V_- = a_3^+a_1 - b_1^+b_3\,, \tag{2.7b}$$

$$U_+ = a_2^+a_3 - b_3^+b_2, \quad U_- = a_3^+a_2 - b_2^+b_3\,, \tag{2.7c}$$

$$T_3 = \tfrac{1}{2}(a_1^+a_1 - a_2^+a_2 - b_1^+b_1 + b_2^+b_2)\,, \tag{2.7d}$$

$$Y = \tfrac{1}{3}(a_1^+a_1 + a_2^+a_2 - 2a_3^+a_3 - b_1^+b_1$$

$$- b_2^+b_2 + 2b_3^+b_3)\,. \tag{2.7e}$$

The IR space is obtained as follows. Making the identification

$$\Psi_m = (a_1^{+p}b_2^{+q}/\sqrt{p!q!})|0\rangle\,, \tag{2.8}$$

$|0\rangle$ being the vacuum state, one notes that

$$V_-^{p+1}\Psi_m = U_+^{q+1}\Psi_m = 0\,.$$

Also the eigenvalues of T_3 and Y for Ψ_m are given by

$$T_3\Psi_m = [(p+q)/2]\Psi_m\,, \tag{2.9a}$$

$$Y\Psi_m = [(p-q)/3]\Psi_m\,. \tag{2.9b}$$

Further,

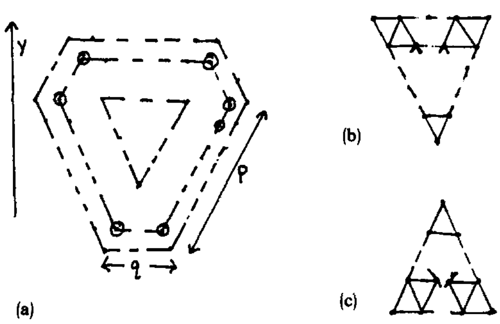


FIG. 1. FDIR's of SU(3). (a) With $p > q = 0$. ⊙ indicates double helicity. (b) With $q = 0$. (c) With $p = 0$.

$$U_3\Psi_m = \tfrac{1}{2}[3Y/2 - T_3]\Psi_m = -(q/2)\Psi_m\,, \quad (2.9c)$$

$$V_3\Psi_m = \tfrac{1}{2}[3Y/2 + T_3]\Psi_m = (p/2)\Psi_m\,. \quad (2.9d)$$

States on the boundary of the IR space consist of only one linear term of the form $a_1^{+r}a_i^{+p-r}b_2^{+q}$ or $a_1^{+p}b_j^{+s}b_2^{+q-s}$ $(i=2,3,\ j=1,3,\ r=1,\ldots,p,\ \ s=1,\ldots,q)$, while states on the next inner layer involve a linear combination of two terms. For instance, operating with V_- on the state $(a_1^{+p}b_3^{+q})/\sqrt{p!q!})\,|0\rangle$, one obtains

$$\Psi_1 = \left[\frac{\sqrt{p}a_1^{+p-1}a_3^{+}p_3^{+q}}{\sqrt{p-1!q!}} - \frac{\sqrt{q}a_1^{+p}b_1^{+}b_3^{+q-1}}{\sqrt{q-1!p!}}\right]|0\rangle. \quad (2.10a)$$

However, operating with T_- on the state $(a_1^{+p}b_2^{+}b_3^{+q-1}/\sqrt{p!q-1!})\,|0\rangle$ [obtained by operation of U_- on $(a_1^{+p}b_3^{+q}/\sqrt{p!q!})\,|0\rangle$], we obtain the state Ψ_2 with the same (T_3,Y) content as Ψ_1, being given by $((p-1)/2,$ $(p+2q)/3-1)$:

$$\Psi_2 = \frac{\sqrt{p}a_2^{+}a_1^{+p-1}b_2^{+}b_3^{+q-1}}{\sqrt{p-1!q-1!}}|0\rangle - \frac{a_1^{+p}b_1^{+}b_3^{+q-1}}{\sqrt{p!q-1!}}|0\rangle. \quad (2.10b)$$

It is easy to see that $\Psi_1 \neq \Psi_2$ and so one can easily check that the layer adjacent to the boundary contains a double multiplicity of states unless either p or q is zero. One chooses linear combinations Ψ_1' and Ψ_2' of Ψ_1 and Ψ_2 characterized by the t values of $(p+1)/2$ and $(p-1)/2$, respectively, i.e., so that

$$T_+\Psi_1' \neq 0, \quad T_+\Psi_2' = 0\,. \quad (2.11)$$

When q or p is zero, the IR space is spanned by the states

$$\frac{a_1^{+i_1}a_2^{+i_2}a_3^{+i_3}}{\sqrt{p!}}|0\rangle \quad \text{or} \quad \frac{b_1^{+j_1}b_2^{+j_2}b_3^{+j_3}}{\sqrt{q!}}|0\rangle\,,$$

where $i_1+i_2+i_3=p$ and $j_1+j_2+j_3=q$, and one of two sets of operators $b_i^+b_j$ or $a_i^+a_j$ $(i,j=1,2,3)$ annihilates the IR space.

In retrospect we collect together the following well-known points regarding the structure of the FDIR space of SU(3) that will have relevance in the context of the IR space of SU(2/1).

(1) $\Psi_{\max}$ is characterized by a U spin value of $q/2$ and a V spin value of $p/2$ [see Eqs. (2.9c) and (2.9d)].

(2) When $p(q)=0$, the $V(U)$ spin content of Ψ_m is zero, and the IR space is elementary with each (T_3,Y) side having single occupancy only.

(3) When both $p,q\neq 0$, the IR is "multilayered," i.e., there are two states on each (T_3,Y) site on the layer adjacent to the boundary, three on the next, and so on until a triangular layer is reached.

(4) This multilayered structure necessitates that we specify each (T_3,Y) site by the eigenvalues of T^2 as well.

B. FDIR space of SU(2/1)

The SU(2/1) or spl(2,1) Lie superalgebra[10] consists of eight generators. The even generators Q_3, $Q_\pm$, and B form the SU(2) $\oplus$ U(1) algebra given below:

$$[Q_3,Q_\pm] = \pm Q_\pm\,, \quad [Q_+,Q_-] = 2Q_3\,, \quad (2.12a)$$

$$[Q_\pm,B] = [Q_3,B] = 0\,. \quad (2.12b)$$

The odd generators V_+, V_-, W_+, and W_- obey the following commutation relations:

$$[Q_\pm,V_\pm] = 0, \quad [Q_\pm,V_\mp] = V_\pm\,, \quad (2.12c)$$

$$[Q_\pm,W_\pm] = 0, \quad [Q_\pm,W_\mp] = W_\pm\,, \quad (2.12d)$$

$$[Q_3,V_\pm] = \pm\tfrac{1}{2}V_\pm\,, \quad [Q_3,W_\pm] = \pm\tfrac{1}{2}W_\pm\,, \quad (2.12e)$$

$$[B,V_\pm] = \tfrac{1}{2}V_\pm\,, \quad [B,W_\pm] = -\tfrac{1}{2}W_\pm\,, \quad (2.12f)$$

$$[V_\pm,W_\pm] = \pm Q_\pm\,, \quad [V_\pm,W_\mp] = -Q_3 \pm B\,, \quad (2.12g)$$

$$[V_\pm,V_\pm] = [V_\pm,V_\mp] = [W_\pm,W_\pm] = [W_\pm,W_\mp] = 0\,. \quad (2.12h)$$

The states of the representation space are taken to be eigenstates of the completely commuting set of operators B, Q^2, and Q_3 and are labeled as $|b,q,q_3\rangle$, where

$$B\,|b,q,q_3\rangle = b\,|b,q,q_3\rangle\,, \quad (2.13a)$$

$$Q^2|b,q,q_3\rangle = q(q+1)(b,q,q_3\rangle\,,$$

$$Q_3|b,q,q_3\rangle = q_3|b,q,q_3\rangle\,. \quad (2.13b)$$

As for SU(3), the maximal state Ψ_m of the FDIR space of SU(2/1) is defined as one with the highest value of Q_3, i.e.,

$$Q_+\Psi_m = V_+\Psi_m = W_+\Psi_m = 0\,, \quad (2.14)$$

as Q_+, V_+, and W_+ are the operators that raise the eigenvalue of Q_3.

From the commutation relations (2.12f) we note that $V_\pm$ raise the eigenvalue b of a representation state by $\frac{1}{2}$, while $W_\pm$ lowers it. Hence if $\Psi = Q_-^m\,\Psi_m$ is that state $|b,q,q_3\rangle$,

$$V_\pm\,|b,q,q_3\rangle = \pm\alpha\sqrt{q\mp q_3}\,|b+\tfrac{1}{2},q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle \quad (2.15a)$$

and

$$W_\pm\,|b,q_1,q_3\rangle = \pm\beta\sqrt{q\mp q_3}\,|b-\tfrac{1}{2},q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle\,, \quad (2.15b)$$

using the facts that $V_\pm$, $W_\pm$ are tensor operators of spin $\frac{1}{2}$ and α and β are constants depending only on b and q.

However, from (2.12h), the representation space cannot accommodate the states whose eigenvalue B is given by $b+1$ or $b-1$, i.e.,

$$V_\pm\,|b+\tfrac{1}{2},q-\tfrac{1}{2},q_3\rangle = W_\pm\,|b-\tfrac{1}{2},q-\tfrac{1}{2},q_3\rangle = 0\,. \quad (2.15c)$$

However,

$$V_\pm\,|b-\tfrac{1}{2},q-\tfrac{1}{2},q_3> = \epsilon\sqrt{q\pm q_3+\tfrac{1}{2}}\,|b,q,q_3\pm\tfrac{1}{2}\rangle \pm \zeta\sqrt{q\mp q_3-\tfrac{1}{2}}\,|b,q-1,q_3\pm\tfrac{1}{2}\rangle\,, \quad (2.16a)$$

and similarly

$$W_\pm\,|b+\tfrac{1}{2},q-\tfrac{1}{2},q_3\rangle = \gamma\sqrt{q\pm q_3+\tfrac{1}{2}}\,|b,q,q_3\pm\tfrac{1}{2}\rangle \pm \delta\sqrt{q\mp q_3-\tfrac{1}{2}}\,|b,q-1,q_3\pm\tfrac{1}{2}\rangle\,, \quad (2.16b)$$

γ, ϵ, δ, and ζ being constants depending on b and q alone.

Hence a state with eigenvalues of B and Q_3 being given by b and $q_3 (q_3 \leqslant q-1)$ has double multiplicity in general.

Finally,

$$V_\pm |b,q-1,q_3\rangle = \tau\sqrt{q\pm q_3}|b+\tfrac{1}{2},q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle , \quad (2.16c)$$

$$W_\pm |b,q-1,q_3\rangle = \omega\sqrt{q\pm q_3}|b-\tfrac{1}{2},q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle , \quad (2.16d)$$

τ and ω being constants depending only on b and q.

It turns out that three of the constants $\alpha, \beta,...,\omega$ are independent. Assuming these to be α, β, and δ, we obtain

$$\gamma = (1/\alpha)(q+b)/2q , \quad (2.17a)$$

$$\epsilon = (1/\beta)(q-b)/2q , \quad (2.17b)$$

$$\zeta = -\alpha\xi/\beta , \quad (2.17c)$$

$$\tau = (1/\delta)(q-b)/2q , \quad (2.17d)$$

$$\omega = -\frac{\beta}{\alpha\delta}\frac{q+b}{2q} = \frac{1}{\zeta}\frac{q+b}{2q} . \quad (2.17e)$$

The FDIR space of SU(2/1) is given in Fig. 2, we shall comment subsequently on its similarity to the FDIR space of SU(3).

The FDIR space becomes "truncated triangular" when the action of either $W_\pm$ or $V_\pm$ on $Q_-^m\ \Psi_m, 0 \leqslant m \leqslant q$, is zero. In this case, the double multiplicity of the full representation space, mentioned above, is absent. It is easy to see when the FDIR space becomes truncated triangular.

If one operates both sides of the second relation (2.12g) on $\Psi_m = |b,q,q\rangle$, one has

$$[V_\pm, W_\mp]|b,q,q\rangle = [-Q_3 \pm B]|b,q,q\rangle$$

$$\Rightarrow V_+W_-|b,q,q\rangle = (-q+b)|b,q,q\rangle$$

$$\text{or } W_+V_-|b,q,q\rangle = (-q-b)|b,q,q\rangle ,$$

using the fact that $V_\pm |b,q,q\rangle = W_+|b,q,q\rangle = 0$. Hence $V_-\Psi_m$ or $W_-\Psi_m$ is zero when $b = -q$ or $b = +q$, respectively.

These "truncated triangular" representations are split parts of the full FDIR space (typical FDIR space) of spl(2,1) and are called "atypical."

The representations of SU(3) and SU(2/1) are remarkably similar, as is to be expected since the two algebras have the same rank, the same number of generators, and a very similar structure. Identifying

$$T_\pm \to Q_\pm ,$$

$$T_3 \to Q_3 ,$$

$$Y \to -B ,$$

$$V_+, U_+ \to V_+, V_-$$

$$U_-, V_- \to W_+, W_- .$$

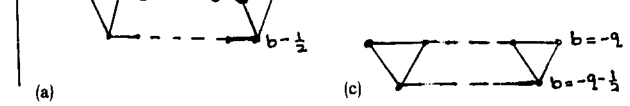


FIG. 2. FDIR's of SU(2/1). (a) Typical FDIR of SU(2/1) with $b \neq \pm q$. ⊙ indicates double helicity. (b) Atypical FDIR of SU(2/1) with $b = q$. (c) Atypical FDIR of SU(2/1) with $b = -q$.

We note that, except for numerical factors occasioned by the commutation relations, the algebras and their representations are identical except for the fact that for the odd generators in spl(2,1), $V_+^2 = W_+^2 = 0$, which prevents them from being applied more than once, and which is responsible for "truncation" of a triangular representation. This is also the reason for no more than double occupancy of a state with eigenvalue of B and Q_3 being given by b and q_3: $|q_3| \leqslant q-1$.

Also we note the fact that atypical representations in the GLA case correspond to elementary representations in the LA case, arising due to the fact that the relevant "V spin" or "W spin" content in the graded case is zero for the particular representation.

Also, as $V_+^2 = W_+^2 = 0$, there is no more than double occupancy for a site on the layer inner to the boundary.

We now find the V spin and W spin content of the maximal state Ψ_m. From (2.15a) and (2.16b),

$$W_+V_-|b,q,q\rangle = \alpha\gamma|b,q,q\rangle . \quad (2.18)$$

Noting that

$$T_+T_-|b,q,q\rangle = 2q|b,q,q\rangle$$

implies that the maximal state Ψ_m has T spin given by q; from (2.17a) we note that the V spin content of $\Psi_m = |b,q,q\rangle$ is $(b+q)/2$.

Similarly the W spin content of Ψ_m is $(b-q)/2$. It is easy to see that when $b = \pm q$, either the V spin or the W spin content of the FDIR is zero, and the FDIR space "splits," i.e., becomes atypical. Thus we see that atypical representations of SU(2/1) correspond to elementary representations of SU(3). It is natural to speculate whether FDIR's of higher rank Lie superalgebras, e.g. SU(n/1), become atypical when the action of one of the odd generators on Ψ_m is zero, so that the "spin" content of that generator is zero. This hope is strengthened by the fact that for osp(1/2), which contains no atypical representations as a special case of the Hochschild–Djokovic theorem[15] (see also Ref. 10), the anticommutation relations of the odd generators read

$$[V_+, V_-] = -\tfrac{1}{2}Q_3 .$$

The action $V_-\Psi_m$ can be zero only for the trivial representation. The same conclusion is arrived at for the osp(1/4) superalgebra if one studies the commutation relations of the same,[18] i.e., the action of the odd generators on the maximal state turn out to be zero only for the trivial representation.

In the next section, we shall obtain FDIR's of SU(3/1) thinking along these lines.

III. FDIR OF THE SU(3/1) SUPERALGEBRA

We begin this section by constructing the SU(3/1) superalgebra (see also Ref. 19). After this we examine the SU(3) content of its FDIR space and comment on the structure of the atypical representations of the superalgebra.

A. Construction of the SU(3/1) superalgebra

The SU(3/1) superalgebra consists of 15 generators, nine generators being even, which form the SU(3) ⊕ U(1) Lie algebra. The rest of the generators, which are six in number, form a $3+3^*$ representation of the even generators. Labeling the even generators as $T_\pm$, $V_\pm$, $U_\pm$, T_3, Y, and B, B being the U(1) generator, we label the two sets of odd generators as $\{X_+, X_-, X_0\}$, and $\{X'_+, X'_-, X'_0\}$, the former forming a 3^* representation of SU(3) and the latter, the 3 representation. The algebra is constructed using one set of three bosonic harmonic oscillators a_i ($i=1,2,3$) and a set of fermionic oscillators b,

$$[a_i,a_j^+]=\delta_{ij}, \quad [b,b^+]=1\,, \tag{3.1}$$

all other commutators (anticommutators) being zero, and the following identification of the generators of the superalgebra:

$$T_+=a_1^+a_2, \quad T_-=a_2^+a_1, \quad V_+=a_3^+a_1, \quad V_-=a_1^+a_3\,,$$
$$U_+=a_2^+a_3, \quad U_-=a_3^+a_2\,; \tag{3.2a}$$
$$T_3=\tfrac{1}{2}(a_1^+a_1-a_2^+a_2), \quad Y=\tfrac{1}{3}(a_1^+a_1+a_2^+a_2-2a_3^+a_3)\,, \tag{3.2b}$$
$$B=\tfrac{1}{3}(a_1^+a_1+a_2^+a_2+a_3^+a_3+3b^+b)\,; \tag{3.2c}$$
$$X_+=a_2b^+, \quad X_-=-a_1b^+, \quad X_0=a_3b^+\,; \tag{3.2d}$$
$$X'_+=a_1^+b, \quad X'_-=a_2^+b, \quad X'_0=a_3^+b\,. \tag{3.2e}$$

It is easy to see that $T_+, T_-, V_+, V_-, U_+, U_-, T_3$, and Y form the familiar SU(3) algebra given in (2.3a) and also that each of these generators commute with B, i.e.,

$$[T_\pm,B]=[V_\pm,B]=[U_\pm,B]$$
$$=[T_3,B]=[Y,B]=0\,. \tag{3.3a}$$

We next check that X_+, X_-, and X_0 form a 3^* representation of SU(3) under the adjoint action, i.e.,

$$[T_\pm,X_\pm]=0, \quad [T_\pm,X_\mp]=X_\pm\,, \tag{3.3b}$$
$$[T_\pm,X_0]=[T_3,X_0]=0, \quad [Y,X_0]=\tfrac{2}{3}X_0\,, \tag{3.3c}$$
$$[T_3,X_\pm]=\pm\tfrac{1}{2}X_\pm\,, \quad [Y,X_\pm]=-\tfrac{1}{3}X_\pm\,, \tag{3.3d}$$
$$[V_\pm,X_\pm]=0=[U_\mp,X_\pm]\,, \tag{3.3e}$$
$$[V_+,X_-]=X_0=-[U_+,X_+]\,, \tag{3.3f}$$
$$[V_-,X_0]=X_-, \quad [U_-,X_0]=-X'_+\,. \tag{3.3g}$$

From (3.3f) and (3.3g), we note that the de Swart phase convention has been built in. We also note that

$$[B,X_\pm]=\tfrac{1}{3}X_\pm\,, \quad [B,X_0]=\tfrac{1}{3}X_0\,. \tag{3.3h}$$

Similarly the commutation (anticommutation) relations of X'_+, X'_-, and X'_0 with the other generators turn out as follows:

$$[T_\pm,X'_\pm]=0=[T_\pm,X'_0]\,, \tag{3.4a}$$
$$[T_3,X'_\pm]=\pm\tfrac{1}{2}X'_\pm\,, \quad [Y,X'_\pm]=\tfrac{1}{3}X'_\pm\,, \tag{3.4b}$$
$$[T_3,X'_0]=0, \quad [Y,X'_0]=-\tfrac{2}{3}X'_0, \tag{3.4c}$$
$$[V_\pm,X'_\pm]=0=[U_\mp,X'_\pm]\,, \tag{3.4d}$$
$$[V_-,X'_+]=X'_0=[U_-,X'_-]\,, \tag{3.4e}$$
$$[V_-,X'_0]=[U_-,X'_0]=0\,, \tag{3.4f}$$
$$[V_+,X'_0]=X'_+\,, \quad [U_+,X'_0]=X'_-\,, \tag{3.4g}$$
$$[B,X'_\pm]=-\tfrac{1}{3}X'_\pm\,, \quad [B,X'_0]=-\tfrac{1}{3}X'_0\,; \tag{3.4h}$$
$$[X_\pm,X'_\mp]=-T_3\pm[Y/2+B],$$
$$[X_0,X'_0]=-Y+B\,, \tag{3.5a}$$
$$[X_\pm,X'_\pm]=\pm T_\pm\,, \quad [X_0,X'_+]=V_+\,,$$
$$[X_0,X'_-]=U_+\,, \tag{3.5b}$$
$$[X'_0,X_+]=U_-, \quad [X'_0,X_-]=-V_-\,, \tag{3.5c}$$
$$[X_\pm,X_\pm]=[X_\pm,X_\mp]=[X_\pm,X_0]=[X'_\pm,X'_\pm]$$
$$=[X'_\pm,X'_0]=[X'_\pm,X'_\mp]=0\,. \tag{3.5d}$$

Note from (3.2) that all the generators have zero supertrace.

B. Structure of the FDIR space of SU(3/1)

We begin by noting that T_+, V_+, U_-, X_+, and X'_+ raise the eigenvalue of T_3 for any state. In addition, we note that the generator X_0 raises the eigenvalue of Y by $\frac{2}{3}$ without changing the eigenvalue of T_3. We define the maximal state Ψ_m to have the maximal value of Y consistent with the maximum eigenvalue of T_3. Hence Ψ_m obeys the relations

$$T_+\Psi_m=V_+\Psi_m=U_-\Psi_m\,, \tag{3.6a}$$
$$X_+\Psi_m=X'_+\Psi_m=X_0\Psi_m\,. \tag{3.6b}$$

We obtain the FDIR space by application of the three lowering generators X_-, X'_-, and X'_0 on the irreducible SU(3) multiplet containing Ψ_m.

If $(j_1,j_2), j_1\geqslant 2, j_2\geqslant 1$, specify the SU(3) multiplet containing Ψ_m, then the (B,T,T_3,Y) content of Ψ_m is

$$\left(b,\frac{j_1+j_2}{2},\frac{j_1+j_2}{2},\frac{j_1-j_2}{3}\right), \quad \text{if } B\Psi_m=b\Psi_m\,. \tag{3.7}$$

It is easy to see that

$$X_-\Psi_m=N_-\Psi_-=N_-\left|\,b+\frac{1}{3},\frac{j_1+j_2-1}{2},\frac{j_1+j_2-1}{2},\right.$$
$$\left.\frac{j_1-j_2-1}{3}\right\rangle, \tag{3.8a}$$
$$X'_0\Psi_m=N'_0\Psi'_0=N'_0\left|\,b-\frac{1}{3},\frac{j_1+j_2}{2},\right.$$
$$\left.\frac{j_1+j_2}{2},\frac{j_1-j_2-2}{3}\right\rangle, \tag{3.8b}$$

N_- and N'_0 being normalization constants depending on b, j_1, and j_2 alone.

Also, as

$$T_+\Psi_-=N_-T_+X_-\Psi_m=N_-[X_-T_+\Psi_m+X_+\Psi_m]=0\,,$$
$$V_+\Psi_-=N_-V_+X_-\Psi_m=N_-[X_-V_+\Psi_m]=0\,,$$
$$U_-\Psi_-=N_-U_-X_-\Psi_m=N_-X_-U_-\Psi_m=0\,,$$

Ψ_- is the maximal state of the SU(3) multiplet characterized by shift operators (j_1-1,j_2).

Similarly, it is easy to see that Ψ'_0 is the maximal state of the SU(3) multiplet with shift operators (j_1-1,j_2+1).

The same does not hold for $X'_-\Psi_m$, even though

$$T_+X'_-\Psi_m=V_+X'_-\Psi_m=0\,,$$
$$U_-X'_-\Psi_m=X'_0\Psi_m\neq 0\,.$$

We may nevertheless write

$$X'_-\Psi_m=\bar{N}'_0U_+X'_0\Psi_m+N'_-\Psi'_-\,, \tag{3.8c}$$

where

$$\Psi'_-:\ U_-\Psi'_-=0\,, \tag{3.9a}$$

$\overline{N}'_0$ and N'_- being normalization constants depending on b, j_1, and j_2 only.

It follows since $T_+X'_-\Psi_m=0=V_+X'_-\Psi_m$, that

$$V_+\Psi'_-=T_+\Psi'_-=0, \tag{3.9b}$$

in addition to (3.9a), and Ψ'_- is the maximal state of a SU(3) multiplet with (B,T,T_3,Y) eigenvalues given by

$$\left(b-\frac{1}{3},\frac{j_1+j_2-1}{2},\frac{j_1+j_2-1}{2},\frac{j_1-j_2+1}{3}\right)$$

and hence characterized by the SU(3) shift operators (j_1,j_2-1).

The representation space is completed using the fact that

$$X_-^2=X_-'^2=X_0'^2=0.$$

Hence

$$X_-\Psi_-=X'_0\Psi_0=0. \tag{3.10a}$$

It also follows that

$$X'_+\Psi'_0\propto X'_+X'_0\Psi_m=-X'_0X'_+\Psi_m=0, \tag{3.10b}$$

$$X_+\Psi'_0\propto X_+X'_0\Psi_m=X'_0X_+\Psi_m+U_-\Psi_m=0. \tag{3.10c}$$

From (3.8a),

$$X'_+X'_-\Psi_m=-\overline{N}'_0U_+X'_0X'_+\Psi_m+N'_-X'_+\Psi'_-$$
$$\Rightarrow -X'_-X'_+\Psi_m=0=X'_+\Psi'_-$$
$$\text{i.e.,}\quad X'_+\Psi'_-=0. \tag{3.10d}$$

From (3.5d) we obtain

$$(X_-X_++X_+X_-)\Psi_m=X_+X_-\Psi_m=0$$
$$\Rightarrow X_+\Psi_-=0, \tag{3.11a}$$

and similarly $X_0\Psi_-=0$. (3.11b)

Hence, from (3.11a), (3.11b), and (3.10a), we note that there exist no states with eigenvalue of B greater than $b+\frac{1}{3}$.

Other IR's of SU(3) contained in the FDIR space of SU(3/1) are as follows:

$$(1)\ X'_0\Psi'_-=N'_{0-}\Psi'_{0-}, \tag{3.12a}$$

N'_{0-} being a normalization constant depending on b, j_1, and j_2 only.

It is easy to check using (3.8c) and (3.10d) that

$$T_+\Psi'_{0-}=V_+\Psi'_{0-}=U_-\Psi'_{0-}=0,$$

and so Ψ'_{0-} is a maximal state of a SU(3) IR with (B,T,T_3,Y) content given by

$$\left(b-\frac{2}{3},\frac{j_1+j_2-1}{2},\frac{j_1+j_2-1}{2},\frac{j_1-j_2-1}{3}\right)$$

and hence characterized by the SU(3) shift operators (j_1,j_2-1). Also Ψ'_{0-} is obtained by the action of X'_- on Ψ'_0 as follows:

$$X'_-\Psi'_0=N'_{-0}\Psi'_{0-}. \tag{3.12b}$$

It follows from (3.10d) that $X'_+\Psi'_{0-}=0$. (3.12c)

(2) Defining normalization constants $\overline{N}_-$ and N_{-0}, depending upon b,j_1, and j_2 alone, one obtains further SU(3) multiplets in the SU(3/1) FDIR as follows:

$$X'_0\Psi_-=\overline{N}_-[V_-,\Psi_m]+N_{-0}\Psi_{-0}, \tag{3.13a}$$

where

$$\Psi_{-0}\colon V_+\Psi_{-0}=0 \tag{3.13b}$$

generates a SU(3) multiplet. It is easy to check that

$$T_+\Psi_{-0}=U_-\Psi_{-0}=0. \tag{3.13c}$$

Thus Ψ_{-0} is a maximal state of the SU(3) IR with (B,T,T_3,Y) content given by

$$\left(b,\frac{j_1+j_2-1}{2},\frac{j_1+j_2-1}{2},\frac{j_1-j_2}{3}-1\right)$$

and hence by the shift operators (j_1-2,j_2+1). It also follows from (3.13a) that

$$X_+\Psi_{-0}=0. \tag{3.13d}$$

(3) The next IR of SU(3) contained in the FDIR's of SU(3/1) is obtained by considering $X'_-X_-\Psi_m$. This state has the (B,T_3,Y) content as

$$\left(b,\frac{j_1+j_2}{2}-1,\frac{j_1-j_2}{3}\right)$$

and can belong to three different SU(3) IR's.

(i) The SU(3) IR generated by Ψ_m itself. Herein, there are two states that occupy this site; we denote the required linear combination as $a_1T_-\Psi_m+a_2\Psi^{(1)}$:

$$T_+\Psi^{(1)}=0.$$

(ii) The SU(3) IR generated by Ψ_{-0}, the required site being $U_+\Psi_{-0}$, which has single occupancy.

(iii) A new multiplet generated by a state we call Ψ_{--}.

Hence

$$X'_-X_-\Psi_m=\overline{N}'_-(a_1T_-\Psi_m+a_2\Psi^{(1)})+\overline{N}'_{-0}U_+\Psi_{-0}$$
$$+N'_{--}\Psi_{--}, \tag{3.14a}$$

where $\overline{N}'_-$, $\overline{N}'_{-0}$, and N'_{--} are normalization constants depending on b,j_1, and j_2 alone, and $T_+\Psi_{--}=0$. (3.14b)

Also $a_1T_-\Psi_m+a_2\Psi^{(1)}$ is a normalized state.

Applying V_+ on both sides of (3.14a), we note that the lhs vanishes as

$$V_+X'_-X_-\Psi_m=X'_-X_0\Psi_m+X'_-X_-V_+\Psi_m=0.$$

On the rhs, $V_+U_+\Psi_{-0}=0$, as Ψ_m and Ψ_{--} belong to different SU(3) multiplets, a_1 and a_2 being determined by the normalized linear combination required, so that[20]

$$V_+[a_1T_-\Psi_m+a_2\Psi^{(1)}]=0. \tag{3.14c}$$

$$\text{Hence } V_+\Psi_{--}=0. \tag{3.14d}$$

Also as

$$U_-[a_1T_-\Psi_m+a_2\Psi^{(1)}]\propto V_-\Psi_m,$$

applying U_- on both sides of (3.14a) it follows using (3.13a), that

$$U_-\Psi_{--}=0.$$

Hence Ψ_{--} is the maximal state of a SU(3) multiplet with (B,T,T_3,Y) content given by

$$\left(b,\frac{j_1+j_2}{2}-1,\frac{j_1+j_2}{2}-1,\frac{j_1-j_2}{3}\right)$$

and described by the shift operators (j_1-1,j_2-1).

(4) The last SU(3) IR contained in the FDIR space of

SU(3/1) is obtained by the action of X'_- on Ψ_{0-} or X'_0 on Ψ_{--}:

$$X'_-\Psi_{-0} = \overline{N}_{--0}[a'_1 T_-\Psi'_0 + a'_2\Psi_0'^{(1)}] + N'_{--0}\Psi_{--0} + \overline{N}'_{--0}V_-\Psi'_- , \tag{3.15a}$$

where

$$T_+\Psi_{--0} = 0 \tag{3.15b}$$

and a'_1, a'_2, $\overline{N}_{--0}$, $\overline{N}'_{--0}$, and N'_{--0} are normalization constants depending on b, j_1, and j_2 above, $a'_1 T_-\Psi'_0 + a'_2\Psi_0'^{(1)}$ being a normalized state.

Applying V_+ on both sides of (3.15a), we obtain

$$0 = \overline{N}_{--0}V_+[a_1 T_-\Psi'_0 + a'_2\Psi_0'^{(1)}] + \overline{N}'_{--0}V_+V_-\Psi'_- + N'_{--0}V_+\Psi_{--0}. \tag{3.15c}$$

We choose a'_1, a'_2: $V_+[a'_1 T_-\Psi'_0 + a'_2\Psi_0'^{(1)}] = 0$ as we did earlier.

Then, it is obvious that $\overline{N}'_{--0} = 0$ if Ψ'_- has nonzero V spin, and that $V_+\Psi_{--0} = 0$ as Ψ_{--0} and Ψ'_- belong to different SU(3) IR's.

Finally, applying U_- on both sides of (3.15a), we obtain, after putting $\overline{N}'_{--0} = 0$,

$$X'_0\Psi_{-0} = \overline{N}_{--0}a''_1 V_-\Psi'_0 + U_-N'_{--0}\Psi_{--0}$$

or, applying X'_0 on both sides of (3.13a), we obtain

$$-\frac{\overline{N}_-}{N_{-0}}V_-X'_0\Psi_m = \overline{N}_{--0}a''_1 V_-\Psi'_0 + U_-N'_{--0}\Psi_{--0}. \tag{3.15d}$$

Again, a''_1 is a normalization constant taking care of the action of U_-. From (3.15d), we note that

$$U_-\Psi_{--0} = 0, \tag{3.15e}$$

there being no other states on either sides of the equation belonging to the same SU(3) IR as Ψ'_{--0} and hence Ψ_{--0} is the maximal state of a SU(3) IR with (B,T,T_3,Y) content given by

$$\left(b - \frac{1}{3}, \frac{j_1+j_2}{2} - 1\ \frac{j_1+j_2}{2} - 1, \frac{j_1-j_2-2}{3}\right)$$

and hence characterized by the SU(3) shift operators (j_1-2, j_2).

The same state can be obtained in a suitable linear combination after applying X'_0 on Ψ_{--}.

In Fig. 3 the action of the SU(3/1) odd generators on the maximal weights of the SU(3) IR's is demonstrated along with the B, T_3, and Y eigenvalues of the maximal and highest states. If one calculates the dimensionality of this IR, it turns out to be $2^{3.1}\times\frac{1}{2}j_1(j_2+1)(j_1+j_2+1)$, which agrees exactly with Ref. 9, as the highest weight of the SU(3/1) FDIR is contained in the SU(3) IR generated by Ψ_- with dimensionality $\frac{1}{2}j_1(j_2+1)(j_1+j_2+1)$. (See Fig. 3.)

C. Analysis of atypical representations

In this section we turn to the main object of this paper, viz., the analysis of atypical FDIR's of SU(3/1).

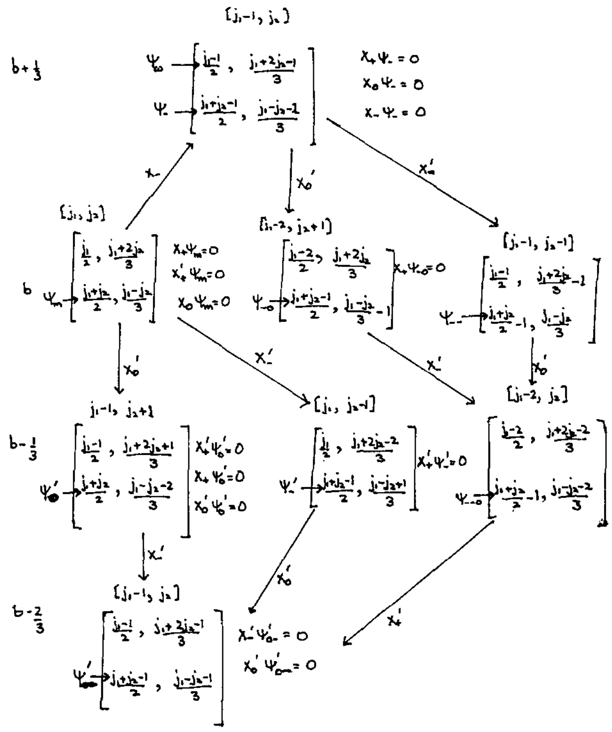


FIG. 3. Typical FDIR's of SU(3/1).

Atypical representations are those IR's of the superalgebra that truncate at some stage, i.e., the product of the odd generators becomes noninvertible, according to Ref. 9. In our analysis presented below, we will find that the decouplings for the independent odd lowering operators take place at the first stage itself, as we find the conditions for the independent operators (to be defined below) to vanish when applied on Ψ_m. For the nonindependent odd lowering operators, the decoupling takes place at a later stage as we shall see. This is a generalization of the SU(2/1) case where both odd lowering operators are independent.

By an independent lowering operator, we mean an operator, the vanishing of whose action does not require simultaneously the vanishing of the action of another lowering operator. We see that the operator X'_- is not one such operator, since

$$X'_-\Psi_m = 0 \tag{3.15a'}$$

implies from (3.8c) that $U_+X'_0\Psi_m = 0$.

Hence $j_2 + 1 = 0$, i.e., $j_2 = -1$, which is impossible. Rewriting (3.15a') as $X'_0U_+\Psi_m = 0$ implies

(1) $U_+\Psi_m = 0$ or (2) $X'_0\Psi_m = 0$.

If $U_+\Psi_m = 0$, $j_2 = 0$, however, in the SU(3) FDIR $\Psi'_0 \propto X'_0\Psi_m$ characterized by the SU(3) shift operators $(j_1-1, 1)$, the state $U_+X'_0\Psi_m$ cannot be obtained. This cannot be allowed; hence we note that $X'_-\Psi_m \neq 0$ independent of $X'_0\Psi_m$.

It may, however, happen that Ψ'_- is not contained in the FDIR space of SU(3/1). This is possible if $j_2 = 0$. However Ψ_{-0} is contained in the representation space, it may decouple if we impose $X'_-\Psi'_0 = 0$. This decoupling and resultant atypicality of the FDIR of SU(3/1) is related with the elementarity of the SU(3) FDIR's containing Ψ_m and will be dealt with subsequently in this subsection.

The case $j_1 = 0$ is not covered by the above case since the vanishing of j_1 is directly related to the vanishing of both odd independent lowering operators $X'_0\Psi_m$ and $X_-\Psi_m$ as we shall see later in this subsection.

Thus our atypicality conditions are

(1) $X_-\Psi_m = 0 \Rightarrow [X'_+, X_-]\Psi_m = 0$,

(2) $X'_0\Psi_m = 0 \Rightarrow [X_0, X'_0]\Psi_m = 0$,

(3) (a) $X'_-\Psi'_0 = 0$ in conjunction with elementarity of Ψ_m, or
(b) $X'_-\Psi_m = 0 \Rightarrow [X_+, X'_-]\Psi_m = 0$ in conjunction with (2).

We now obtain the atypical representations of SU(3/1).

Case 1:

$$X_-\Psi_m = 0 \Rightarrow [X'_+, X_-]\Psi_m = 0 \tag{3.16a}$$

$$\Rightarrow [-T_3 - Y/2 - B]\Psi_m = 0$$

$$\text{i.e., } b = -(2j_1 + j_2)/3. \tag{3.16b}$$

In this case Ψ_- and the SU(3) IR's generated by it, namely, Ψ_{-0}, Ψ_{--}, and Ψ_{-0}, are not contained in the FDIR space that contains only the SU(3) multiplets generated by Ψ_m, Ψ'_0, Ψ'_-, and Ψ'_{0-}, pictorially represented in Fig. 4(a).

(i) If, in addition to (3.16a),

$$X'_0\Psi_m = 0, \tag{3.17a}$$

then

$$[X_0, X'_0]\Psi_m = 0, \text{ i.e., } b = (j_1 - j_2)/3. \tag{3.17b}$$

Combining (3.17b) and (3.16b), we see that

$$j_1 = 0.$$

Hence the FDIR space of SU(3/1) reduces to the two SU(3) IR's [denoted by (b, j_1, j_2)] $(-j_2/3, 0, j_2)$ and $(j_2/3 - \frac{1}{3}, 0, j_2 - 1)$, pictorially represented in Fig. 4(b). This representation corresponds to the elementary representation $j_3\Lambda_3$ of SU(4), Λ_3 being the 4* IR's of SU(4) (see the Appendix).

(ii) If, in addition to (3.16a), $j_1 = 1, j_2 = 0$, then Ψ'_- is not contained in the FDIR space of SU(3/1). Thus (3.8c) reduces to

$$X'_-\Psi_m = N_0 U_+ X'_0 \Psi_m \tag{3.17c}$$

since Ψ'_- is not contained in the FDIR space of SU(3/1). The latter reduces to the two components $(\frac{1}{3}, 1, 0)$ and $(0, 0, 1)$ corresponding to the six-dimensional elementary IR's of SU(4), Λ_2 [see Fig. 4(c) and Appendix]. When $j_1 > 1$ one obtains the atypical IR's with the SU(3) IR's generated by Ψ_m and Ψ'_0 characterized by shift operators $(j_1, 0)$ and $(j_1 - 1, 1)$, respectively.

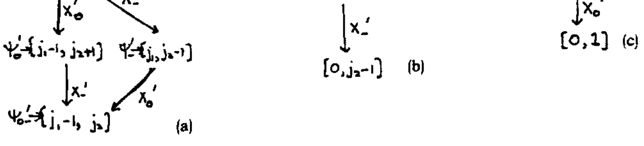


FIG. 4. Atypical FDIR's of SU(3/1) when $x - \Psi_m = 0$.

We do not obtain an exact analog for the representation $j_2\Lambda_2$, $j_2 > 1$, since the odd generators cannot be applied more than once.

(iii) If, in addition to (3.16a),

$$X'_-\Psi_m = 0, \tag{3.18a}$$

then

$$[X_+, X'_-]\Psi_m = 0$$

$$\Rightarrow \left[-T_3 + \frac{Y}{2} + B\right]\Psi_m = 0, \text{ i.e., } b = \frac{j_1 + 2j_2}{3}. \tag{3.18b}$$

Combining (3.16b) with (3.18b), we see that

$$j_1 = -j_2, \tag{3.18c}$$

which can be satisfied only if $j_1 = j_2 = 0$. Hence the FDIR space of SU(3/1) is trivial, making $X'_0\Psi_m = 0$. Hence $X'_-\Psi_m$ is not zero, independent of $X'_0\Psi_m$.

Case 2:

$$X'_0\Psi_m = 0 \Rightarrow [X_0, X'_0]\Psi_m = 0 \tag{3.19a}$$

$$\Rightarrow [-Y + B]\Psi_m = 0, \text{ i.e., } b = (j_1 - j_2)/3. \tag{3.19b}$$

Hence (3.8c) reduces to

$$X'_-\Psi_m = N'_-\Psi_-. \tag{3.19c}$$

From (3.13a), we note that

$$X'_0\Psi_- \propto X'_0 X_-\Psi_m = [X'_0, X_-]\Psi_m = V_-\Psi_m$$

using (3.19a), implying that the SU(3) IR generated by Ψ_{-0} is not contained in the FDIR space of SU(3/1). Similarly, as

$$\Psi'_{0-} \propto X'_0 X'_-\Psi_m = -X'_- X'_0\Psi_m = 0,$$

Ψ'_{0-} is also not contained in the FDIR space of SU(3/1).

Finally, we note that Ψ'_{--0} is also not contained in the FDIR space if (3.19a) holds. Applying X'_0 on both sides of (3.14), we obtain

$$X'_0 X'_- X_-\Psi_m = X'_0 N_-[a_1 T_-\Psi_m + a_2\Psi^{(1)}]$$

$$+ N'_{--} X'_0 \Psi_{--}. \tag{3.20a}$$

Here N'_{-0} has been set equal to zero since Ψ_{-0} is not contained in the FDIR space. Using the commutation relations (3.5c) of SU(3/1), (3.20a) yields, using (3.19a),

$$X'_- V_-\Psi_m = N_- a_2 X'_0\Psi^{(1)} + N'_{--} X'_0\Psi_{--}. \tag{3.20b}$$

We now note that $X'_0\Psi^{(1)} \propto X'_- V_-\Psi_m$.

From (3.19a),

$$V_- X'_0\Psi_m = X'_0 V_-\Psi_m = 0. \tag{3.20c}$$

Now $U_+ V_-\Psi_m$ contains the two states $\Psi^{(1)}$ and $T_-\Psi_m$:

$$X'_0 T_-\Psi_m = T_- X'_0\Psi_m = 0,$$

$$X'_0 U_+ V_-\Psi_m = +X'_- V_-\Psi_m + U_+ X'_0 V_-\Psi_m$$

$$= X'_- V_-\Psi_m \quad [\text{using (3.20c)}].$$

$X'_0\Psi^{(1)}$ is contained in $X'_-\ V_-\Psi_m$, i.e., belongs to the SU(3) IR's generated by Ψ'_-. Hence $X'_0\Psi_{--}$ is contained only in Ψ'_-, and Ψ_{--0} is not contained in the FDIR space of SU(3/1).

The full atypical IR space shown in Fig. 5(a) is thus

$$(b,j_1,j_2),\ (b,j_1-1,j_2-1),$$
$$(b+\tfrac{1}{3},j_1-1,j_2),\ (b-\tfrac{1}{3},j_1,j_2-1),$$

where $b=(j_1-j_2)/3$. It contains four SU(3) IR's.

The adjoint representation of SU(3/1) characterized by (0,1,1) is an example of such an atypical representation. It contains the (b,j_1,j_2) components given below: $(0,1,1)\oplus(0,0,0)$ constitute the even part of the superalgebra SU(3) $\oplus B$. The odd generators $(X_\pm,X_0)$ belong to the SU(3) IR $(\tfrac{1}{3},0,1)$, while the generators $(X'_\pm,X'_0)$ belong to the SU(3) IR$(-\tfrac{1}{3},1,0)$.

(i) If, in addition to (3.19a),

$$X'_-\Psi_m=0, \tag{3.21a}$$

we obtain

$$[X_+,X'_-]\Psi_m=0\Rightarrow[-T_3+Y/2+B]\Psi_m=0,$$
$$\text{i.e., } b=(j_1+2j_2)/3. \tag{3.21b}$$

Comparing (3.19b) and (3.21b), one obtains

$$j_2=0,$$

i.e., the atypical representation contains only the components $(j_1/3,j_1,0)$ and $((j_1+1)/3,j_1-1,0)$ corresponding to $j_1\Lambda_1$, Λ_1 being the elementary four-dimensional IR's of SU(4) [see the Appendix and Fig. 5(b)].

We note that in this case Ψ_{-0} cannot be contained in the IR space. This may be seen by applying the second relation of (3.5c) and (3.13a) on Ψ_m, which immediately yields $N_{-0}=0$, showing that Ψ_{-0} is not contained in the IR space.

Case 3:

(1) $X'_-\Psi'_0=0$ in connection with $j_2=0$.

(2) $X'_-\Psi_m=0$ in connection with cases 1 and 2.

The second condition has been already dealt with in cases 1 and 2; we only deal with the first condition in what follows. We note, however, when Ψ_m is characterized by $j_2=0$, b arbitrary, the FDIR space of SU(3/1) does not contain the SU(3) IR generated by Ψ'_- and Ψ_{--} (Fig. 6) as j_2-1 becomes negative. This case is not analogous to $j_1=0$, however, because $j_1=0$ implies that the action of both independent operators X_- and X'_0 on Ψ_m become zero automatically.

The resultant representation becomes identical to the atypical representation discussed in case 1(i). The condition

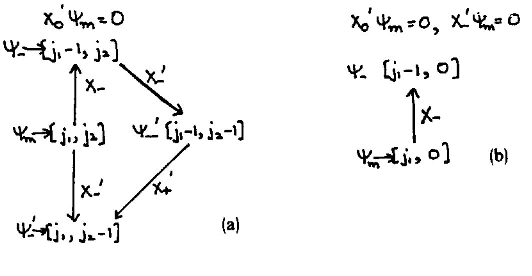


FIG. 5. Atypical FDIR's of SU(3/1) when $X'_0X_m=0$.

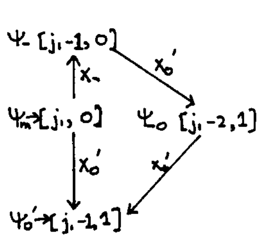


FIG. 6. Atypical FDIR's of SU(3/1) when $j_2=0$, $x'_-\Psi'_0=0$.

$j_2=0$ does not entail any such restriction even on the dependent operator X'_-, i.e., $X'_-\Psi_m\neq0$.

(i) The SU(3/1) FDIR's with Ψ_m generating a SU(3) IR with $j_2=0$ becomes atypical if

$$X'_-\Psi'_0=0, \tag{3.22a}$$

i.e.,

$$[X_+,X'_-]\Psi'_0=0\Rightarrow[-T_3+Y/2+B]\Psi'_0=0,$$

i.e.,

$$b=(j_1+2)/3. \tag{3.22b}$$

Hence Ψ'_{0-} is also not contained in this atypical IR in addition to Ψ'_- and Ψ_{--}.

It so happens that imposition of (3.22a) results in the atypical representation consisting of the SU(3) IR's generated by Ψ_m, Ψ'_0, Ψ_-, and Ψ_{-0}. It is a rather involved calculation to demonstrate explicitly that Ψ_{--0} drops out. Here one uses an intuitive argument, viz., that the boundary states of the SU(3/1) FDIR's have unit SU(3) multiplicity corresponding to unit SU(3) multiplicity of the boundary states of SU(4) and unit SU(2) multiplicity of the boundary states of SU(2/1) and SU(3). Hence Ψ'_0 for this FDIR is the sole occupant of the floor with eigenvalue of B, $b=(j_1+1)/3$. This also gives the FDIR, a symmetry under a certain rotation about the B axis [corresponding to 120° rotation for the SU(3) LA]. The explicit computation of the matrix elements of $X_\pm$, $X'_\pm$, X_0, and X'_0 is left to a later communication.

IV. COMPARATIVE STUDY OF FDIR OF SU(4) AND SU(3/1)

The SU(4) LA and the structure of its FDIR are discussed in the Appendix. We note the following equivalence among the non-SU(3) generators of both algebras:

$$T_{41},T_{42},T_{43}\leftrightarrow(X_-,X_+,X_0),$$
$$T_{14},T_{24},T_{34}\leftrightarrow(X'_+,X'_-,X'_0),$$
$$-T_4\leftrightarrow B,$$

apart from numerical factors between the generators T_4 and B.

Because of the subsidiary condition $X_-^2=X'^2_-=X'^2_0=0$, the lowering operators of SU(3/1) cannot be applied more than once.

The points of similarity between the FDIR's of SU(4) and SU(3/1) are obvious. Like T_{41}, X_- reduces the value of j_1 for the SU(3) IR generated by Ψ_m, it cannot be applied more than once.

Like T_{43}, X'_0 reduces the value of j_1 by one unit and raises j_2 by the same amount; it cannot be applied more than once.

Finally X'_- operates on the state $X_0\,\Psi_m$ to yield a SU(3) IR characterized by (j_1, j_2-1); again it cannot be applied more than once.

Further, for the SU(4) IR's under consideration, the floor containing Ψ_m contains three SU(3) IR's, one generated by the vector with the same (T_2,T_3,T_4) content as $T_-\,\Psi_m$ and another with the same content as $V_-\,\Psi_m$ corresponding to the products $T_{41}T_{24}\,\Psi_m$ and $T_{41}T_{34}\,\Psi_m$, respectively. The multiplicity as one goes inward goes on increasing in the upgraded case; for the graded case it stops at the first stage.

For the next stage in the graded case, the (T_2,T_3,B) content of the SU(3) FDIR's (occupying the same floor as $X'_0\,\Psi_m$) is the same as that of $X'_0\,\Psi_m$, $T_-\,X'_0\,\Psi_m$, and $U_+X'_0\,\Psi_m$ corresponding to the products of the odd generators $X_-\,X'_-\,X'_0\,\Psi_m$ and $X'_-\,\Psi_m$, respectively.

It transpires that the atypical FDIR's in the GLA case are analogous to the FDIR's in the LA case when one of the shift operators j_1 or j_2 is zero. This is seen as follows.

(1) $X_-\,\Psi_m=0 \leftrightarrow T_{41}\,\Psi_m=0$, i.e., $j_1=0$.

The SU(3) content of the atypical FDIR's of SU(3/1) is shown alongside the max SU(3) content of SU(4) FDIR's [characterized by shift operators (j_1,j_2,j_3)] below:

$$
B\uparrow\quad
\begin{tikzcd}[column sep=small]
& \mathrm{SU}(3/1) & \\
& \Psi_m=(j_1,j_2) \arrow[d, "X'_0"] \arrow[dr, "X'_-"] & \\
& (j_1-1,j_2+1) \arrow[d, "X'_-"] & (j_1,j_2-1) \arrow[dl, "X'_0"] \\
& (j_1-1,j_2) &
\end{tikzcd},
$$

$$
-T_4\uparrow\quad
\begin{tikzcd}
\mathrm{SU}(4) \\
\Psi_m=(j_2,j_3) \arrow[d, "T_{34}^{j_2}"] \\
(0,j_2+j_3)+\text{other SU(3) IR's} \arrow[d, "T_{24}^{j_3}"] \\
(0,j_2)
\end{tikzcd}.
$$

(2) $X'_0\,\Psi_m=0 \leftrightarrow T_{34}\,\Psi_m=0$, i.e., $j_2=0$.

The SU(3) content of the atypical FDIR's of SU(3/1) is shown alongside that of SU(4) in the following diagrams:

$$
B\uparrow\quad
\begin{tikzcd}[column sep=small]
\mathrm{SU}(3/1) & \\
(j_1-1,j_2) \arrow[dr, "X'_-"] & \\
\Psi_m=(j_1,j_2) \arrow[u, "X_-"] \arrow[d, "X'_-"] & (j_1-1,j_2-1) \arrow[dl, "X'_+"] \\
(j_1,j_2-1) &
\end{tikzcd},
$$

$$
-T_4\uparrow\quad
\begin{tikzcd}
\mathrm{SU}(4) \\
(0,j_3) \\
\Psi_m=(j_1,j_3)+\text{other SU(3) IR's} \arrow[u, "T_{41}^{j_1}"] \arrow[d, "T_{24}^{j_3}"] \\
(j_1,0)
\end{tikzcd}.
$$

(3) $X'_-\,\Psi'_0=0$, $j_2=0 \leftrightarrow T_{32}\,\Psi_m=0$, $T_{24}\,\Psi'_2=0$, i.e., $j_3=0$.

This is the most interesting case. Since there is no analog of j_3 in the GLA case, the additional condition $X'_-\,\Psi'_0=0$ has to be imposed. The SU(3) FDIR's content of SU(3/1) versus SU(4) is shown in the following diagrams:

$$
B\uparrow\quad
\begin{tikzcd}[column sep=small]
\mathrm{SU}(3/1) & \\
(j_1-1,0) \arrow[dr, "X'_0"] & \\
\Psi_m=(j_1,0) \arrow[u, "X_-"] \arrow[d, "X'_0"] & (j_1-2,1) \arrow[dl, "X'_+"] \\
(j_1-1,1) &
\end{tikzcd},
$$

$$
-T_4\uparrow\quad
\begin{tikzcd}
\mathrm{SU}(4) \\
(j_2,0) \\
\Psi_m=(j_1+j_2,0)+\text{other SU(3) IR's} \arrow[u, "T_{41}^{j_1}"] \arrow[d, "T_{34}^{j_2}"] \\
(j_1,j_2)
\end{tikzcd}.
$$

In the case of SU(4), $j_3=0$, i.e., elementarity of the SU(3) IR's generated by Ψ_m is sufficient to ensure that $T_{42}\,\Psi'_2=0$. This is related to the fact that T_4 can have only eigenvalues proportional to $p/4$, p being an integer. However, B need obey no such restriction so when one drops the condition $X'_-\,\Psi'_0=0$, one obtains in the graded case a FDIR with no analog in the ungraded case. The condition $X'_-\,\Psi'_0=0$ must be imposed to set the eigenvalue of B at the requisite rational analog of $p/4$.

The similarity in the representation content of the LA versus the GLA case is subject to the restriction that any change in the shift operator content (j_1,j_2) of Ψ_m in the LA case, although accompanied by a corresponding reduction in the GLA case, cannot in the latter take place by more than one unit.

Finally, a few words are in order regarding a comparison between the elementary IR's of SU(3/1) versus those of SU(4):

(1) $X_-\Psi_m=X'_0\,\Psi_m=0 \quad\leftrightarrow\quad T_{41}\Psi_m=T_{34}\Psi_m=0$,

i.e., $j_1=0$, $b=-j_2/3$, $\qquad$ i.e., $j_1=j_2=0$,

$$
B\uparrow\quad
\begin{tikzcd}
\mathrm{SU}(3/1) \\
\Psi_m=(0,j_2) \arrow[d, "X'_-"] \\
(0,j_2-1)
\end{tikzcd}
\qquad
-T_4\uparrow\quad
\begin{tikzcd}
\mathrm{SU}(4) \\
\Psi_m=(0,j_3) \arrow[d, "T_{24}^{j_3}"] \\
(0,0)
\end{tikzcd};
$$

$$
\begin{array}{lccl}
 & (2)\ X'_{-}\Psi_m = X_0\Psi_m = 0 & \leftrightarrow & T_{24}\Psi_m = 0 = T_{34}\Psi_m \\
 & \text{i.e., } j_2 = 0,\quad b = j_1/3 & -T_4 \uparrow & \Rightarrow j_2 = 0 = j_3 = 0 \\
B \uparrow & SU(3/1) & & SU(4) \\
 & (j_1 - 1,0) & & (0,0) \\
 & \uparrow X_{-} & & \uparrow T_{41}^{j_1} \\
 & \Psi_m = (j_1,0) & & \Psi_m = (j_1,0)
\end{array} ;
$$

$$
\begin{array}{lccl}
 & (3)\ j_1 = 1,\quad j_2 = 0,\quad \text{and } X_{-}\Psi_m = 0 & \leftrightarrow & T_{41}\Psi_m = 0 = T_{23}\Psi_m \\
 & & -T_4 \uparrow & \text{i.e., } j_1 = j_3 = 0 \\
B \uparrow & SU(3/1) & & SU(4) \\
 & \Psi_m = (1,0) & & \Psi_m = (j_2,0) \\
 & \downarrow X'_0 & & \downarrow T_{34}^{j_2} \\
 & (0,1) & & (0,j_2)
\end{array} .
$$

In the graded case there is no analog of $j_2\Lambda_2$ [Λ_2 being the six-dimensional elementary IR of SU(4), $j_2 > 1$].

V. CONCLUSION

We conclude that typical FDIR's of SU(3/1) are analogous to FDIR's of SU(4) with all shift operators j_1, j_2, j_3 nonzero. Atypical FDIR's are like those of FDIR's of SU(4) for which one or more of the shift operators are zero, and have a similarly reduced structure.

Our atypicality conditions read as follows:

$$(1)\ X_{-}\Psi_m = 0 \Rightarrow [X'_{+}, X_{-}]\Psi_m = 0, \tag{5.1}$$

$$(2)\ X'_0\Psi_m = 0 \Rightarrow [X'_0, X_0]\Psi_m = 0, \tag{5.2}$$

$$(3)\ X'_{-}\Psi'_0 = 0 \quad \text{in conjunction with } j_2 = 0. \tag{5.3}$$

Some comments on the philosophy of this technique may be in order at this stage. Instead of starting conventionally from the highest state Ψ_w, one starts from a state midway between the highest and the lowest.[21] Two advantages of this technique that deserve mention are as follows.

(1) Operation of the odd lowering operators on a maximal state of a SU(3) FDIR yields another maximal state of new SU(3) IR's, making it relatively easy to determine the shift operator content. This is not so for the highest state, as is evident from Fig. 3, since the operation of X'_0 on the highest state of SU(3/1) characterized by

$$(T_2, T_3, B) = \left(\frac{j_1 - 1}{2}, \frac{j_1 + 2j_2 - 1}{3}, b + \frac{1}{3}\right)$$

does not yield the highest state of $X'_0\Psi_{-}$.

(2) Starting from some "in between" state makes it easier to assess the various spin contents of the FDIR's and hence easier to determine when the FDIR's get truncated without any knowledge of the matrix elements of the operators, which is the problem under investigation.

It must also be mentioned that this technique does not show explicitly why some SU(3) IR's drop out, e.g., in the third case, why Ψ_{--0} is truncated and one then has to resort to symmetry arguments as we did to "see" the SU(3) content. This may get complicated for higher n in a generalization to SU(n/1).

Besides, $j_2 = 0$ for Ψ_m in SU(3/1) does not automatically, in the third atypicality case, imply that Ψ_{--0} drops out, unless $X'_{-}\Psi'_0 = 0$ is imposed, which sets a restriction on the eigenvalue of B. This is not required in the SU(4) case as the range of T_4 is fixed to be of the form $p/4$, p being an integer. In the other two atypicality conditions, the restriction on B is automatically imposed and no other restrictions are necessary.

We plan to generalize the results of this paper to SU(n/1) in a forthcoming work.

APPENDIX: FDIR SPACE OF SU(4)

The SU(4) Lie algebra is constructed by the use of the four sets of bosonic harmonic oscillator creation and annihilation operators a_i, a_j^+:

$$[a_i, a_j^+] = \delta_{ij}, \quad i,j = 1,\ldots,4.$$

The SU(3) operators are

$$T_{ij} = a_i^+ a_j, \quad i \neq j = 1,\ldots,4, \tag{A1a}$$

$$T_i = \frac{1}{i}\left[\sum_{k=1}^{i-1} a_k^+ a_k - (i-1)a_i^+ a_i\right], \quad i = 2,\ldots,4. \tag{A1b}$$

The SU(4) LA is obtained as follows:

$$
\begin{aligned}
[T_{ij}, T_{mn}] &= [a_i^+ a_j, a_m^+ a_n] \\
&= \delta_{jm} T_{in} - \delta_{in} T_{mj},
\end{aligned} \tag{A2a}
$$

unless $i = n$, $m = j$ simultaneously.

$$[T_j, T_{im}] = 0, \quad j < i < m, \tag{A2b}$$

$$= -[(j-1)/j]T_{im}, \quad i = j < m, \tag{A2c}$$

$$= (1/j)T_{im}, \quad i < j < m, \tag{A2d}$$

$$= 0, \quad i < m < j, \tag{A2e}$$

$$= 0, \quad j < m < i, \tag{A2f}$$

$$= [(j-1)/j]T_{im}, \quad m = j < i, \tag{A2g}$$

$$= -(1/j)T_{im}, \quad m < j < i, \tag{A2h}$$

$$= 0, \quad m < i < j, \tag{A2i}$$

$$
\begin{aligned}
[T_{ij}, T_{ji}] &= -T_i + T_{i+1}/i \\
&\quad + \cdots + jT_j/(j-1), \quad i \leqslant 2,
\end{aligned} \tag{A2j}
$$

$$[T_{ij}, T_{jl}] = T_2 + \cdots + jT_j/(j-1). \tag{A2k}$$

From (A2b)–(A2i) it is seen that T_{12}, T_{13}, T_{32}, T_{14}, and T_{42} are the operators that raise the eigenvalue of T_2, T_{13}, T_{23}, T_{14}, and T_{24} are the operators that raise the eigenvalue of T_3; and T_{14}, T_{24}, and T_{34} are the operators that raise the eigenvalue of T_4. We define our maximal state Ψ_m as one for which

$$T_{1j}\Psi_m = 0, \quad j>1, \tag{A3a}$$

$$T_{ij}\Psi_m = 0, \quad i\geqslant j\geqslant 2. \tag{A3b}$$

The SU(4) FDIR space is generated by the action of the generators T_{41}, T_{24}, and T_{34} on Ψ_m and the SU(3) FDIR's generated by Ψ_m.

Before we obtain the FDIR space of SU(4), it is convenient to examine the elementary IR's of SU(4). These are three in number.

(1) Λ_1: For this IR, in addition to (A3a) and (A3b),

$$T_{ij}\Psi_m = 0, \quad i<j. \tag{A4a}$$

In this case Ψ_m is the maximal as well as the highest state of the SU(4) IR. The IR, characterized by only one shift operator, has four states, Ψ_m, $T_{21}\Psi_m$, $T_{31}\Psi_m$, and $T_{41}\Psi_m$. Now Ψ_m, $T_{21}\Psi_m$, and $T_{31}\Psi_m$ form a 3 representation of SU(3), while $T_{41}\Psi_m$ forms a 1 representation of SU(3). Hence the SU(3) (j_1,j_2) content of the 4 representation of SU(4) is given by

$$4 = (1,0) \oplus (0,0), \tag{A4b}$$

$$T_4\Psi_m = \tfrac{1}{4}\Psi_m, \tag{A4c}$$

$$T_4[T_{41}\Psi_m] = -\tfrac{3}{4}T_{41}\Psi_m. \tag{A4d}$$

(2) Λ_3: For this IR, in addition to (A3a) and (A3b),

$$T_{j1}\Psi_m = 0, \quad j\geqslant 3, \tag{A5a}$$

$$T_{jk}\Psi_m = 0, \quad 3=j<k. \tag{A5b}$$

This IR, which we denote by 4*, contains the four states Ψ_m, $T_{24}\Psi_m$, $T_{23}\Psi_m$, and $T_{21}\Psi_m$. The SU(3) shift operator content of 4* is given by

$$4^* = (0,1) \oplus (0,0). \tag{A5c}$$

Also

$$T_4\Psi_m = -\tfrac{1}{4}\Psi_m, \tag{A5d}$$

$$T_4[T_{24}\Psi_m] = \tfrac{3}{4}T_{24}\Psi_m. \tag{A5e}$$

This IR is the conjugate representation to Λ_1.

(3) Λ_2: This IR is conjugate to itself. In addition to (A3a) and (A3b), the following conditions on Ψ_m are imposed:

$$T_{41}\Psi_m = 0, \tag{A6a}$$

$$T_{23}\Psi_m = 0. \tag{A6b}$$

This IR consists of the six states Ψ_m, $T_{21}\Psi_m$, $T_{31}\Psi_m$, $T_{24}\Psi_m$, $T_{34}\Psi_m$, and $T_{21}T_{34}\Psi_m$. We denote it by 6. The 6 IR has SU(3) content given by

$$6 = (1,0) \oplus (0,1). \tag{A6c}$$

Also

$$T_4\Psi_m = -\tfrac{1}{2}\Psi_m, \tag{A6d}$$

$$T_4[T_{34}\Psi_m] = \tfrac{1}{2}[T_{34}\Psi_m]. \tag{A6e}$$

The FDIR space of SU(4) in general contains a combination of all the 3 Λ_i's, applied j_i times, the j_i being integers: $j_i\geqslant 0$, $i=1,2,3$. Hence T_{41} can be applied at most j_1 times on Ψ_m, T_{32} can be applied j_3 times on Ψ_m, and T_{34} can be applied at most j_2 times on Ψ_m.

The last statement follows from the fact that one basis has been chosen so that the operators T_{41}, T_{34}, and T_{23} are independent, i.e.,

$$T_{41}|\Lambda_3\rangle = T_{34}|\Lambda_3\rangle = 0,$$

$$T_{34}|\Lambda_1\rangle = T_{23}|\Lambda_1\rangle = 0,$$

$$T_{41}|\Lambda_2\rangle = T_{23}|\Lambda_2\rangle = 0,$$

$|\Lambda_i\rangle$ being the maximal state of Λ_i. Also we note that

$$[T_{14},T_{41}]\Psi_m = [T_2 + T_3/2 + \tfrac{4}{3}T_4]\Psi_m = j_1\Psi_m, \tag{A7a}$$

$$[T_{34},T_{43}]\Psi_m = [-T_3 + \tfrac{4}{3}T_4]\Psi_m = -j_2\Psi_m, \tag{A7b}$$

$$[T_{23},T_{32}]\Psi_m = [-T_2 + \tfrac{3}{2}T_3]\Psi_m = -j_3\Psi_m. \tag{A7c}$$

Hence

$$\begin{aligned}[T_{13},T_{31}]\Psi_m &= [T_2 + 3T_3/2]\Psi_m \\ &= [[T_{41},T_{41}] - [T_{34},T_{43}]]\Psi_m \\ &= (j_1+j_2)\Psi_m,\end{aligned} \tag{A7d}$$

and so, as T_{31} and T_{23} can be operated on Ψ_m, (j_1+j_2) and j_3 times, respectively, the SU(3) IR generated by Ψ_m is characterized by the shift operators (j_1+j_2,j_3).

Let us consider the states

$$\Psi_1' = T_{41}^{j_1}\Psi_m, \tag{A8a}$$

$$\Psi_2' = T_{34}^{j_2}\Psi_m, \tag{A8b}$$

$$\Psi_3' = T_{24}^{j_3}\Psi_2'. \tag{A8c}$$

It is easy to check that Ψ_1', Ψ_2', Ψ_3' are the maximal states of the SU(3) IR's they generate, which are characterized by (j_2,j_3), (j_1,j_2+j_3), and (j_1,j_2), respectively.

The FDIR space of SU(4) characterized by (j_1,j_2,j_3) has the maximum SU(3) content ranging as given in the following diagram. The expression maximum SU(3) content is used, because there are more SU(3) IR's in each floor. The maximum SU(3) IR is characterized by the highest value of j_1:

$$T_4 \uparrow \quad \begin{array}{c} \Psi_3' \to (j_1,j_2) \\ \uparrow T_{24}^{j_3} \\ \Psi_2' \to (j_1,j_2+j_3) \\ \uparrow T_{34}^{j_2} \\ \Psi_m \to (j_1+j_2,j_3) \\ \downarrow T_{41}^{j_1} \\ \Psi_1' \to (j_2,j_3) \end{array}.$$

We now reobtain the elementary IR's of SU(4) as follows.

(1) If $j_1=0$, the SU(4) IR contains the SU(3) IR's ranging as shown below:

$$T_4 \uparrow \quad \begin{array}{c} \Psi_3' \to (0,j_2) \\ \uparrow T_{24}^{j_3} \\ \Psi_2' \to (0,j_2+j_3) \\ \uparrow T_{34}^{j_2} \\ \Psi_m \to (j_2,j_3) \end{array}.$$

(2) If $j_3 = 0$, the IR contains the SU(3) IR's ranging as shown below:

$$\begin{array}{l} (j_1,j_2) \\ \uparrow T_{34}^{j_2} \\ \Psi_m \to (j_1+j_2,0)\,. \\ \downarrow T_{41}^{j_1} \\ (j_2,0) \end{array}$$

(3) If $j_2 = 0$, the FDIR of SU(4) contains SU(3) IR's ranging as follows:

$$\begin{array}{l} (j_1,0) \\ \uparrow T_{24}^{j_3} \\ \Psi_m \to (j_1,j_3)\,. \\ \downarrow T_{41}^{j_1} \\ (0,j_3) \end{array}$$

(4) If $j_2 = j_3 = 0$, the SU(4) IR contains SU(3) IR's ranging as shown below, yielding the FDIR $j_1\Lambda_1$:

$$\begin{array}{l} \Psi_m \to (j_1,0) \\ \downarrow T_{41}^{j_1}\quad . \\ (0,0) \end{array}$$

(5) If $j_1 = j_3 = 0$, the FDIR contains SU(3) IR's ranging as shown below and yielding the FDIR $j_2\Lambda_2$:

$$\begin{array}{l} (0,j_2) \\ \uparrow T_{34}^{j_2}\quad . \\ \Psi_m \to (j_2,0) \end{array}$$

(6) If $j_1 = j_2 = 0$ the FDIR space contains SU(3) IR's ranging as shown below yielding the IR $j_3\Lambda_3$:

$$\begin{array}{l} (0,0) \\ \uparrow T_{24}^{j_3}\quad . \\ \Psi_m \to (0,j_3) \end{array}$$

Before concluding this account of the FDIR space of SU(4), a word must be mentioned about the multiplicity of SU(3) IR's at a particular floor. It turns out that the site occupied by $T_-\ \Psi_m$ is spanned by three different SU(3) IR's, in general, obtained because $T_{41}T_{24}\Psi_m$, $T_{41}T_{34}\Psi_m$, and $T_{41}T_{24}\Psi_m$ are contained in three different SU(3) IR's, one being the SU(3) IR generated by Ψ_m itself. The two others have their maximal state at the same site as $T_-\ \Psi_m$ and $V_-\Psi_m$, respectively. The multiplicity of SU(3) IR's increases as one goes inward into each floor.

Finally, one must mention that the same SU(3) IR is obtained after a more conventional analysis of the FDIR space on the lines of Refs. 22 and 23. We present this analysis in order to relate to the FDIR of SU(3/1) and to establish our notation.

[1]P. Fayet and S. Ferrara, Phys. Rep. **32**, 249 (1977).

[2]P. D. Dondi and P. D. Jarvis, Phys. Lett. B **84**, 75 (1980); J. G. Taylor, *ibid.* **83**, 331 (1979); E. J. Squires, *ibid.* **82**, 395 (1979); D. B. Fairlie, *ibid.* **82**, 97 (1979); Y. Neeman, *ibid.* **81**, 190 (1979).

[3]F. Iachello, Phys. Rev. Lett. **44**, 772 (1980); A. Balantekin and I. Bars, J. Math. Phys. **23**, 1239 (1982).

[4]P. van Niewenhuizen, Phys. Rep. **68**, 192 (1981).

[5]V. Kac, Adv. Math. **26**, 8 (1977).

[6]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626,1640 (1976).

[7]M. Scheunert, *Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979).

[8]V. Kac, in *Lecture Notes in Mathematics*, Vol. 676, edited by K. Bleuler, H. R. Petry, and A. Reetz (Springer, Berlin, 1978).

[9]J. P. Hurni and B. Morel, J. Math. Phys. **24**, 197 (1983).

[10]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).

[11]J. van der Jeugt, J. Math. Phys. **25**, 3334 (1984).

[12]J. de Swart, Rev. Mod. Phys. **35**, 916 (1963).

[13]S. Gasiorowicz, *Elementary Particle Analysis* (Wiley, New York, 1966).

[14]By an independent lowering operator, we mean a lowering operator whose action may vanish without simultaneously requiring the action of other lowering operators to vanish as well.

[15]G. Hochschild, Ill. J. Math. **20**, 107 (1976); G. Hochschild and D. Djokovic, *ibid.* **20**, 124 (1976).

[16]J. Schwinger, in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965).

[17]T. Takabayashi, Prog. Theor. Phys. **32**, 981 (1964); N. Arisaka, *ibid.* **47**, 1758 (1972).

[18]J. Blank, M. Havliček, M. Bednar, and W. Lassner, "Canonical representations of the Lie superalgebra osp(1/4)," ICTP preprint IC|81|72.

[19]I. Bars and M. Gunyadin, Commun. Math. Phys. **91**, 31 (1983).

[20]X'_- does not change the V spin of the state on which it acts. $X_-\ \Psi_m$ has V spin $(v,v_3) = \{(j_1-1)/2, (j_1-1)/2\}$. The state $a_1T_-\Psi_m + a_2\Psi^{(1)}$ can belong to two possible V spin multiplets $v = (j_1+1)/2, (j_1-1)/2$. X'_- cannot change v by its action; thus the conclusion follows.

[21]R. T. Sharp and H. von Baeyer, J. Math. Phys. **7**, 1105 (1966).

[22]J. F. Cornwell, *Group Theory in Physics*, Vol. II, *Techniques of Physics*, Vol. 7 (Academic, New York, 1984).

[23]N. Jacobson, *Lie Algebras* (Dover, New York, 1979).

Conservation of angular momentum for systems of charged particles

Barton L. Willis
Department of Mathematics, University of North Carolina at Charlotte, Charlotte, North Carolina 28223

(Received 28 June 1988; accepted for publication 2 November 1988)

It is shown that the interaction part of the electromagnetic angular momentum density of a system of charged particles is integrable over any forward light cone. A covariant definition of the mechanical momentum is then given and it is shown that angular momentum is conserved.

I. INTRODUCTION

In Ref. 1, I showed that the interaction part of the energy-momentum tensor is integrable over any forward light cone. In this paper, I will show that the same is true for the angular momentum density. The interaction part of the angular momentum density $M_{(\mathrm{int})}$ are those terms of the angular momentum tensor that involve products of fields of distinct particles.[1] It is given by Jackson[2] that

$$M^{\alpha\beta\gamma}_{(\mathrm{int})} = T^{\alpha\beta}_{(\mathrm{int})} X^{\gamma} - T^{\alpha\gamma}_{(\mathrm{int})} X^{\beta}, \tag{1.1}$$

where $T_{(\mathrm{int})}$ is the interaction part of the energy-momentum tensor. One can show that the interaction part of the angular momentum tensor has zero divergence. That is, $M^{\mu\alpha\beta}_{(\mathrm{int}),\mu}(x) = 0$ for x not on a world line. The usual definition of the electromagnetic angular momentum[3,4]

$$L^{\alpha\beta}(\sigma) = \frac{1}{c}\int_{\sigma} M^{\mu\alpha\beta}\, d\sigma_{\mu}, \tag{1.2}$$

where σ is a spacelike hyperplane, has several problems. Most importantly, the angular momentum tensor M is $O(1)$ towards spatial infinity because the null fields vanish at spatial infinity like $O(r^{-1})$. Thus the integral defining the field angular momentum diverges for point particles. This has been discussed in a recent paper.[5] In Ref. 1, I showed that the interaction part of the energy-momentum tensor is integrable over any forward light cone. In Sec. II, I will show the same for the angular momentum tensor. In Sec. III, I will prove conservation of angular momentum for systems of charged particles interacting by retarded fields. Unlike other relativistic treatments of conservation laws for systems of particles,[6] I define the mechanical angular momentum as a function defined over all Minkowski space instead of as a function of the world lines of each particle.

I will use the following notation and conventions: The pseudometric g is defined by $g^{\mu\nu} = \mathrm{diag}\{1, -1, -1, -1\}$, velocities have unit length, and I will denote the forward and backward light cones with apex at x by $L^{+}(x)$ and $L^{-}(x)$, respectively.

II. INTEGRABILITY

I consider two charged particles whose world lines W and W^* are parametrized by $Z(\cdot)$ and $Z^*(\cdot)$, respectively. The electric charges associated with W and W^* will be denoted by e and e^*, respectively, and the electromagnetic fields by F and F^*, respectively.

In this section, I will show that $L^{\alpha\beta}$ defined by

$$L^{\alpha\beta}(x_0) = \frac{1}{c}\int_{L^{+}(x_0)} M^{\mu\alpha\beta}_{(\mathrm{int})}\, d\sigma_{\mu} \tag{2.1}$$

is finite. Without loss of generality, I will assume that $x_0 = 0$. The field F is given by

$$F^{\alpha\beta}(x) = e(\Gamma^{\alpha} M^{\beta} - \Gamma^{\beta} M^{\alpha}), \tag{2.2a}$$

where

$$\Gamma^{\alpha} = X^{\alpha} - Z^{\alpha}(\tau_0), \tag{2.2b}$$

$$\rho = \Gamma\cdot\dot{Z}(\tau_0), \tag{2.2c}$$

$$Q = \Gamma\cdot\ddot{Z}(\tau_0), \tag{2.2d}$$

$$M^{\alpha} = (1/\rho^2)\ddot{Z}^{\alpha} + [(1-Q)/\rho^3]\dot{Z}^{\alpha}, \tag{2.2e}$$

and τ_0 is the retarded proper time. The quantities Γ^*, ρ^*, Q^*, and M^* are defined similarly. I will assume that W and W^* intersect $L^{\pm}(x_0)$ so that unique retarded times exist for all $X \in L^{+}(x_0)$.[7]

The light cone $L^{+}(0)$ can be parametrized by

$$X^{\alpha} = \langle\sqrt{u^2+v^2+w^2}, u, v, w\rangle, \quad u, v, w \in R, \tag{2.3a}$$

$$= \langle r, u, v, w\rangle, \tag{2.3b}$$

where $r = \sqrt{u^2+v^2+w^2}$. The surface element $d\sigma_{\mu} \equiv J_{\mu}\, d\sigma$ is given by[1]

$$d\sigma_{\mu} = (1/r) X_{\mu}\, du\, dv\, dw = J_{\mu}\, d\sigma. \tag{2.4}$$

In Ref. 1, I found that

$$\begin{aligned} T^{\alpha\beta}_{(\mathrm{int})} J_{\beta} = {} & (ee^*/4\pi r)[(\Gamma^*\cdot M)(X\cdot M^*) - (M\cdot M^*)(\Gamma^*\cdot X)]\Gamma^{\alpha} \\ & + (ee^*/4\pi r)[(\Gamma\cdot M^*)(\Gamma^*\cdot X) - (\Gamma^*\cdot\Gamma)(M^*\cdot X)]M^{\alpha} \\ & + (ee^*/4\pi r)[(\Gamma\cdot M^*)(X\cdot M) - (M^*\cdot M)(X\cdot\Gamma)]\Gamma^{*\alpha} \\ & + (ee^*/4\pi r)[(\Gamma^*\cdot M)(\Gamma\cdot X) - (\Gamma^*\cdot\Gamma)(M\cdot X)]M^{*\alpha} \\ & + (ee^*/4\pi r)[(\Gamma\cdot\Gamma^*)(M\cdot M^*) - (\Gamma\cdot M^*)(\Gamma^*\cdot M)]X^{\alpha}. \end{aligned} \tag{2.5}$$

 0022-2488/89/030732-03$02.50

Thus

$$M^{\mu\alpha\beta}_{(\mathrm{int})}J_\mu = T^{\mu\alpha}_{(\mathrm{int})}X^\beta J_\mu - T^{\mu\beta}_{(\mathrm{int})}X^\alpha J_\mu \tag{2.6a}$$

$$\begin{aligned} &= (ee^*/4\pi r)[(\Gamma^*\cdot M)(X\cdot M^*) - (M\cdot M^*)(\Gamma^*\cdot X)][\Gamma^\alpha X^\beta - \Gamma^\beta X^\alpha] \\ &\quad + (ee^*/4\pi r)[(\Gamma\cdot M^*)(X\cdot M) - (M\cdot M^*)(\Gamma\cdot X)][\Gamma^{*\alpha}X^\beta - \Gamma^{*\beta}X^\alpha] \\ &\quad + (ee^*/4\pi r)[(\Gamma\cdot M^*)(\Gamma^*\cdot X) - (\Gamma^*\cdot\Gamma)(M^*\cdot X)][M^\alpha X^\beta - M^\beta X^\alpha] \\ &\quad + (ee^*/4\pi r)[(\Gamma^*\cdot M)(\Gamma\cdot X) - (\Gamma^*\cdot\Gamma)(M\cdot X)][M^{*\alpha}X^\beta - M^{*\beta}X^\alpha]. \end{aligned} \tag{2.6b}$$

In Ref. 1, I also showed that for $X \in L^+(x_0)$

$$\Gamma\cdot\Gamma^*,\quad \Gamma\cdot X,\quad \Gamma^*\cdot X = O(1),\quad r\to\infty, \tag{2.7a}$$

$$\Gamma\cdot M,\quad \Gamma^*\cdot M,\quad X\cdot M,\quad X\cdot M^* = O(r^{-2}),\quad r\to\infty. \tag{2.7b}$$

These order relations are not valid for X on some spacelike hyperplane. The assumption that $X \in L^\pm(x_0)$ is crucial. Furthermore,

$$\Gamma^\alpha X^\beta - \Gamma^\beta X^\alpha = Z^\beta X^\alpha - Z^\alpha X^\beta = O(r),\quad r\to\infty, \tag{2.8a}$$

$$\Gamma^{*\alpha}X^\beta - \Gamma^{*\beta}X^\alpha = Z^{*\beta}X^\alpha - Z^{*\alpha}X^\beta = O(r),\quad r\to\infty. \tag{2.8b}$$

One may verify that each term in Eq. (2.5) is $O(r^{-4})$ at infinity. Therefore, $M_{(\mathrm{int})}$ is integrable towards spatial infinity. The singularities of $M_{(\mathrm{int})}$ at the point of intersection of the world lines W and W^* and the light cone $L^+(0)$ are integrable because these singularities are no more singular than r^{-2}. Therefore, the interaction part of the angular momentum tensor for a system of charged particles is integrable over any forward light cone.

I now show that Eq. (2.1) is a reasonable definition of the field angular momentum by showing conservation of angular momentum.

III. CONSERVATION OF ANGULAR MOMENTUM

Let W, W^* intersect $L^\pm(x_0)$. In particular, assume that $Z(\tau^\pm), Z^*(\tau^{*\pm}) \in L^\pm(x_0)$. Define the mechanical angular momentum $L_{(\mathrm{mech})}(x_0)$ by

$$\begin{aligned} L^{\alpha\beta}_{(\mathrm{mech})}(x_0) &= mZ^\alpha(\tau^+)\dot Z^\beta(\tau^+) + m^*Z^{*\alpha}(\tau^{*+})\dot Z^{*\beta}(\tau_0^*) \\ &\quad - [\alpha\leftrightarrow\beta], \end{aligned} \tag{3.1}$$

where m and m^* are the masses associated with the world lines W and W^*, respectively. From now on, I will write Z for $Z(\tau^+)$, etc. The gradient of $L_{(\mathrm{mech})}$ is given by

$$\begin{aligned} L^{\alpha\beta,\gamma}_{(\mathrm{mech})} &= (m/\rho)Z^\alpha\ddot Z^\beta\Gamma^\gamma + (m^*/\rho^*)Z^{*\alpha}\ddot Z^{*\beta}\Gamma^{*\gamma} \\ &\quad - [\alpha\leftrightarrow\beta], \end{aligned} \tag{3.2}$$

where $\Gamma = x_0 - Z$, $\Gamma^* = x_0 - Z^*$, $\rho = \Gamma\cdot\dot Z$, and $\rho^* = \Gamma^*\cdot\dot Z^*$ [see Eq. (2.2b)]. I now compute the directional derivative of $L_{(\mathrm{int})}$ in some timelike direction l. Thus

$$L^{\alpha\beta,\mu}_{(\mathrm{int})}l_\mu = \frac{1}{c}\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\left[\int_{L^+(x_0+\epsilon l)} - \int_{L^+ + (x_0)} M^{\mu\alpha\beta}\,d\sigma_\mu\right].\right. \tag{3.3}$$

Let T_Δ and T^*_Δ be tubes of radius Δ that surround the world lines W and W^*, respectively, and let S_R be the portion of the spacelike hyperplane $X^0 = R$ that lies between the cones $L^+(x_0+\epsilon l)$ and $L^+(x_0)$. Choose the normal to S_R to have positive zeroth component. Then Eq. (3.3) becomes

$$\begin{aligned} &L^{\alpha\beta,\mu}_{(\mathrm{int})}(x_0)l_\mu \\ &= \lim_{\epsilon\to 0}\lim_{\Delta\to 0}\lim_{R\to\infty}\left[\left[\int_{L^+(x_0+\epsilon l)} - \int_{L^+(x_0)} - \int_{T_\Delta} - \int_{T^*_\Delta}\right.\right. \\ &\quad \left. + \int_{S_R} + \int_{T_\Delta} + \int_{T^*_\Delta} - \int_{S_R} M^{\mu\alpha\beta}\,d\sigma_\mu\right]. \end{aligned} \tag{3.4}$$

The first four integrals combine to an integral over a closed and bounded surface. Since $M^{\mu\alpha\beta}_{(\mathrm{int}),\mu} = 0$ on the interior of this surface, Gauss's theorem implies that these integrals sum to zero. I now compute

$$\lim_{\Delta\to 0}\int_{T_\Delta} M^{\mu\alpha\beta}\,d\sigma_\mu. \tag{3.5}$$

In Ref. 1, I gave a parametrization of the surface T_Δ, computed its surface element, and evaluated integrals of products of Γ's. From this work it follows that

$$\lim_{\Delta\to 0}\int_{T_\Delta} M^{\mu\alpha\beta}_{(\mathrm{int})}\,d\sigma_\mu = -\frac{e}{c}\int_{\tau_1}^{\tau_2} F^{*\alpha\mu}(Z(s))\dot Z_\mu(s)Z^\beta(s)ds - [\alpha\leftrightarrow\beta], \tag{3.6a}$$

$$\lim_{\Delta\to 0}\int_{T^*_\Delta} M^{\mu\alpha\beta}_{(\mathrm{int})}\,d\sigma_\mu = -\frac{e^*}{c}\int_{\tau_1^*}^{\tau_2^*} F^{\alpha\mu}(Z^*(s))\dot Z^*_\mu(s)Z^{*\beta}(s)ds - [\alpha\leftrightarrow\beta], \tag{3.6b}$$

where τ_1 and τ_2 are defined by $Z(\tau_1)\in L^+(x_0)$ and $Z(\tau_2)\in L^+(x_0+\epsilon l)$; τ_1^* and τ_2^* are defined similarly. These integrals can be identified as line integrals along each world line. However, $\tau_2 - \tau_1 = \Gamma\cdot l + O(\epsilon)$ and $\tau_2^* - \tau_1^* = \Gamma^*\cdot l + O(\epsilon)$. Thus

$$\begin{aligned} L^{\alpha\beta,\mu}_{(\mathrm{int})}(x_0)l_\mu &= -\frac{e}{c}(\Gamma\cdot l)F^{*\alpha\mu}\dot Z_\mu Z^\beta - \frac{e^*}{c}(\Gamma^*\cdot l)F^{\alpha\mu}\dot Z^*_\mu Z^{*\beta} \\ &\quad - [\alpha\leftrightarrow\beta] - \lim_{\epsilon\to 0}\lim_{R\to\infty}\int_{S_R} M^{\mu\alpha\beta}_{(\mathrm{int})}\,d\sigma_\mu, \end{aligned} \tag{3.7a}$$

$$= -L^{\alpha\beta,\mu}_{(\mathrm{mech})}l_\mu - \lim_{\epsilon\to 0}\lim_{R\to\infty}\int_{S_R} M^{\mu\alpha\beta}_{(\mathrm{int})}\,d\sigma_\mu. \tag{3.7b}$$

The equations of motion

$$mc\ddot{Z}^{\alpha} = (e/c)F^{*\alpha\beta}\dot{Z}_{\beta}, \tag{3.8a}$$

$$m^{*}c\ddot{Z}^{*\alpha} = (e^{*}/c)F^{\alpha\beta}\dot{Z}^{*}_{\beta}, \tag{3.8b}$$

were used to derive Eq. (3.7b) from Eq. (3.7a). Thus

$$[L^{\alpha\beta}_{(\mathrm{int})} + L^{\alpha\beta}_{(\mathrm{mech})}]^{,\mu} l_{\mu} = -\lim_{\epsilon\to 0}\lim_{R\to\infty}\int_{S_R} M^{\mu\alpha\beta}_{(\mathrm{int})}\, d\sigma_{\mu}. \tag{3.9}$$

The right-hand side of Eq. (3.9) is identified with the angular momentum that escapes to infinity. One should not expect the right-hand side of Eq. (3.9) to be zero because of the dissipative nature of this problem. Equation (3.9) is a relativistic generalization of the conservation of angular momentum for a multiparticle system.

IV. CONCLUSION

Conservation of angular momentum for systems of charged particles interacting by retarded electromagnetic fields has been proven in a manner similar to the conservation proof given in Ref. 1. The lack of a proof of integrability, plus the difficulty in defining the electromagnetic and the mechanical angular momentum in a consistent fashion, have been stumbling blocks for many treatments of this and similar problems.

[1]B. L. Willis, J. Math. Phys. **30**, 197 (1989).
[2]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed., p. 604, Eq. (12.109).
[3]F. Rohrlich, *Classical Charged Particles* (Addison–Wesley, Reading, MA, 1965), p. 94, Eq. (4–152).
[4]A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Dover, New York, 1980), p. 110, Eq. (3.77).
[5]A. Lozanda and P. Torres, J. Math. Phys. **29**, 1361 (1988).
[6]Ch. G. van Weert, Phys. Rev. D **9**, 339 (1974).
[7]See Ref. 1, Proposition 1.

Solution to the inverse problem of the 1-D wave equation

Jau-Inn Huang
Aldridge Laboratory of Applied Geophysics, Henry Krumb School of Mines, Columbia University, New York, New York 10027

(Received 15 March 1988; accepted for publication 9 November 1988)

A general solution for the determination of a variable coefficient in the 1-D wave equation from an additional boundary condition is presented. The Gopinath–Sondhi equation is shown to be a special case of this general solution. One application of this inversion method is to determine the structure of vertically inhomogeneous media from the backscattered waves generated by arbitrary incident waves. When *a priori* information about the unknown structure is available, a method is presented for the incorporation of this information to stabilize the inversion algorithm.

I. INTRODUCTION

The forward problem for the wave equation is the determination of the wave field as a function of space and time when the wave equation together with appropriate initial and boundary conditions are given. Various numerical algorithms are available to solve the forward problem. The inverse problem, on the other hand, is to determine unknown variable coefficients in the wave equation when a boundary condition, in addition to those that are required to solve the forward problem, is available. This additional boundary condition is usually provided by recording backscattered waves generated by an incident wave.

There are practical applications of the inverse problem for wave equations. One example is the determination of the elastic structure of the earth by processing backscattered waves recorded on the Earth's surface. For three-dimensionally inhomogeneous media, no exact solution of the inverse problem is known. However, for media that are inhomogeneous only in the vertical direction, several inversion schemes are available.

In the literature there are many exact methods devised to solve slightly different inverse problems. The pioneering work of Gel'fand and Levitan[1] was originally developed to determine the Schrödinger potential from the spectral function. The equation of Marchenko[2] recovers the potential function from a frequency-dependent reflection coefficient. Gopinath and Sondhi[3] treated the inverse problem on a transmission line. Burridge[4] summarized these methods in the time domain and discussed the applications to geophysical inverse problems. The time domain formulation of the Marchenko equation was also considered by Kay,[5] Ware and Aki,[6] and Balanis.[7] Carroll and Santosa[8] generalized the Gel'fand–Levitan method and relaxed the smoothness requirement of the unknown impedance function. A survey of works by many other contributors can be found in Ref. 9.

As pointed out by Balanis,[10] the time domain formulation, when compared to the frequency domain formulation, yields more natural and intuitive derivation of key inversion results. This paper, inspired by Balanis' work, considers the 1-D wave equation in the time domain with arbitrary boundary conditions that do not necessarily involve the Dirac δ function.

The second-order wave equation is first decomposed into a simultaneous system of two first-order equations. The representation of the wave field involves the boundary conditions and two auxiliary functions that are uniquely determined (and vice versa) by the unknown variable coefficient in the original wave equation. When the causality principle is invoked, two integral equations are obtained to relate boundary conditions to the unknown coefficient. These two integral equations are not independent of each other, and either one can be the basis of the inversion scheme. Furthermore, the Gopinath–Sondhi equation is shown to be a special case of one integral equation derived here.

In practical applications the inversion schemes are unstable for bandlimited incident waves. Tikhonov regularization procedures[11] can mathematically restore the stability, but they usually also introduce physically unacceptable artifacts. When *a priori* information about the unknown medium is available, a method, which will be presented here, to incorporate this information can be used to stabilize the inversion algorithm. This method has proven successful on real seismic data and has yielded useful geological information.

Finally, in the concluding remarks (Sec. VI) we briefly discuss the possibility of extending the inversion method presented here to dissipative wave equations.

II. FORMULATION OF THE INVERSE PROBLEM

We shall consider the equation

$$U_{tt} - U_{zz} - q(z)\,U_z = 0,$$

for $-\infty < t < \infty$ and $z \geqslant 0$. The subscripts denote differentiations. The term $q(z)$, a continuous function, is related to the acoustic impedance $a(z)$ by $q(z) = a_z/a$. Hereafter, $a(z)$ is normalized such that $a(0) = 1$.

After being decomposed into a system of two first-order equations, the above equation becomes

$$\mathbf{V}_t + \mathsf{A}(z)\mathbf{V}_z = 0, \tag{1a}$$

where

$$\mathbf{V}(z,t) \equiv \begin{pmatrix} -a(z)\,U_z \\ U_t \end{pmatrix},$$

and

$$\mathsf{A}(z) \equiv \begin{pmatrix} 0 & a(z) \\ 1/a(z) & 0 \end{pmatrix}. \tag{1b}$$

 0022-2488/89/030735-05$02.50

The inverse problem is then formulated as follows: Given the initial condition $\mathbf{V}(z,t) = 0$, for $t<0$; and boundary conditions

$$\mathbf{V}(0,t) = \begin{pmatrix} g(t) \\ f(t) \end{pmatrix}, \qquad (1c)$$

for $t \geqslant 0$; find the unknown function $a(z)$.

For the forward problem $a(z)$ is known, and the quiescent initial condition together with either $f(t)$ or $g(t)$ will be sufficient to uniquely determine the wave field $\mathbf{V}(z,t)$. For the inverse problem $a(z)$ is unknown, and it has to be determined by both $f(t)$ and $g(t)$. For artificially created $f(t)$ and $g(t)$, there may be no solution for $a(z)$. If f and g are provided by physical measurements, the solution for $a(z)$ exists but may not be stable.

III. DERIVATION OF THE INTEGRAL EQUATIONS

To facilitate the derivation of the solution of the inverse problem, we define two auxiliary functions $\mathbf{V}_1(z,t)$ and $\mathbf{V}_2(z,t)$ as solutions of Eq. (1a) with conditions[12]

$$\mathbf{V}_1(0,t) = \begin{pmatrix} \delta(t) \\ \delta(t) \end{pmatrix}, \quad \mathbf{V}_2(0,t) = \begin{pmatrix} \delta(t) \\ -\delta(t) \end{pmatrix},$$

for $t \in (-\infty, \infty)$.

The terms $\mathbf{V}_1$ and $\mathbf{V}_2$ are nonvanishing in the region $z \geqslant |t|$, and are uniquely determined by $a(z)$ when there is only one spatial dimension. For higher spatial dimensions, the one-to-one correspondence between $a(z)$ and the auxiliary functions breaks down.

Here, $\mathbf{V}$ can be expressed in terms of $\mathbf{V}_1$ and $\mathbf{V}_2$ as

$$\mathbf{V}(z,t) = \int_{-\infty}^{\infty} I(t-y)\mathbf{V}_1(z,y)dy - \int_{-\infty}^{\infty} R(t-y)\mathbf{V}_2(z,y)dy, \qquad (2)$$

where

$$\begin{pmatrix} I(t) \\ R(t) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} f(t)+g(t) \\ f(t)-g(t) \end{pmatrix}$$

for $t \geqslant 0$, and both $I(t)$ and $R(t)$ vanish for $t<0$. Due to physical considerations, we can call $I(t)$ the incident wave and $R(t)$ the reflected or backscattered wave.

A heuristic argument to justify the above representation is to transform Eq. (1a) to the frequency domain. Equation (1a) becomes a system of two first-order linear ordinary differential equations, with the transformed $\mathbf{V}_1$ and $\mathbf{V}_2$ as two independent solutions. Therefore $\mathbf{V}$ can be expressed as a linear combination of $\mathbf{V}_1$ and $\mathbf{V}_2$ with coefficients given in Eq. (2).

Superficially Eq. (2) is of little use because there are too many unknowns. However, when we eventually restrict our attention to the area outside the light cone, i.e., $z > |t|$, the causality principle will force the physical field on the left-hand side of the equation to be zero. The resultant equation then relates boundary conditions (I and R) to auxiliary functions ($\mathbf{V}_1$ and $\mathbf{V}_2$), which in turn are related to $a(z)$.

A few useful properties of $\mathbf{V}_1$ and $\mathbf{V}_2$ are listed in the following. The proofs are given in the Appendix.

Property 1: Let

$$\mathbf{V}_1(z,t) \equiv \begin{pmatrix} p(z,t) \\ w(z,t) \end{pmatrix},$$

then

$$\mathbf{V}_2(z,t) = \begin{pmatrix} p(z,-t) \\ -w(z,-t) \end{pmatrix}.$$

Property 2: For continuous $a(z)$, $\mathbf{V}_1(z,t)$ has the form

$$\mathbf{V}_1(z,t) = \delta(z-t)\mathbf{D}(z) + H(z-t)\mathbf{B}(z,t), \qquad (3)$$

where H is the Heaviside step function, and $\mathbf{D}$ and $\mathbf{B}$ are continuous vector functions, with the support of $\mathbf{B}(z,t)$ in the region $z \geqslant |t|$. Here $a_z(0) = 0$ is assumed.

Property 3: $\mathbf{D}(z)$ is directly related to $a(z)$ by

$$\mathbf{D}(z) = \begin{pmatrix} \sqrt{a(z)} \\ 1/\sqrt{a(z)} \end{pmatrix}. \qquad (4)$$

Property 4: Let

$$\mathbf{B}(z,t) \equiv \begin{pmatrix} b_1(z,t) \\ b_2(z,t) \end{pmatrix},$$

then $\mathbf{B}$ is the solution of the following Goursat problem in the region $z \geqslant |t|$:

$$\mathbf{B}_t(z,t) + \mathsf{A}(z)\mathbf{B}_z(z,t) = 0, \qquad (5a)$$

$$(2/\sqrt{a(z)})b_1(z,z) - 2\sqrt{a(z)}b_2(z,z) = -a_z(z)/a(z), \qquad (5b)$$

$$\mathbf{B}(z,-z) = 0. \qquad (5c)$$

Property 5: Defining $\mathbf{S}(z) \equiv \int_{-z}^{z} \mathbf{B}(z,t)dt$, (6a)

then

$$\mathbf{S}(z) \equiv \begin{pmatrix} s_1(z) \\ s_2(z) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \mathbf{D}(z) = \begin{pmatrix} 1-\sqrt{a(z)} \\ 1-1/\sqrt{a(z)} \end{pmatrix}. \qquad (6b)$$

With these properties, we can derive the integral equations to recover $a(z)$. Inserting Eq. (3) into Eq. (2) and using Property 1, it follows that in the region $z \geqslant |t|$:

$$0 = -R(t+z)\begin{pmatrix} \sqrt{a(z)} \\ -1/\sqrt{a(z)} \end{pmatrix} + \int_{-z}^{t} I(t-y)\begin{pmatrix} b_1(z,y) \\ b_2(z,y) \end{pmatrix} dy - \int_{-t}^{z} R(t+y)\begin{pmatrix} b_1(z,y) \\ -b_2(z,y) \end{pmatrix} dy. \qquad (7)$$

Or, written in two scalar equations,

$$0 = R(t+z) - \int_{-z}^{t} I(t-y)\left[\frac{b_1(z,y)}{\sqrt{a(z)}}\right] dy + \int_{-t}^{z} R(t+y)\left[\frac{b_1(z,y)}{\sqrt{a(z)}}\right] dy, \qquad (8a)$$

$$0 = R(t+z) + \int_{-z}^{t} I(t-y)\left[\sqrt{a(z)}b_2(z,y)\right] dy + \int_{-t}^{z} R(t+y)\left[\sqrt{a(z)}b_2(z,y)\right] dy, \qquad (8b)$$

for $z \geqslant |t|$.

Since I and R are given, one can solve Eq. (8a) for $b_1(z,y)/\sqrt{a(z)}$ and obtain $s_1(z)/\sqrt{a(z)}$ as defined in Eq. (6a). It follows from Eq. (6b) that $a(z)$ is recovered by

$$a(z) = \left[1 + s_1(z)/\sqrt{a(z)}\right]^{-2}. \qquad (9a)$$

Similarly, solving Eq. (8b) one obtains $\sqrt{a(z)}b_2(z,y)$ and hence $\sqrt{a(z)}s_2(z)$. The term $a(z)$ is then recovered by

$$a(z) = [1 + \sqrt{a(z)}s_2(z)]^2. \tag{9b}$$

Either Eq. (8a) or Eq. (8b) can be used as the basis of an inversion scheme. It should be pointed out that these two integral equations are not independent of each other. A little algebra will confirm that Eq. (8a) multiplied by $\sqrt{a(z)}$, then differentiated with respect to z, is identical to Eq. (8b), differentiated with respect to t.

IV. GOPINATH–SONDHI EQUATION AS A SPECIAL CASE

The inverse problem posed and the solution provided by Gopinath and Sondhi[3] can be summarized as follows:

Problem: $\mathbf{V}_t + \mathsf{A}\mathbf{V}_z = 0$, for $z \geqslant 0$. Initial condition $\mathbf{V}(z,t) = 0$, for $t < 0$. Boundary conditions

$$\mathbf{V}(0,t) = \begin{pmatrix} \delta(t) \\ h(t) \end{pmatrix},$$

for $t \geqslant 0$. Here, A is defined as in Eq. (1b). Here $h(t)$ is the backscattered wave generated by a Dirac-δ incident wave.

Solution: $a(z) = G^2(z,z)$. Here $G(z,t)$ is obtained by solving

$$G(z,t) + \frac{1}{2}\int_{-z}^{z} \bar{h}(|t-y|)G(z,y)dy = 1,$$

in the region $z \geqslant |t|$. Here $\bar{h} \equiv h - \delta$.

We shall show that Eq. (8a) can be reduced to the Gopinath–Sondhi equation. The boundary conditions given by Gopinath and Sondhi imply that $R = \frac{1}{2}(h - \delta) = \frac{1}{2}\bar{h}$ and $I = \frac{1}{2}(h + \delta) = R + \delta$. Denoting $b_1(z,y)/\sqrt{a(z)}$ by $b(z,y)$, Eq. (8a) becomes

$$0 = R(t+z) - \int_{-z}^{t} [R(t-y) + \delta(t-y)]b(z,y)\,dy + \int_{-t}^{z} R(t+y)b(z,y)dy,$$

or, after rearranging terms and replacing t by t',

$$0 = R(t'+z) - b(z,t') - \int_{-z}^{t'} R(t'-y)[b(z,y) - b(z,-y)]\,dy.$$

Integrating over t' from $-z$ to t (t is restricted by $|t| < z$), the above equation leads to

$$0 = \bar{R}(t+z) - F(z,t) - \int_{-z}^{t} dt' \int_{-z}^{t'} dy\, R(t'-y)[b(z,y) - b(z,-y)],$$

where $\bar{R}$ and F are defined as

$$\bar{R}(t) \equiv \int_0^t R(s)ds,$$

$$F(z,t) \equiv \int_{-z}^{t} b(z,y)\,dy.$$

Changing the order of integration, it follows that

$$0 = \bar{R}(t+z) - F(z,t) - \int_{-z}^{t} dy\, \bar{R}(t-y)[b(z,y) - b(z,-y)].$$

Using $b(z,y) = F_y(z,y)$ and integrating by parts, one obtains

$$0 = \bar{R}(t+z)[1 + F(z,z)] - F(z,t) - \int_{-z}^{t} R(t-y)[F(z,y) + F(z,-y)]\,dy.$$

Replacing t by $-t$ in the above equation yields

$$0 = \bar{R}(-t+z)[1 + F(z,z)] - F(z,-t) - \int_{t}^{z} R(-t+y)[F(z,y) + F(z,-y)]\,dy.$$

The sum of the last two equations yields

$$0 = -[\bar{R}(t+z) + \bar{R}(-t+z)][1 + F(z,z)] + [F(z,t) + F(z,-t)] + \int_{-z}^{z} R(|t-y|)[F(z,y) + F(z,-y)]dy.$$

Recognizing

$$\int_{-z}^{z} R(|t-y|)dy = \bar{R}(t+z) + \bar{R}(z-t),$$

and defining

$$G(z,y) \equiv 1 - [F(z,y) + F(z,-y)]/[1 + F(z,z)],$$

it follows that

$$1 = G(z,t) + \int_{-z}^{z} R(|t-y|)G(z,y)\,dy = G(z,t) + \frac{1}{2}\int_{-z}^{z} \bar{h}(|t-y|)G(z,y)\,dy,$$

which is the Gopinath–Sondhi equation.

Finally, since $F(z,z) = s_1(z)/\sqrt{a(z)}$ by definition and $s_1(z) = 1 - \sqrt{a(z)}$, from the definition of G it follows that $G^2(z,z) = a(z)$, which completes the reduction.

V. INCORPORATING *A PRIORI* INFORMATION

In shorthand notation both Eqs. (8a) and (8b) can be expressed as

$$0 = R + \mathsf{K}b, \tag{10}$$

where K is the given kernel of the integral equation, and $b = b_1/\sqrt{a}$ for Eq. (8a) and $\sqrt{a b_2}$ for Eq. (8b). In practical applications the solutions of the integral equations are usually unstable. By imposing bounds on b or its derivatives, the Tikhonov's regularization procedure reformulates the integral equation as an optimization problem: find the function b that minimizes the functional

$$\phi[b] = \|\mathsf{K}b + R\|^2 + (\epsilon/M)^2\|\mathsf{B}b\|^2,$$

where B is a constraint operator that has a bounded inverse, ϵ is estimated from the signal-to-noise ratio in the recorded data, and M is the bound imposed on $\mathsf{B}b$, i.e., $\|\mathsf{B}b\| < M$. Although B can be the identity operator or derivatives of any order, experience with real seismic data indicate that the identity operator is usually sufficient to restore the stability, and derivative operators seldom add any improvement to the quality of the inversion results.

The arbitrariness in selecting the operator B and the upper bound M implies that some mathematical artifacts that bear no physical relevance can be introduced into the final inversion results. When *a priori* information about the medium is available, we shall present a more satisfactory way to incorporate the *a priori* information that also stabilizes the inversion algorithm.

Based on the *a priori* information, a resonable impedance function $a_0(z)$ can be constructed. The backscattered waves generated from this test impedance function may not be compatible with the physical measurements. The following inversion algorithm, which is stable, can improve the test impedance function to yield a final model that is consistent with both physical measurements and the *a priori* information.

By replacing the unknown $a(z)$ with the given $a_0(z)$ in Eq. (5), the Goursat problem yields a unique solution

$$\mathbf{B}_0(z,t) = \begin{pmatrix} b^0{}_1(z,t) \\ b^0{}_2(z,t) \end{pmatrix}.$$

In conforming to our shorthand notation in Eq. (10), the original inverse problem is reformulated as finding the function $b(z,t)$ that simultaneously satisfies the following inequalities:

$$\|R + \mathsf{K}b\| < \epsilon, \tag{11a}$$

$$\|b - b_0\| < M, \tag{11b}$$

where $b_0 = b^0{}_1/\sqrt{a_0}$ or $\sqrt{a_0 b^0{}_2}$ depending on whether Eq. (8a) or Eq. (8b) is used in the inversion scheme. Here ϵ is again estimated from the signal-to-noise ratio of the real data, and M reflects our confidence level in the *a priori* information. The more confidence that we have, the smaller is the value of M that can be used.

Equivalent to Eq. (11), an integral equation of the second kind

$$\mathsf{K}^+\mathsf{K}b + (\epsilon/M)^2(b - b_0) = 0, \tag{12}$$

is solved in practical applications. Here K^+ denotes the operator adjoint to K. The resultant solution b via Eqs. (6) and (9) gives the final impedance function $a(z)$. The usefulness of this inversion method has been tested on a real seismic data set.[13] A travel time inversion method that does not account for the amplitudes and waveforms in the data provides the *a priori* model. The application of this inversion method, constrained by the *a priori* model, confirms the existence of a controversial low-velocity zone at the base of the oceanic crust in the North Atlantic.

VI. CONCLUSIONS AND FINAL REMARKS

A general solution to the inverse problem of the one-dimensional wave equation is derived, and a method for incorporating *a priori* information to stabilize the inversion algorithm is presented. While the derivation presented here is rather simple and straightforward, it nevertheless provides a general framework that includes the Gopinath–Sondhi equation as a special case. A reexamination of the derivation shows that the success of the 1-D inversion hinges on two things: the availability of the wave field representation [Eq. (2)] and the causality principle. Equation (2) is available for 1-D space-time because there the roles of space and time are interchangeable. For higher-dimensional inverse problems, the causality principle still holds, but the benefit of the representation similar to Eq. (2) is lost.

A possible generalization of this inversion method is to recover the velocity and the electric conductivity profiles in a stratified medium that supports electromagnetic waves. The nonzero conductivity introduces one additional term U_t to the wave equation. Auxiliary functions are still available to give a representation of the wave field, and it seems possible to solve this inverse problem for dissipative wave equations following a procedure similar to that outlined in this paper. The properties of the auxiliary functions and the form of the integral equations may be altered due to the presence of the dissipative term. Research on this inverse problem is in progress.

ACKNOWLEDGMENTS

Acknowledgments are due to Yu-Hua Chu for invaluable discussions. L. Alsop read the manuscript and made helpful comments.

This research was supported by Project MIDAS sponsored by a consortium of oil and geophysical companies.

APPENDIX: SKETCHY PROOFS OF PROPERTIES

Property 1: If $\mathbf{V}_1 = \binom{p(z,t)}{w(z,t)}$, it can be easily verified that $\binom{p(z,-t)}{-w(z,-t)}$ and $\mathbf{V}_2$(z,t) satisfy the same equation and conditions; hence they are identical.

Property 2: For continuous $a(z)$, the most singular part of $\mathbf{V}_1(z,t)$ along $z = t$ will contain a delta-function term followed by the next singular part that contains the Heaviside function.[14]

Property 3: Substituting Eq. (3) into Eq. (1a), it follows that

$$-\delta'(z-t)[\mathbf{D} - \mathsf{A}\mathbf{D}] - \delta(z-t)[\mathbf{B} - \mathsf{A}\mathbf{D}_z - \mathsf{A}\mathbf{B}]$$
$$+ H(z-t)[\mathbf{B}_t + \mathsf{A}\mathbf{B}_z] = 0.$$

Setting the coefficients of δ', δ, and H to zeros, one obtains

$$\mathsf{A}(z)\mathbf{D}(z) = \mathbf{D}(z), \tag{A1}$$

$$\mathbf{B}(z,z) - \mathsf{A}(z)\mathbf{D}_z(z) - \mathsf{A}(z)\mathbf{B}(z,z) = 0, \tag{A2}$$

and

$$\mathbf{B}_t(z,t) + \mathsf{A}(z)\mathbf{B}_z(z,t) = 0, \quad \text{in } z \geqslant |t|. \tag{A3}$$

To obtain Eq. (4), $\mathbf{D}$ is first solved as the eigenvector of A with eigenvalue 1, which gives

$$\mathbf{D}(z) = \begin{pmatrix} \sqrt{a(z)} \\ 1/\sqrt{a(z)} \end{pmatrix} d(z), \tag{A4}$$

where $d(z)$ is a scalar function to be determined.

Multiplying Eq. (A2) from the left by $[1/\sqrt{a(z)} \;\; \sqrt{a(z)}]$, it follows that

$$[1/\sqrt{a(z)} \;\; \sqrt{a(z)}]\mathbf{D}_z(z) = 0.$$

When combined with Eq. (A4), the above equation yields $d_z(z) = 0$. The term $d(z)$ is therefore a constant, and this constant is determined to be 1 from Eq. (3) and the boundary condition of $\mathbf{V}_1$ at $z = 0$.

Property 4: Equation (5a) has already been proven as Eq. (A3). Inserting Eq. (4) into Eq. (A2) results in Eq. (5b). Because $\mathbf{V}_1(0,t) = 0$ for $t<0$ and the Cauchy data propagating along $z = -t$ originating from the origin is also zero, $\mathbf{V}_1(z,-z) = 0$. Equation (5c) therefore follows by setting $t = -z$ in Eq. (3).

Property 5: Integrating Eq. (5a) over t from $-z$ to z, and making use of

$$\mathbf{S}_z(z) = \mathbf{B}(z,z) + \int_{-z}^{z} \mathbf{B}_z(z,t)\,dt,$$

which is obtained by differentiating Eq. (6a), one obtains

$$\mathbf{B}(z,z) + \mathsf{A}(z)[\mathbf{S}_z(z) - \mathbf{B}(z,z)] = 0.$$

Combining the above equation and Eq. (A2), it follows that $\mathbf{S}(z) + \mathbf{D}(z) = \text{const}$. This constant is determined by the initial values $\mathbf{S}(0) = \binom{0}{0}$ and $\mathbf{D}(0) = \binom{1}{1}$. Hence $\mathbf{S}(z) = \binom{1}{1} - \mathbf{D}(z)$, which is Eq. (6b).

[1]I. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Transl. Ser. 2 **1**, 253 (1955).

[2]V. A. Marchenko, Dokl. Akad. Nauk. SSSR **72**, 457 (1950).

[3]B. Gopinath and M. M. Sondhi, Proc. IEEE **59**, 383 (1971).

[4]R. Burridge, Wave Motion **2**, 305 (1980).

[5]I. Kay, Commun. Pure Appl. Math. **13**, 371 (1960).

[6]J. A. Ware and K. Aki, J. Acoust. Soc. Am. **45**, 911 (1969).

[7]G. N. Balanis, J. Math. Phys. **13**, 1001 (1972).

[8]R. Carroll and F. Santosa, Math. Methods Appl. Sci. **3**, 145 (1981).

[9]*Inverse Problems of Acoustic and Elastic Waves*, edited by F. Santosa, Y.-H. Pao, W. W. Symes, and C. Holland (SIAM, Philadelphia, 1984).

[10]G. N. Balanis, J. Math. Phys. **23**, 2562 (1982).

[11]A. Tikhonov, Am. Math. Soc. Transl., Ser. 2 **70**, 222 (1968).

[12]These auxiliary functions are similar to those introduced by Balanis in Ref. 7.

[13]J.-I. Huang, Y.-H. Chu, and J. T. Kuo, Trans. Am. Geophys. Union **67**, 44 (1986).

[14]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. 2, p. 622.

Algebraic construction of one-dimensional quasiperiodic tilings

Daniel B. Litvin
Department of Physics, Pennsylvania State University, Berks Campus, P. O. Box 7009, Reading, Pennsylvania 19610-6009

(Received 28 June 1988; accepted for publication 19 October 1988)

For an arbitrary one-dimensional quasiperiodic tiling constructed via the *grid method* with periodically spaced grids, an algebraic equation is derived for the positions of the vertices dependent on a single variable, the cardinal position of the vertices.

I. INTRODUCTION

Interest in quasiperiodic tiling predates the 1984 discovery by Shechtman *et al.*[1] of an alloy of aluminum and manganese that exhibits an electron diffraction pattern with icosahedral symmetry. In 1981 de Bruijn[2] developed a *grid method* to construct the prototypic two-dimensional quasiperiodic Penrose tilings,[3–5] and showed that these tilings could also be constructed by a *projection method* from a higher-dimensional space. Mackay[6,7] constructed a three-dimensional icosahedral quasiperiodic tiling using a pair of *unit cells*. The projection method was generalized to the construction of three-dimensional icosahedral quasiperiodic tilings by Kramer and Neri.[8] With the discovery by Shechtman *et al.*[1] much work has been published on the construction of three-dimensional quasiperiodic tilings using grid methods[9–11] and projection methods.[12–20] Gahler and Rhyner[11] have shown the equivalence of the two methods. We shall consider in this paper one-dimensional quasiperiodic tilings constructed via the grid method with periodically spaced grids. Those tilings constructed using quasiperiodically spaced grids or tilings generated by inflation rules not obtainable by the projection method[21] will not be considered.

In 1986, using special sequences of ones and zeros introduced by de Bruijn,[22] Litvin and Litvin[23] derived an algebraic equation for the positions of the vertices of a one-dimensional quasiperiodic tiling consisting of two basic tilings with the equation dependent on a single index, the cardinal position of the vertices. In this paper, we shall derive an analogous algebraic equation for the positions of the vertices of a one-dimensional quasiperiodic tiling consisting of an arbitrary number of basic tilings. In Sec. II we define an arbitrary one-dimensional quasiperiodic tiling using the grid method and state the algebraic equation for the vertices of this tiling. In Sec. III we give an inductive proof of this equation. In Sec. IV we compare this work with the results of Ref. 23.

II. ARBITRARY ONE-DIMENSIONAL QUASIPERIODIC TILING

We construct an arbitrary one-dimensional quasiperiodic tiling with p basic tilings using the grid method as follows: On a number line we plot the p set of points

$$\{nt_i + \gamma_i | n \in Z\}, \quad i = 1,2,\ldots,p, \tag{1}$$

where we shall assume that the ratios t_i/t_j, $i \neq j$, $i,j = 1,2,\ldots,p$, are irrational and the γ_i are constants such that $\gamma_i < t_i$ for $i = 1,2,\ldots,p$. This construct divides the number line into segments that can be characterized by a p-tuple of integers $(M_1,M_2,\ldots,M_p)$, where the $(M_1,M_2,\ldots,M_p)^{\text{th}}$ segment is defined by

$$\{x | M_1 t_1 \leqslant x < (M_1+1)t_1 \cap M_2 t_2 \leqslant x < (M_2+1)t_2 \cap \cdots \cap M_p t_p \leqslant x < (M_p+1)t_p\}. \tag{2}$$

For each p-tuple $(M_1,M_2,\ldots,M_p)$ determined above we construct on a second number line for a given p-tuple of real numbers $(a_1,a_2,\ldots,a_p)$ the point

$$X(M_1,M_2,\ldots,M_p) = M_1 a_1 + M_2 a_2 + \cdots + M_p a_p. \tag{3}$$

The problem that we consider is the derivation of an algebraic equation dependent on a single index, the cardinal position of the vertices, for the set of points defined by Eq. (3) for all p-tuples of integers $(M_1,M_2,\ldots,M_p)$ defined by Eqs. (1) and (2). We shall prove the following theorem.

Theorem: Let $N = M_1 + M_2 + \cdots + M_p$ be the cardinal position of the point $X(M_1,M_2,\ldots,M_p)$ defined by Eq. (3). Then the position $X(N)$ is given by

$$X(N) = \sum_{j=1}^{p} \left\lfloor \left(N + 1 + \sum_{m=1}^{p} \frac{\gamma_m - \gamma_j}{t_m}\right)\left(\sum_{k=1}^{p} \frac{t_j}{t_k}\right)^{-1} \right\rfloor a_j, \tag{4}$$

where $\lfloor \ \rfloor$ denotes the integer function, $\lfloor y \rfloor$ is the greatest integer less than or equal to y.

III. PROOF OF EQ. (4)

We shall give an inductive proof of Eq. (4). We shall prove that the theorem is correct for the cases of $p = 2$ and $p = 3$, and then for an arbitrary integer g, prove that it is correct for the case $p = g + 1$ assuming the equation is valid for the case $p = g$.

A. $p=2$

The vertices of the tilings in the case $p = 2$ are given by

$$X(M_1,M_2) = M_1 a_1 + M_2 a_2. \tag{3'}$$

We subdivide this set of points into two subsets; the first, the vertices at the end of tilings of length a_1 and the second, the vertices at the end of tilings of length a_2. We denote the points of these two subsets, respectively, as $X(N_1)$ and $X(N_2)$, where N_1 and N_2 denote the cardinal positions of the vertices in the tiling. We have that

$$X(N_1) = M_1 a_1 + \lfloor (M_1 t_1 + \gamma_1 - \gamma_2)/t_2 \rfloor a_2, \tag{5a}$$

$$X(N_2) = \lfloor (M_2 t_2 + \gamma_2 - \gamma_1)/t_1 \rfloor a_1 + M_2 a_2, \tag{5b}$$

where

 0022-2488/89/030740-04$02.50

$$N_1 = M_1 + \lfloor (M_1 t_1 + \gamma_1 - \gamma_2)/t_2 \rfloor, \quad (6a)$$

$$N_2 = M_2 + \lfloor (M_2 t_2 + \gamma_2 - \gamma_1)/t_1 \rfloor. \quad (6b)$$

We can invert Eqs. (6): Since $\lfloor y \rfloor = y - \Delta$, where $0 \leqslant \Delta < 1$, we can rewrite Eq. (6a) as

$$N_1 = M_1 + (M_1 t_1 + \gamma_1 - \gamma_2)/t_2 - \Delta$$

and consequently,

$$N_1 + 1 = M_1[(t_1 + t_2)/t_2] + (\gamma_1 - \gamma_2)/t_2 + (1 - \Delta),$$

$$(N_1 + 1)\frac{t_2}{t_1 + t_2} + \left(\frac{\gamma_2 - \gamma_1}{t_1 + t_2}\right) = M_1 + (1 - \Delta)\frac{t_2}{t_1 + t_2}.$$

Since M_1 is an integer and $0 \leqslant (1 - \Delta)t_2/(t_1 + t_2) < 1$, we have on taking the integer function of both sides of the previous equation that

$$M_1 = \lfloor (N_1 + 1)[t_2/(t_1 + t_2)] + (\gamma_2 - \gamma_1)/(t_1 + t_2) \rfloor,$$

$$M_1 = \left\lfloor \left(N_1 + 1 + \frac{\gamma_2 - \gamma_1}{t_2}\right)\left(\sum_{k=1}^{2} \frac{t_1}{t_k}\right)^{-1} \right\rfloor. \quad (7a)$$

In an analogous manner we derive from Eq. (6b) that

$$M_2 = \left\lfloor \left(N_2 + 1 + \frac{\gamma_1 - \gamma_2}{t_1}\right)\left(\sum_{k=1}^{2} \frac{t_2}{t_k}\right)^{-1} \right\rfloor. \quad (7b)$$

The coefficient of a_2 in Eq. (5a) is, using Eq. (6a), equal to $N_1 - M_1$, and using Eq. (7a) and the relationship that $-\lfloor y \rfloor = \lfloor -y \rfloor + 1$, we have

$$N_1 - M_1 = N_1 + 1 + \left\lfloor -\left(N_1 + 1 + \frac{\gamma_2 - \gamma_1}{t_2}\right) \times \left(\sum_{k=1}^{2} \frac{t_1}{t_k}\right)^{-1} \right\rfloor,$$

$$N_1 - M_1 = \left\lfloor \left(N_1 + 1 + \frac{\gamma_1 - \gamma_2}{t_1}\right)\left(\sum_{k=1}^{2} \frac{t_2}{t_k}\right)^{-1} \right\rfloor. \quad (8a)$$

The coefficient of a_1 in Eq. (5b) is, using Eq. (6b), equal to $N_2 - M_2$, and using Eq. (7b) we have

$$N_2 - M_2 = \left\lfloor \left(N_2 + 1 + \frac{\gamma_2 - \gamma_1}{t_2}\right)\left(\sum_{k=1}^{2} \frac{t_1}{t_k}\right)^{-1} \right\rfloor. \quad (8b)$$

We can now rewrite Eqs. (5) as

$$X(N_1) = \left\lfloor \left(N_1 + 1 + \frac{\gamma_2 - \gamma_1}{t_2}\right)\left(\sum_{k=1}^{2} \frac{t_1}{t_k}\right)^{-1} \right\rfloor a_1 + \left\lfloor \left(N_1 + 1 + \frac{\gamma_1 - \gamma_2}{t_1}\right)\left(\sum_{k=1}^{2} \frac{t_2}{t_k}\right)^{-1} \right\rfloor a_2,$$

$$X(N_2) = \left\lfloor \left(N_2 + 1 + \frac{\gamma_2 - \gamma_1}{t_2}\right)\left(\sum_{k=1}^{2} \frac{t_1}{t_k}\right)^{-1} \right\rfloor a_1 + \left\lfloor \left(N_2 + 1 + \frac{\gamma_1 - \gamma_2}{t_1}\right)\left(\sum_{k=1}^{2} \frac{t_2}{t_k}\right)^{-1} \right\rfloor a_2$$

and since the cardinal positions of the two subsets of vertices are mutually exclusive we can write a single equation for the positions of the vertices

$$X(N) = \left\lfloor \left(N + 1 + \frac{\gamma_2 - \gamma_1}{t_2}\right)\left(\sum_{k=1}^{2} \frac{t_1}{t_k}\right)^{-1} \right\rfloor a_1 + \left\lfloor \left(N + 1 + \frac{\gamma_1 - \gamma_2}{t_1}\right)\left(\sum_{k=1}^{2} \frac{t_2}{t_k}\right)^{-1} \right\rfloor a_2,$$

which proves Eq. (4) for the case where $p = 2$.

B. $p=3$

The vertices of the tilings in the case $p = 3$ are given by

$$X(M_1, M_2, M_3) = M_1 a_1 + M_2 a_2 + M_3 a_3. \quad (3'')$$

We subdivide this set of points into three subsets; the first, the vertices at the end of tilings of length a_1, the second, the vertices at the end of tilings of length a_2, and the third, the vertices at the end of tilings of length a_3. We denote these subsets of points, respectively, as $X(N_1)$, $X(N_2)$, and $X(N_3)$, where N_1, N_2, and N_3 denote the cardinal positions of the vertices in the tiling. We have that

$$X(N_1) = M_1 a_1 + \lfloor (M_1 t_1 + \gamma_1 - \gamma_2)/t_2 \rfloor a_2 + \lfloor (M_1 t_1 + \gamma_1 - \gamma_3)/t_3 \rfloor a_3, \quad (9a)$$

$$X(N_2) = \lfloor (M_2 t_2 + \gamma_2 - \gamma_1)/t_1 \rfloor a_1 + M_2 a_2 + \lfloor (M_2 t_2 + \gamma_2 - \gamma_3)/t_3 \rfloor a_3, \quad (9b)$$

$$X(N_3) = \lfloor (M_3 t_3 + \gamma_3 - \gamma_1)/t_1 \rfloor a_1 + \lfloor (M_3 t_3 + \gamma_3 - \gamma_2)/t_2 \rfloor a_2 + M_3 a_3, \quad (9c)$$

where

$$N_1 = M_1 + \lfloor (M_1 t_1 + \gamma_1 - \gamma_2)/t_2 \rfloor + \lfloor (M_1 t_1 + \gamma_1 - \gamma_3)/t_3 \rfloor, \quad (10a)$$

$$N_2 = M_2 + \lfloor (M_2 t_2 + \gamma_2 - \gamma_1)/t_1 \rfloor + \lfloor (M_2 t_2 + \gamma_2 - \gamma_3)/t_3 \rfloor, \quad (10b)$$

$$N_3 = M_3 + \lfloor (M_3 t_3 + \gamma_3 - \gamma_1)/t_1 \rfloor + \lfloor (M_3 t_3 + \gamma_3 - \gamma_2)/t_2 \rfloor. \quad (10c)$$

We can invert Eqs. (10): Since[24]

$$\lfloor \alpha + \beta \rfloor - 1 \leqslant \lfloor \alpha \rfloor + \lfloor \beta \rfloor \leqslant \lfloor \alpha + \beta \rfloor, \quad (11)$$

from Eq. (10a) we have

$$\left\lfloor M_1\left(\frac{t_1}{t_2} + \frac{t_1}{t_3}\right) + \frac{\gamma_1 - \gamma_2}{t_2} + \frac{\gamma_1 - \gamma_3}{t_3} \right\rfloor - 1 \leqslant N_1 - M_1 \leqslant \left\lfloor M_1\left(\frac{t_1}{t_2} + \frac{t_1}{t_3}\right) + \frac{\gamma_1 - \gamma_2}{t_2} + \frac{\gamma_1 - \gamma_3}{t_3} \right\rfloor.$$

From this it follows that

$$M_1 - 1 \leqslant \left\lfloor \left(N_1 + 1 + \frac{\gamma_2 - \gamma_1}{t_2} + \frac{\gamma_3 - \gamma_1}{t_3}\right) \times \left(\sum_{k=1}^{3} \frac{t_1}{t_k}\right)^{-1} \right\rfloor \leqslant M_1.$$

Since $M_1 - 1$ and M_1 are consecutive integers we have

$$M_1 = \left\lfloor \left(N_1 + 1 + \frac{\gamma_2 - \gamma_1}{t_2} + \frac{\gamma_3 - \gamma_1}{t_3}\right) \times \left(\sum_{k=1}^{3} \frac{t_1}{t_k}\right)^{-1} \right\rfloor + I, \quad (12)$$

where $I = 0$ or 1. This constant I is determined by taking the limit of Eq. (12) then t_3 goes to infinity. In this limit we obtain the $p = 2$ case and Eq. (12) must become identical with Eq. (7a). We find that $I = 0$ and

$$M_1 = \left\lfloor \left(N_1 + 1 + \frac{\gamma_2 - \gamma_1}{t_2} + \frac{\gamma_3 - \gamma_1}{t_3}\right)\left(\sum_{k=1}^{3} \frac{t_1}{t_k}\right)^{-1} \right\rfloor. \quad (13a)$$

In an analogous manner we invert Eq. (10b) and (10c) and obtain

$$M_2 = \left\lfloor \left(N_2 + 1 + \frac{\gamma_1 - \gamma_2}{t_1} + \frac{\gamma_3 - \gamma_2}{t_3}\right)\left(\sum_{k=1}^{3} \frac{t_2}{t_k}\right)^{-1} \right\rfloor , \tag{13b}$$

$$M_3 = \left\lfloor \left(N_3 + 1 + \frac{\gamma_1 - \gamma_3}{t_1} + \frac{\gamma_2 - \gamma_3}{t_2}\right)\left(\sum_{k=1}^{3} \frac{t_3}{t_k}\right)^{-1} \right\rfloor . \tag{13c}$$

Equations (13) are then substituted in Eq. (9). Denoting the coefficient of a_2 in Eq. (9a) by C_{12}, we have

$$\begin{aligned} C_{12} = {} & \Bigg\lfloor \Bigg\lfloor \left(N_1 + 1 + \frac{\gamma_2 - \gamma_1}{t_2} + \frac{\gamma_3 - \gamma_1}{t_3}\right) \\ & \times \left(\sum_{k=1}^{3} \frac{t_1}{t_k}\right)^{-1} \Bigg\rfloor \frac{t_1}{t_2} + \frac{\gamma_1 - \gamma_2}{t_2} \Bigg\rfloor , \\ C_{12} = {} & \Bigg\lfloor \left(N_1 + 1 + \frac{\gamma_1 - \gamma_2}{t_1} + \frac{\gamma_3 - \gamma_2}{t_3}\right) \\ & \times \left(\sum_{k=1}^{3} \frac{t_2}{t_k}\right)^{-1} - \Delta \frac{t_1}{t_2} \Bigg\rfloor , \\ C_{12} = {} & \Bigg\lfloor \left(N_1 + 1 + \frac{\gamma_1 - \gamma_2}{t_1} + \frac{\gamma_3 - \gamma_2}{t_3}\right) \\ & \times \left(\sum_{k=1}^{3} \frac{t_2}{t_k}\right)^{-1} \Bigg\rfloor + I, \end{aligned}$$

where I is some integer. On taking the limit of t_3 going to infinity we obtain the $p = 2$ case and this coefficient must then become identical with Eq. (8a). Consequently, $I = 0$ and

$$C_{12} = \left\lfloor \left(N_1 + 1 + \frac{\gamma_1 - \gamma_2}{t_1} + \frac{\gamma_3 - \gamma_2}{t_3}\right)\left(\sum_{k=1}^{3} \frac{t_2}{t_k}\right)^{-1} \right\rfloor .$$

In an analogous manner the coefficients C_{ij} of a_j in Eqs. (9), $i = 1,2,3$ corresponding to Eqs. (9a), (9b), and (9c), respectively, and $j = 1,2,3$, can be shown to be given by

$$C_{ij} = \left\lfloor \left(N_i + 1 + \sum_{m=1}^{3} \frac{\gamma_m - \gamma_j}{t_m}\right)\left(\sum_{k=1}^{3} \frac{t_j}{t_k}\right)^{-1} \right\rfloor .$$

Since the cardinal positions N_1, N_2, and N_3 of Eqs. (9) are mutually exclusive, we then can write a single equation for the positions of the vertices given by Eqs. (9),

$$X(N) = \sum_{j=1}^{3} \left\lfloor \left(N + 1 + \sum_{m=1}^{3} \frac{\gamma_m - \gamma_j}{t_m}\right)\left(\sum_{k=1}^{3} \frac{t_j}{t_k}\right)^{-1} \right\rfloor a_j, \tag{14}$$

proving the theorem, Eq. (4), for the case $p = 3$.

C. $p=g+1$

In the general case we subdivide the set of points representing the vertices of the tiling into p subsets $X(N_i)$, $i = 1,2,\ldots,p$, where N_i is the cardinal positions of the vertices at the end of tilings of length a_i. We have

$$X(N_i) = \sum_{j=1}^{p} C_{ij} a_j, \quad i = 1,2,\ldots,p, \tag{15}$$

where

$$C_{ij} = \lfloor (M_i t_i + \gamma_i - \gamma_j) t_j \rfloor , \quad i,j = 1,2,\ldots,p, \tag{16}$$

and

$$N_i = \sum_{j=1}^{p} C_{ij}, \tag{17}$$

Assuming that the theorem, Eq. (4), is valid for the case $p = g$, we have for the case $p = g$ that

$$C_{ij} = \left\lfloor \left(N_i + 1 + \sum_{m=1}^{g} \frac{\gamma_m - \gamma_j}{t_m}\right)\left(\sum_{k=1}^{g} \frac{t_j}{t_k}\right)^{-1} \right\rfloor , \quad i,j = 1,2,\ldots,g, \tag{18}$$

and in particular that $M_i = C_{ii}$ for $i = 1,2,\ldots,g$.

For the case $p = g + 1$, from Eqs. (16) and (17) we have

$$N_i = \sum_{j=1}^{g+1} \left\lfloor \frac{M_i t_i + \gamma_i - \gamma_j}{t_j} \right\rfloor ,$$

and therefore

$$N_i + 1 = M_i \sum_{j=1}^{g+1} \frac{t_i}{t_j} + \sum_{j=1}^{g+1} \frac{\gamma_i - \gamma_j}{t_j} + 1 - \sum_{j=1}^{g+1} \Delta_j,$$

where Δ_j is defined by

$$\lfloor (M_i t_i + \gamma_i - \gamma_j)/t_j \rfloor = (M_i t_i + \gamma_i - \gamma_j)/t_j - \Delta_j .$$

It follows that

$$\left\lfloor \left(N_i + 1 + \sum_{j=1}^{g+1} \frac{\gamma_j - \gamma_i}{t_j}\right)\left(\sum_{k=1}^{g+1} \frac{t_i}{t_k}\right)^{-1} \right\rfloor = M_i + I, \tag{19}$$

where I is some integer. Equation (19) in the limit that t_q goes to infinity, $q \neq i$, must become the expression for $M_i = C_{ii}$ given by Eq. (18), for $i = j$, in the case $p = g$. Consequently, $I = 0$ and for the case $p = g + 1$,

$$M_i = \left\lfloor \left(N_i + 1 + \sum_{m=1}^{g+1} \frac{\gamma_m - \gamma_i}{t_m}\right)\left(\sum_{k=1}^{g+1} \frac{t_i}{t_k}\right)^{-1} \right\rfloor .$$

Substituting this into Eq. (16) in the case of $p = g + 1$ we have

$$\begin{aligned} C_{ij} = {} & \Bigg\lfloor \Bigg\lfloor \left(N_i + 1 + \sum_{m=1}^{g+1} \frac{\gamma_m - \gamma_i}{t_m}\right) \\ & \times \left(\sum_{k=1}^{g+1} \frac{t_i}{t_k}\right)^{-1} \Bigg\rfloor \frac{t_i}{t_j} + \frac{\gamma_i - \gamma_j}{t_j} \Bigg\rfloor , \\ C_{ij} = {} & \Bigg\lfloor \Bigg\lfloor \left(N_i + 1 + \sum_{m=1}^{g+1} \frac{\gamma_m - \gamma_i}{t_m}\right) \\ & \times \left(\sum_{k=1}^{g+1} \frac{t_i}{t_k}\right)^{-1} \Bigg\rfloor \frac{t_i}{t_j} + \frac{\gamma_i - \gamma_j}{t_j} - \Delta \frac{t_i}{t_j} \Bigg\rfloor , \\ C_{ij} = {} & \left\lfloor \left(N_i + 1 + \sum_{m=1}^{g+1} \frac{\gamma_m - \gamma_j}{t_m}\right)\left(\sum_{k=1}^{g+1} \frac{t_j}{t_k}\right)^{-1} - \Delta \frac{t_i}{t_j} \right\rfloor , \\ C_{ij} = {} & \left\lfloor \left(N_i + 1 + \sum_{m=1}^{g+1} \frac{\gamma_m - \gamma_j}{t_m}\right)\left(\sum_{k=1}^{g+1} \frac{t_j}{t_k}\right)^{-1} \right\rfloor + I, \end{aligned}$$

where I is some integer. This equation is the limit that t_q, $q \neq i,j$ goes to infinity must become the expression for C_{ij} given in Eq. (18) in the case $p = g$. Consequently, $I = 0$. Since the sets of cardinal positions N_i, $i = 1,2,\ldots,g + 1$, are mutually exclusive, Eqs. (15), for $p = g + 1$ can be written as a single equation

$$X(N)=\sum_{j=1}^{g+1}\left\lfloor\left(N+1+\sum_{m=1}^{g+1}\frac{\gamma_m-\gamma_j}{t_m}\right)\left(\sum_{k=1}^{g+1}\frac{t_j}{t_k}\right)^{-1}\right\rfloor a_j,$$

proving the theorem given in Eq. (4).

IV. COMPARISON WITH THE RESULTS OF LITVIN AND LITVIN[23]

The positions $X'(N)$ of the vertices of a quasiperiodic tiling with $p=2$ basic tilings of lengths $a_1=\sin\theta$ and $a_2=\cos\theta$, constructed via a projection method was given in Eq. (9) of Ref. 23,

$$X'(N)=Na_2+(\lfloor\gamma^*+N/\alpha^*\rfloor-\lfloor\gamma^*\rfloor)(a_1-a_2),$$

where α^* and γ^* are constants. This can be rewritten as

$$X'(N)=\lfloor N/\alpha^*+\gamma^*-\lfloor\gamma^*\rfloor\rfloor a_1+(N-\lfloor N/\alpha^*+\gamma^*-\lfloor\gamma^*\rfloor\rfloor)a_2. \tag{20}$$

From Eq. (4) for $p=2$ we have

$$X(N)=\left\lfloor\left(N+1+\frac{\gamma_2-\gamma_1}{t_2}\right)\left(1+\frac{t_1}{t_2}\right)^{-1}\right\rfloor a_1+\left\lfloor\left(N+1+\frac{\gamma_1-\gamma_2}{t_1}\right)\left(1+\frac{t_2}{t_1}\right)^{-1}\right\rfloor a_2,$$

from which we derive

$$X(N)=\left\lfloor\left(N+1+\frac{\gamma_2-\gamma_1}{t_2}\right)\left(\frac{t_1+t_2}{t_2}\right)^{-1}\right\rfloor a_1+\left\{N-\left\lfloor\left(N+1+\frac{\gamma_2-\gamma_1}{t_2}\right)\left(\frac{t_1+t_2}{t_2}\right)^{-1}\right\rfloor\right\}a_2,$$

$$X(N)=\left\lfloor N\left(\frac{t_1+t_2}{t_2}\right)^{-1}+\frac{t_2+\gamma_2-\gamma_1}{t_1+t_2}\right\rfloor a_1+\left\{N-\left\lfloor N\left(\frac{t_1+t_2}{t_2}\right)^{-1}+\frac{t_2+\gamma_2-\gamma_1}{t_1+t_2}\right\rfloor\right\}a_2,$$

and

$$\begin{aligned}X(N)=&\left\lfloor N\left(\frac{t_1+t_2}{t_2}\right)^{-1}+\frac{t_2+\gamma_2-\gamma_1}{t_1+t_2}-\left\lfloor\frac{t_2+\gamma_2-\gamma_1}{t_1+t_2}\right\rfloor\right\rfloor a_1\\&+\left\{N-\left\lfloor N\left(\frac{t_1+t_2}{t_2}\right)^{-1}+\frac{t_2+\gamma_2-\gamma_1}{t_1+t_2}-\left\lfloor\frac{t_2+\gamma_2-\gamma_1}{t_1+t_2}\right\rfloor\right\rfloor\right\}a_2\\&+\left\lfloor\frac{t_2+\gamma_2-\gamma_1}{t_1+t_2}\right\rfloor(a_1+a_2).\end{aligned}\tag{21}$$

Comparing Eqs. (20) and (21) we have

$$\alpha^*=(t_1+t_2)/t_2,$$

$$\gamma^*=(t_2+\gamma_2-\gamma_1)/(t_1+t_2),$$

and the positions of the vertices given by Eqs. (20) and (21) are related by a change in origin

$$X(N)=X'(N)+\lfloor(t_2+\gamma_2-\gamma_1)/(t_1+t_2)\rfloor(a_1+a_2).$$

ACKNOWLEDGMENT

The author would like to thank N. G. de Bruijn for correspondence.

[1] D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn, Phys. Rev. Lett. **53**, 1951 (1984).
[2] N. G. de Bruijn, Nederl. Akad. Wetensch. Proc. Ser. A **84**, 39, 53 (1981).
[3] R. Penrose, Bull. Inst. Math. **10**, Appl. 266 (1974).
[4] M. Gardner, Sci. Am. **236**, 110 (1977).
[5] R. Penrose, Math. Intelligencer **2**, 32 (1979).
[6] A. L. Mackay, Sov. Phys. Cryst. **26**, 517 (1981).
[7] A. L. Mackay, Physica (Amsterdam) A **114**, 609 (1982).
[8] P. Kramer and R. Neri, Acta Cryst. Sect. A **40**, 580 (1984).
[9] J. E. S. Socolar, P. J. Steinhardt, and D. Levine, Phys. Rev. B **32**, 5547 (1985).
[10] F. Gahler and J. Rhyner, Phys. Rev. Lett. **55**, 2369 (1985).
[11] F. Gahler and J. Rhyner, J. Phys. A **19**, 267 (1986).
[12] D. Levine and P. J. Steinhardt, Phys. Rev. Lett. **53**, 2477 (1984).
[13] P. A. Kalugin, A. Yu. Kitaev, and L. S. Levitov, JETP Lett. **41**, 145 (1985).
[14] P. A. Kalugin, A. Yu. Kitaev, and L. S. Levitov, J. Phys. Lett. (Paris) **46**, 601 (1985).
[15] R. K. P. Zia and W. J. Dallas, J. Phys. A **18**, 341 (1985).
[16] M. Duneau and A. Katz, Phys. Rev. Lett. **54**, 2688 (1985).
[17] A. Katz and M. Duneau, J. Phys. (Paris) **47**, 181 (1986).
[18] V. Elser, Acta Cryst. Sect. A **42**, 36 (1986).
[19] P. J. Steinhardt and S. Ostlund, *The Physics of Quasicrystals* (World Scientific, Singapore, 1988).
[20] *Introduction to Mathematics of Quasi-crystals*, edited by M. V. Jaric (Academic, Boston, 1988).
[21] E. Bombieri and J. E. Taylor, J. Phys. (Paris) **47**, 19 (1986).
[22] N. G. de Bruijn, Nederl. Akad. Wetensch. Proc. Ser. A **84**, 27 (1981).
[23] S. Y. Litvin and D. B. Litvin, Phys. Lett. A **116**, 39 (1986).
[24] J. E. Shockley, *Introduction to Number Theory* (Holt, Rinehart, and Winston, New York, 1967), p. 19.

A nonlinear eigenvalue problem in astrophysical magnetohydrodynamics: Some properties of the spectrum

John A. Adam
Department of Mathematics and Statistics, Old Dominion University, Norfolk, Virginia 23529

(Received 11 March 1988; accepted for publication 9 November 1988)

The equations of ideal magnetohydrodynamics (MHD) with an external gravitational potential—a "magnetoatmosphere"—are examined in detail as a singular nonlinear eigenvalue problem. Properties of the spectrum are discussed with specific emphasis on two systems relevant to solar magnetohydrodynamics. In the absence of a gravitational potential, the system reduces to that of importance in MHD and plasma physics, albeit in a different geometry. This further reduces to a form isomorphic to that derived in the study of plasma oscillations in a cold plasma, Alfvén wave propagation in an inhomogeneous medium, and MHD waves in a sheet pinch. In cylindrical geometry, the relevant model equations are those for a diffuse linear pinch. The full system, including gravity, has been applied to the study of flare-induced coronal waves, running penumbral waves in sunspots, and linear wave coupling in a highly inhomogeneous medium. The structure of the so-called MHD critical layer and its contribution to the continuous spectrum is examined in detail for a model magnetoatmosphere, based on properties of the hypergeometric differential operator. The relationship of this singular region to critical layers in classical linear hydrodynamic stability theory is also discussed in the light of a specific model (in the Appendix).

I. INTRODUCTION

In this paper, a magnetoatmosphere is defined as a system, which, when linearly perturbed about a stable equilibrium, may support wave motion due to the combined restoring forces of compressibility, buoyancy, and magnetic fields.[1-3] Systems of this type have been of interest in a number of areas in solar physics: that of running penumbral waves in sunspots,[4] flare-induced coronal waves,[5] waves in "magnetic flux tubes,"[6,7] and the associated problem of coronal heating.[8,9] Indeed, there has been a resurgence of interest in the latter problem in recent years, because the outer solar atmosphere has been shown to be highly inhomogeneous inasmuch as it consists of myriads of magnetic flux tubes in complicated field configurations.[9] Such structures are of interest because they are believed to be capable of supporting waves that may be crucial in determining energy balance in the solar corona (and, by implication, the coronae of at least other stars similar in spectral type to the sun).

In the corona the effects of buoyancy and gravity-induced density stratification may often be neglected, and the ideal magnetohydrodynamic (MHD) equations form the basis for models of flux tubes and their stability.[7] Similar configurations on a smaller scale are of great importance in plasma physics, particularly in connection with nuclear fusion devices and their MHD stability.[10-12] In this area, there has also been interest in the problem of MHD stability in the presence of an external gravitational potential[13] (the Rayleigh–Taylor stability problem), not because gravitational forces *per se* play an important role in laboratory plasmas (usually they do not), but because acceleration forces that act on many plasma configurations can be simulated by a gravitational-type term in the equations.[14]

Another aspect of the formulation of the Rayleigh–Taylor problem, namely that of stable systems, is ideally suited to magnetoatmospheric wave propagation problems. In Secs. IV and VII specific attention will be paid to the spectral analysis of the differential operator arising in the study of an isothermal atmosphere permeated by a uniform horizontal magnetic field. The solution of this system has been utilized in a model of the low corona–chromosphere transition region, for comparison with observations of flare-induced coronal waves.[5] In that model, horizontally propagating disturbances were found to exist in a waveguide formed by the sudden density increase into the chromosphere below, and by the rapidly increasing Alfvén speed above in the corona.

Another topic of interest in solar physics for which models of this type are relevant is that of running penumbral waves.[4] These waves have been observed in H_α propagating horizontally outward across sunspot penumbrae. It appears that these are essentially gravity-modified magnetoacoustic waves (the ω_+ modes discussed in Sec. VI). A considerable number of papers have appeared on magnetoacoustic-gravity waves; detailed accounts and references may be found in the reviews by Thomas,[2] Campos,[15] and also the paper by Zhugzda and Dzhalilov.[16]

It should also be realized that magetoatmospheric or magnetoacoustic-gravity (MAG) systems can be supplemented by plane or rotational *shear* (and for such systems we will adopt the acronym SMAG). Some theoretical work has been carried out for SMAG waves (that is shear-modified MAG waves).[17-22] The study of shear in incompressible fluids is the basis for much of classical linear hydrodynamic stability theory, of interest in oceanography and meteorology alike.[23] So-called "critical layers" are associated with those levels z_c within the fluid such that the horizontal phase speed $\hat{c}$ is equal to the local flow speed $U(z_c)$. The topic is not only one of the relevance to geophysical fluid dynamics.[24] The study of waves in SMAG systems leads to a generalized critical layer that may be of significance in the study of Evershed flow in sunspots[18] (though it must be pointed out

 0022-2488/89/030744-13$02.50

that the physical significance of MHD critical layers is still in dispute).[15]

In general, in all these types of systems, the normal modes have a continuous as well as a discrete spectral component.[10–12,25–28] The former is not of the type associated with a spatially infinite operator domain[29] (though this may occur), but it comes from the presence of a frequency-dependent singularity in the governing ordinary differential equations in the finite (or infinite) spatial domain. Associated with this singular point (or set of such points) are singular normal modes,[30] or improper eigenfunctions.[31] The inclusion of such modes is necessary for completeness in a mathematical sense. An alternative to the normal mode approach in problems of this type is the initial value problem (IVP). Careful analysis shows, in general, that they yield identical results,[25,31] but that the IVP is often more straightforward from a physical point of view. The time-harmonic, Fourier-transformed system of partial differential equations, or the Fourier–Laplace-transformed system for the IVP, gives rise, in general, to an ordinary differential operator of the form

$$\frac{d}{dz}\left\{A(z,\hat{c})\frac{d}{dz}\right\} + B(z,\hat{c}) \tag{1.1}$$

together with such boundary conditions and/or initial conditions as may be appropriate. In (1.1) the terms $A(z,\hat{c})$ and $B(z,\hat{c})$ may be highly nonlinear in the eigenvalue $\hat{c}$. (Here $\hat{c}$ is a horizontal phase speed. In Sec. II the eigenvalue is $\omega^2 = \hat{c}^2k^2$, where k is a Fourier wavenumber.) However, for the case of incompressible shear flow A has the misleadingly simple form[30,32]

$$A(z,\hat{c}) = (U(z) - \hat{c})^2. \tag{1.2}$$

In the case of incompressible magnetic shear flow,[32]

$$A(z,\hat{c}) = \{(U(z) - \hat{c})^2 - a^2k^2\}, \tag{1.3}$$

$a(z)$ being the Alfvén speed (defined below). Clearly, MHD systems, of which (1.3) is illustrative and typical but by no means general, contain the hydrodynamic systems in the limit $a \to 0$. The difference between the solutions for the two systems is characterized by the nature of the coefficients A: in (1.2) for real $\hat{c}$, A is always non-negative, whereas in (1.3) A can change sign in the interior of the domain (see Ref. 33 for a detailed discussion of a specific example in this context). The solutions in the neighborhood of a critical layer at $z = z_c$ have been discussed elsewhere.[32,34] Another major difference between MAG problems and SMAG problems is that in the former case, as noted above, the operator $\mathbf{F}(\mathbf{r})$ is formally self-adjoint, and the eigenvalues ω^2 are therefore real, whereas in any shear-type problem the operator is formally non-self-adjoint, and eigenvalues may be complex. There is much in the literature on complex eigenvalue bounds for such systems, the most well known being Howard's celebrated "semi-circle theorem" for incompressible plane parallel shear flow under gravity.[35] There are a number of interesting modifications of the extensions to this result: further details may be found in Ref. 21.

Mathematically, the state of a magnetoatmospheric, MHD, or plasma medium can be represented as an element in Hilbert space, and the powerful theory of self-adjoint linear operators becomes available, together with many fruitful analogies in quantum mechanics.[10] Ironically, the latter subject can be regarded as more straightforward (relatively, at least!) and "classical" than when compared with the exceedingly rich structure inherent in the MHD systems discussed above.

There are several classification schemes for the spectrum of self-adjoint linear operators in Hilbert space, and, while some differences exist, for the purpose of this paper they are essentially equivalent. Two will be stated here since each one is germane to topics introduced in the main body of the paper.

In Eq. (2.11) we denote the operator $\rho_0^{-1}\mathbf{F} + \omega^2$ by

$$T_\lambda = T - \lambda I, \tag{1.4}$$

where I is the identity operator on the domain of $T = \rho_0^{-1}\mathbf{F}$ and $\lambda = -\omega^2$ here is real (since T is formally self-adjoint). If T_λ has an inverse

$$R_\lambda(T) = T_\lambda^{-1} = (T - \lambda I)^{-1}, \tag{1.5}$$

$R_\lambda(T)$ is called the resolvent operator.

If H is a Hilbert space and $T: D(T) \to X$, a regular value λ of T is a number (in general complex) such that (i) $R_\lambda(T)$ exists, (ii) $R_\lambda(T)$ is bounded, and (iii) $\overline{D(R_\lambda)} = X$, i.e., $R_\lambda(T)$ is defined on a set which is dense in X. The resolvent set $\rho(T)$ of T is the set of all regular values λ of T, and its complement $\sigma(T) = \mathscr{R} - \rho(T)$ in the real line is called the spectrum of T. A partition of the spectrum is as follows.[36,37]

(i) The point spectrum $P\sigma(T)$ is the set of λ such that $R_\lambda(T)$ does not exist. Equivalently, it is the set of λ such that T_λ is not injective, i.e., its null space $N(T_\lambda)$ is nontrivial. The point spectrum is thus the set of all eigenvalues.

(ii) The continuous spectrum $C\sigma(T)$ is the set of λ such that $R_\lambda(T)$ exists, its domain is dense in X, but $R_\lambda(T)$ is unbounded. Thus T_λ is an injective operator.

Returning to the original operator, the initial value problem would have the form

$$[\rho_0^{-1}\mathbf{F} + \omega^2]\boldsymbol{\phi} = \mathbf{S}, \tag{1.6}$$

where $\mathbf{S}$ contains the initial data. Here, ω^2 being in the point spectrum of $T = \rho_0^{-1}\mathbf{F}$ means that

$$[\rho_0^{-1}\mathbf{F} + \omega^2]\boldsymbol{\phi} = \mathbf{0} \tag{1.7}$$

possesses a nontrivial solution.

It is worth noting that the *discrete spectrum* $D\sigma(T)$ [often identified as being synonymous with $P\sigma(T)$] is the set of all isolated spectral points, excluding eigenvalues of infinite multiplicity. Thus $D\sigma(T) \subset P\sigma(T)$. Furthermore, if an element of $P\sigma(T)$ is not isolated from the $C\sigma(\mathrm{T})$, it is said to be in the point-continuous spectrum, $PC\sigma(T)$.

Consider the self-adjoint operator $Ty = (p(z)y')' + q(y)$, $a \leqslant z \leqslant b < \infty$. The two linearly independent solutions of

$$T_\lambda(y) = T(y) - \lambda s(z)y = 0 \tag{1.8}$$

are $\phi(z,\lambda)$ and $\psi(z,\lambda)$, and these satisfy the initial conditions

$$\begin{aligned} \phi(a,\lambda) &= -\alpha_2, \quad \phi'(a,\lambda) = \alpha_1/r(a),\\ \psi(a,\lambda) &= \alpha_1, \quad \psi'(a,\lambda) = \alpha_2/r(a), \end{aligned} \tag{1.9}$$

where $\alpha_1 \in R$, $\alpha_2 \in R$, and $|\alpha_2| + |\alpha_2| \neq 0$. Each solution of (1.8), except multiples of ψ, can be expressed as a multiple of

$$Y = \phi + m\psi, \quad m \in \mathscr{C}.$$

If, at a point $b_0 < \infty$, the boundary condition

$$\beta_1 y(b_0,\lambda) + \beta_2 r(b_0) y'(b_0,\lambda) = 0, \tag{1.10}$$

where $\beta_1 \in R$, $\beta_2 \in R$, and $|\beta_1| + |\beta_2| \neq 0$, is imposed, then for $\mathrm{Im}(\lambda) \neq 0$, as the ratio β_1/β_2 takes on all real values, m describes a circle in the complex λ plane, with center

$$S = -W(\phi,\psi;b_0)/W(\psi,\psi^*;b_0) \tag{1.11}$$

and radius

$$R = \frac{(\alpha_1^2 + \alpha_2^2)^{1/2}}{2|\mathrm{Im}(\lambda)| \int_a^{b_0} s(z) |\psi|^2 \, dz}. \tag{1.12}$$

Here W is the Wronskian of its arguments and ψ^* denotes the complex conjugate of ψ. As b_0 increases with λ fixed, R decreases and then the circle either approaches a *limit circle* $C_\infty(\lambda)$ or a *limit point* $m_\infty(\lambda)$.[38] In the former case all solutions are square-integrable over (a,∞) relative to the weight function $s(z)$. In the latter case only one solution, $y = \phi + m_\infty \psi$, is so square-integrable.

The paper is arranged as follows. Having introduced areas of application and spectral theoretic terminology in the present section, Sec. II deals with the basic equations of ideal MHD, and the fundamental reduced wave equations (2.16)–(2.18) is stated. In Sec. III the so-called exponential magnetoatmospheric wave equation is expressed in canonical hypergeometric form. The spectral problem is studied in Sec. IV, including a discussion of the number of zeros of the dispersion relation, and a point of accumulation of such zeros is identified. Conditions are derived under which the end points for the operator domain are limit point or limit circle in nature (in the sense of Weyl and Titchmarsh).

In Sec. V the effect of a *finite* upper boundary in the exponential magnetoatmospheric problem is discussed. Another important case—the constant parameter magnetoatmosphere—is examined in detail via its governing dispersion relation. In Sec. VI the effect of gravity on the hierarchical ordering of characteristic frequencies is identified, as are points of accumulation for the various discrete subspectra that are present in the system. Conditions under which the discrete spectra are Sturmian and anti-Sturmian are given. In the light of these properties, the spectrum of the exponential atmosphere is discussed further, and a fuller description of the nature of the spectrum is presented in Sec. VII.

Finally, the Appendix deals briefly with other material of relevance to the paper: the comparison of the spectral analysis of a shear flow system with that in Sec. IV.

In what follows, $p_0(z)$, $\rho_0(z)$, $\mathbf{B}_0(z)$, and $\psi_0(z)$ represent the equilibrium distributions of pressure, density, magnetic field intensity, and external potential, respectively. Linear perturbations to these basic field variables have no subscript (e.g., $p, \rho, \mathbf{B}, \psi$). The velocity perturbation field is $\mathbf{v}$, the ratio of specific heats for the medium is denoted by γ, the sound speed $c_0(z) = (\gamma p_0/\rho_0)^{1/2}$, and the Alfvén speed $a_0(z) = |\mathbf{B}_0|/\rho_0^{1/2}$, $a_x = B_x/\rho_0^{1/2}$, etc. Other variables are introduced as needed.

II. THE IDEAL MAGNETOHYDRODYNAMIC EQUATIONS

The equations of ideal magnetohydrodynamics are those of momentum, induction, isentropy, and continuity, respectively:

$$\rho\left[\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right] = -\nabla p - \mathbf{B}\times(\nabla\times\mathbf{B}) - \rho\,\nabla\psi, \tag{2.1}$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla\times(\mathbf{v}\times\mathbf{B}), \tag{2.2}$$

$$\left[\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right](p\rho^{-\gamma}) = 0, \tag{2.3}$$

$$\left[\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right]\rho = -\rho\nabla\cdot\mathbf{v}. \tag{2.4}$$

The equations are linearized about a static equilibrium defined by

$$\nabla p_0 + \mathbf{B}_0\times(\nabla\times\mathbf{B}_0) = -\rho_0\,\nabla\psi_0. \tag{2.5}$$

If the external potential ψ_0 corresponds to a uniform gravitational field, as assumed below, then $\nabla\psi_0 = -\mathbf{g} = -(0,0,-g)$, and the right-hand side of Eq. (2.5) is replaced by the vector $\rho_0\mathbf{g}$.

In terms of the linear Lagrangian displacement field $\boldsymbol{\xi}(\mathbf{r},t)$, defined by

$$\mathbf{v} = \frac{\partial \boldsymbol{\xi}}{\partial t}, \tag{2.6}$$

the equation of motion of ideal MHD is obtained by integrating and eliminating all dependent variables except $\boldsymbol{\xi}$. Thus

$$\frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} = \rho_0^{-1}\mathbf{F}(\boldsymbol{\xi}), \tag{2.7}$$

where the "force" operator $\mathbf{F}(\boldsymbol{\xi})$ is defined as

$$\mathbf{F}(\boldsymbol{\xi}) = \nabla(\gamma p_0 \nabla\cdot\boldsymbol{\xi} + \boldsymbol{\xi}\cdot\nabla p_0) - \mathbf{B}_0\times(\nabla\times\mathbf{Q}) - \mathbf{Q}\times(\nabla\times\mathbf{B}_0) + \nabla\cdot(\rho_0\boldsymbol{\xi})\nabla\psi_0, \tag{2.8}$$

where

$$\mathbf{Q} = \nabla\times(\boldsymbol{\xi}\times\mathbf{B}_0). \tag{2.9}$$

The operator $\rho_0^{-1}\mathbf{F}$ can be proved to be self-adjoint for many boundary conditions of interest in MHD.[39]

We now regard $\boldsymbol{\xi}(x,y,z,t)$ as an integral superposition of harmonic terms, with $\boldsymbol{\xi}$ and its first x derivative vanishing at infinity: thus

$$\boldsymbol{\xi} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\boldsymbol{\phi}(z;k)e^{ikx}\,dk, \tag{2.10}$$

where the horizontal wave vector $\mathbf{k}$ is described with respect to oriented axes such that $\mathbf{k} = (k,0,0)$. Thus, in general, the horizontal magnetic field has components given by

$$\mathbf{B}_0 = (B_x(z), B_y(z), 0).$$

Now Eq. (2.7) becomes

$$-\rho_0\omega^2\boldsymbol{\phi} = \mathbf{F}(\boldsymbol{\phi},k). \tag{2.11}$$

Eliminating all dependent variables except ϕ_z we obtain an equation of the form[40]

$$\frac{d}{dz}\left[A(z,\omega)\frac{d\phi_z}{dz}\right] + B(z,\omega)\phi_z = 0, \tag{2.12}$$

where, after some algebra, the coefficients $A(z,\omega)$ can be expressed as

$$A(z,\omega) = \frac{\rho_0(a_x^2 + a_y^2 + c_0^2)(\omega^2 - \omega_1^2(z))(\omega^2 - \omega_2^2(z))}{(\omega^2 - \omega_3^2(z))(\omega^2 - \omega_4^2(z))} \quad (2.13)$$

and

$$B(z,\omega) = \rho_0\left\{(\omega^2 - \omega_1^2(z)) - \frac{k^2 g^2(\omega^2 - \omega_1^2(z))}{(\omega^2 - \omega_3^2)(\omega^2 - \omega_4^2)} \frac{-k^2 g}{\rho_0} \times \frac{d}{dz}\left[\frac{\rho_0(a_y^2 + c_0^2)(\omega^2 - \omega_5^2)}{(\omega^2 - \omega_3^2)(\omega^2 - \omega_4^2)}\right]\right\}, \quad (2.14)$$

where

$$\omega_1^2(z) = a_x^2 k^2, \quad (2.15)$$

$$\omega_2^2(z) = a_x^2 c_0^2 k^2 (a_x^2 + c_0^2)^{-1}, \quad (2.16)$$

$$\omega_3^2(z) = \tfrac{1}{2}[k^2(a_x^2 + a_y^2 + c_0^2) - \Delta], \quad (2.17)$$

$$\omega_4^2(z) = \tfrac{1}{2}[k^2(a_x^2 + a_y^2 + c_0^2) + \Delta], \quad (2.18)$$

$$\omega_5^2(z) = a_x^2 c_0^2 k^2 (a_y^2 + c_0^2)^{-1}, \quad (2.19)$$

where

$$\Delta^2 = k^4(a_x^2 + a_y^2 + c_0^2)^2 - 4a_x^2 c_0^2 k^4. \quad (2.20)$$

The expressions ω_3^2 and ω_4^2 are thus seen to be the roots of the quadratic in ω^2 given by

$$\omega^4 - \omega^2 k^2(a_y^2 + a_y^2 + c_0^2) + a_x^2 c_0^2 k^4 = 0. \quad (2.21)$$

III. THE "EXPONENTIAL" MAGNETOATMOSPHERE

We now restrict ourselves further in system (2.12) by considering the special case of an isothermal atmosphere ($c_0 = c =$ constant) with a uniform magnetic field $(B_x, 0, 0)$. Under these circumstances the equilibrium density field is given in terms of a constant scale height $H\,[H = -\rho_0'(z)/\rho_0(z)]$ as

$$\rho_0(z) = \rho_0(0)e^{-z/H} \quad (3.1)$$

and the Alfvén speed

$$a_x(z) = a_x(0)e^{z/2H}. \quad (3.2)$$

(Hereafter, in this section we drop the subscript x for simplicity.) This magnetoatmosphere was examined by Nye and Thomas in connection with a model of flare-induced coronal waves,[5] and by Adam[41] in a study of critical layer behavior (see also Ref. 16). Equation (2.12) simplifies considerably [on multiplying by $\rho_0^{-1}(\omega^2 - c^2k^2)$] to become the equation

$$\{\omega^2c^2 + (\omega^2 - c^2k^2)a^2(0)e^{z/H}\}\phi'' - c^2\omega^2 H^{-1}\phi' + \{(\omega^2 - c^2k^2)(\omega^2 - a^2(0)e^{z/H}k^2) - g(g - c^2/H)k^2\}\phi = 0. \quad (3.3)$$

Let us make the following transformations of dependent and independent variables:

$$\phi(z) = e^{-kz}f(w), \quad (3.4)$$

$$w = w_0 e^{-z/H}, \quad (3.5)$$

where we have noted that since k^2 appears in Eq. (3.3), we can replace k by $|k|$ without loss of generality, so

$$K = |k|H, \quad (3.6)$$

and

$$w_0 = \omega^2c^2/a^2(0)(\tilde{\omega}_3^2 - \omega^2) \quad (3.7a)$$

$$= \{\beta^2(c^2k^2/\omega^2 - 1)\}^{-1}, \quad (3.7b)$$

where $\beta^2 = a^2(0)/c^2$. (See Fig. 1.)

Under these transformations Eq. (3.3) appears in the canonical form of the hypergeometric differential equation, namely,

$$w(1-w)\frac{d^2f}{dw^2} + \{c - (a+b+1)w\}\frac{df}{dw} - abf = 0, \quad (3.8)$$

or, in terms of Riemann's P symbol,

$$f = P\begin{bmatrix} 0 & \infty & 1 & \\ 0 & K + \tfrac{1}{2} - S & 0 & W \\ -2K & K + \tfrac{1}{2} + S & 0 & \end{bmatrix}. \quad (3.9)$$

Thus

$$a + b = c = 2K + 1 \quad (3.10)$$

and

$$ab = H^2\omega^2/c^2 + K + (\gamma - 1)K^2c^2/\gamma^2H^2\omega^2. \quad (3.11)$$

In Eq. (3.9), $S(K,\omega)$ is defined as [with $\mathrm{Re}(S) > 0$]

$$S = \{K^2(1 - N^2/\omega^2) + \tfrac{1}{4}(1 - \omega^2/\omega_c^2)\}^{1/2} \quad (3.12a)$$

$$= \{\tfrac{1}{4} - R(K,\omega)\}^{1/2}, \quad (3.12b)$$

where

$$R = K^2(N^2/\omega^2 - 1) + \omega^2/4\omega_c^2, \quad (3.13)$$

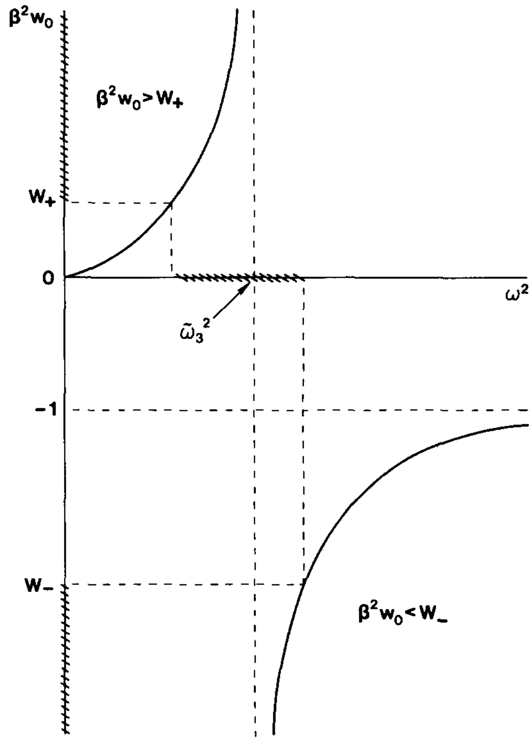


FIG. 1. The mapping $\omega^2 \to \beta w_0$ defined by Eq. (3.7b), where $\beta^2 = a^2(0)/c^2$. The expressions for $W_\pm$ are found in Sec. VII [Eq. (7.20)]. The hatched region on the ω^2 axis corresponds to the union of subacoustic and superacoustic horizontal phase speeds, which corresponds to (semi-infinite) intervals of $\beta^2 w_0$ (see Sec. VII for details).

and we have written S and R (called the *propagation number*) in terms of the Brunt–Vaisala frequency N, where

$$N^2 = -g\rho_0'/\rho_0 - g^2/c^2$$

$[= (\gamma - 1)g^2/c^2$ for this isothermal atmosphere], and the acoustic cutoff frequency $\omega_c = \gamma g/2c$. (See Fig. 2 for the physical significance of S^2 and its magnetic counterpart in Fig. 3.)

In terms of S, then

$$a = k - S + \tfrac{1}{2} \tag{3.14}$$

and

$$b = K + S + \tfrac{1}{2}. \tag{3.15}$$

There has been much interest in the literature concerning the significance of the singularities $\omega^2 = \omega_1^2$ and $\omega^2 = \omega_2^2$ in the differential equation (2.12) with A given by (2.13).[15,41,42] In the system considered in this section it is the latter singularity that is of concern, namely, when

$$(a^2 + c^2)\omega^2 = a^2c^2k^2, \tag{3.16}$$

or, in terms of w, $w = 1$.

The hypergeometric differential equation possesses regular singularities at $w = 0$, 1, and ∞, and that at $w = 1$ is often referred to as the cusp singularity in the MHD literature.[43] It is in some ways analogous to the classical "critical layer" singularity in hydrodynamic stability theory. A particular example of the critical layer problem is compared and contrasted with the cusp singularity in the Appendix.

The boundary conditions for the exponential magnetoatmosphere are taken to be the simple forms required by Nye and Thomas[5]; namely, a semi-infinite ($0 \leqslant z < \infty$) medium bounded below by the rigid plane $z = 0$, with the vertical

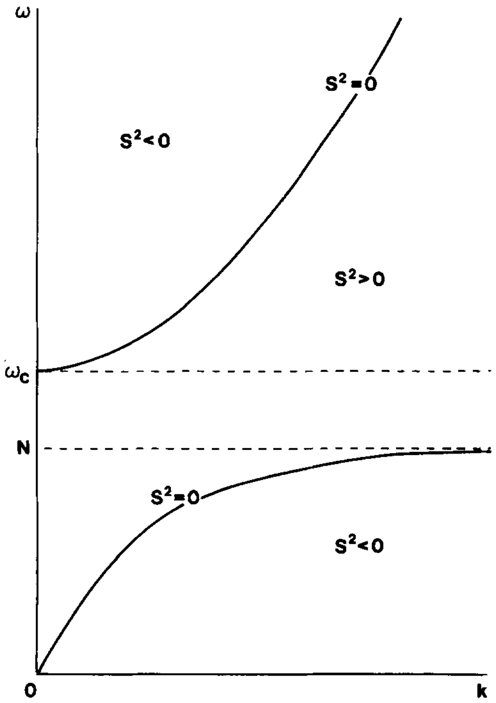


FIG. 2. The locus $S^2 = 0$ in the $\omega - k$ plane delineating regions of vertically propagating acoustic-gravity waves ($S^2 < 0$) from vertically nonpropagating waves ($S^2 > 0$), where S is defined by Eq. (3.12a).

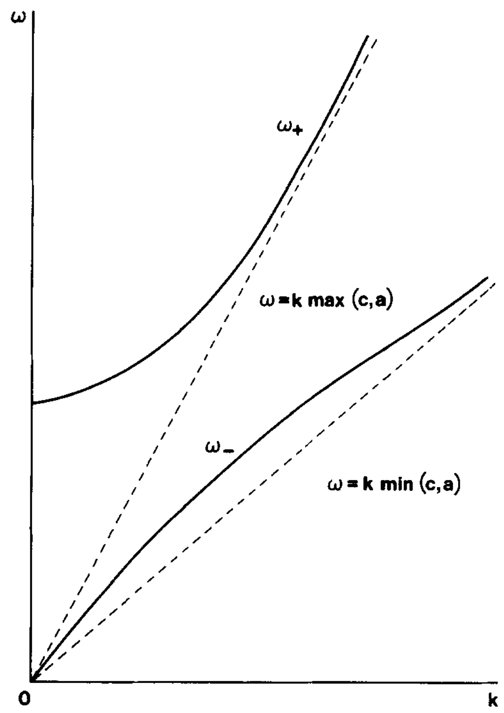


FIG. 3. The propagation–nonpropagation curves for the constant parameter magnetoatmosphere (Sec. VI). The curves are appropriate roots of the quadratic in ω^2 with $k_z = 0$ in Eq. (6.2) [see (6.4)]. Note the asymptotic behavior as illustrated by the broken lines. The ω_- curve may approach the asymptote from below (see Fig. 7).

displacement remaining finite as $z \to \infty$. The general solution of Eq. (3.8) convergent in the range $|w| < 1$ is, for constants E and D,

$$f(w) = D\,{}_2F_1(a,b;c;w) + Ew^{1-c} \times {}_2F_1(a - c + 1, b - c + 1; 2 - c; w) \tag{3.17}$$

provided $1 - c$ is not zero or a positive integer. Other representations of the solution will be used for $|w| > 1$. The boundedness condition at infinity implies that $f(w)$ must be regular at the origin $w = 0$, and since $1 - c < 0$ this means E must be zero. The other condition, $\phi(0) = 0$, yields $f(w_0) = 0$, i.e., for nontrivial f,

$${}_2F_1(a,b;c;w_0) = 0. \tag{3.18}$$

Hence we have that the roots of the eigenvalue equation correspond to the zeros of the hypergeometric function ${}_2F_1(a,b;c;w_0)$ in a plane cut along the interval $1 < w_0 < \infty$. Let us note several properties of the mapping (3.7b) (see Fig. 1): it is clearly a one-to-one mapping of the right $\hat{c}$ plane ($\hat{c} \neq c$) into the w_0 plane, with real $\hat{c}$ mapping onto real w_0 according to

$$\hat{c} \in (0^+, c^-) \to w_0 \in (0^+, \infty^-),$$

$$\hat{c} \in (c^+, \infty) \to w_0 \in (-\infty, [-\beta^{-2}]^-),$$

where $\hat{c} = \omega/k$ is a horizontal phase speed. Again, for real ω we may take $\omega > 0$ without loss of generality.

The domain of $w_0(\hat{c})$ is thus $\hat{c} \in (0^+, c^-) \cup (c^+, \infty)$, and the range is $w_0 \in (-\infty, -\beta^{-2}) \cup (0, \infty)$. For $w_0 \in (-\beta^{-2}, 0)$, $\hat{c}^2 < 0$.

IV. THE SPECTRAL PROBLEM FOR THE EXPONENTIAL MAGNETOATMOSPHERE

A. $S^2>0$

We utilize the notation and results of Van Vleck[44] to investigate the eigenvalue problem corresponding to Eq. (3.8). To this end we define $\lambda_1=|1-c|=2K$, $\lambda_2=|c-a-b|=0$, and $\lambda_3=|a-b|=2S$. Here, $E[q]$ is defined to be the integral part of q if $q>0$, and to be zero if $q\leqslant 0$. We now summarize the pertinent results of Van Vleck without proof in the form of a theorem.

Theorem 4.1: If $1-c<0$ and $S^2>0$, the number of roots of ${}_2F_1\ (a,b,c;w_0)$ in

(a) $(0,1)$ is $E\{\frac{1}{2}(\lambda_3-\lambda_1-\lambda_2+1)\}$;

(b) $(-\infty,0)$ is $E\{\frac{1}{2}(\lambda_2-\lambda_1-\lambda_3+1)\}$;

(c) $(1,\infty)$ is zero, unless $\frac{1}{2}\{\lambda_2+\lambda_3-\lambda_1+1\}$ is a positive integer.

If this is the case, there are four situations to consider: Let $m_i=E(\lambda_i)$, $i=1,2,3$. The number of complex roots of $F(a,b,;w_0)$ within the half-plane $\mathrm{Re}(w_0)>0$ if

(i) $m_1>m_2+m_3$, is zero;

(ii) $m_2>m_1+m_3$, is $E\{U/2\}$,

where

$$U=E\{\tfrac{1}{2}(\lambda_2+\lambda_3-\lambda_1+1)\}-E\{\tfrac{1}{2}(\lambda_2-\lambda_3-\lambda_1+1)\},$$

unless $\lambda_2-\lambda_1-\lambda_3$ is an odd integer, in which case the number of complex roots is zero;

(iii) $m_3>m_1+m_2$

[this is result (ii) with subscripts 2 and 3 interchanged];

(iv) $m_i\leqslant m_j+m_k$, for all $i,j,k=1,2,3$, is

$$E\{\tfrac{1}{2}(\lambda_2+\lambda_3-\lambda_1+1)\}.$$

We now examine each of (a), (b), and (c) in turn for $R<\frac{1}{4}$ and hence S real.

Lemma 4.1: $F(w_0)$ has no zeros in $(0,1)$.

Proof: The number of roots in $0<w_0<1$ is given by $E[S-K+\frac{1}{2}]$. For exactly n real roots, it is necessary that

$$n+1>S-K+\tfrac{1}{2}\geqslant n\geqslant 1. \tag{4.1}$$

It can be shown that this inequality is inconsistent with the model under consideration for which $N^2>0$ and $K\geqslant 0$. Hence there are no roots of $F(w_0)$ for $w_0\in(0,1)$.

Lemma 4.2: $F(w_0)$ has no zeros in $(-\infty,0)$.

Proof: The number of roots in $-\infty<w_0<0$ is given by $E[\frac{1}{2}-S-K]$. Again, for exactly n real roots, we require $n+1>\frac{1}{2}-S-K\geqslant n$.

This clearly cannot be satisfied for any integer $n\geqslant 1$ since both S and K are non-negative. The result follows.

Lemma 4.3: $F(w_0)$ has no zeros in $(1,\infty)$.

Proof: There are no zeros unless $S-K+\frac{1}{2}=2p+1$, $p=0,1,2,\ldots$, i.e.,

$$S=p+K+\tfrac{1}{2}. \tag{4.2}$$

As noted in Theorem 4.1, there are four cases to consider. Since, however, the operator $\rho_0^{-1}\mathbf{F}(\xi)$ in Eq. (2.7) is formally self-adjoint, we know the eigenvalues ω^2 are real, but we can show that there are no real eigenvalues either in each category. Details can be found elsewhere.[45] Hence in summary, we have the following theorem.

Theorem 4.2: $F(w_0)$ has no zeros in $w_0\in(0,1)\cup(1,\infty)\cup(-\infty,0)$ for $S^2(\omega,K)>0$.

B. $S^2<0$

Theorem 4.3: Let $R(K)>\frac{1}{4}$, and $S=i\mu$, $\mu=(R-\frac{1}{4})^{1/2}$. Then there exists an infinite number of zeros of (3.18) for $w_0\in(1,\infty)$, and they accumulate at (or diverge to) ∞.

Proof: Since we are interested in $|w_0|>1$ we use the linear transformation formula[46] [valid for $|\arg(-w_0)|<\pi|$], noting that, since w_0 is real, the second term is the complex conjugate of the first, so we may rewrite the eigenvalue equation (3.18) as

$$\mathrm{Re}\{A(-w_0)^{i\mu-(K+1/2)}\,{}_2F_1(K+\tfrac{1}{2}-i\mu,\tfrac{1}{2}-K-i\mu,1-2i\mu;w_0^{-1})\}=0, \tag{4.3}$$

where

$$A=\Gamma(2K+1)\Gamma(2i\mu)/\{\Gamma(K+\tfrac{1}{2}+i\mu)\}^2,\quad \text{for } |w_0|>1. \tag{4.4}$$

It can be shown that, neglecting real multiplicative factors, this result reduces to

$$\mathrm{Re}[[1+Bw_0^{-1}]\times\exp i\{\mu\ln w_0-\pi(K+1)+O(\mu^3)\}]=0, \tag{4.5}$$

where $B=B_r+iB_i$ with

$$B_r=(\alpha+2\mu^2)/(1+4\mu^2),$$
$$B_i=\mu(2\alpha-1)/(1+4\mu^2),$$

and

$$\alpha=\tfrac{1}{4}-K^2-\mu^2;$$

thus, on identifying

$$\phi=\mu\ln w_0-\pi(K+1)+O(\mu^3), \tag{4.6}$$

Eq. (4.5) is satisfied when

$$\cot\phi=\frac{B_i}{w_0+B_r}=\frac{\mu(2\alpha-1)}{\alpha+2\mu^2+(1+4\mu^2)w_0}. \tag{4.7}$$

This is a limiting eigenvalue relation for $\mu>0$ ($\mu\ll 1$ for simplicity, but this is not necessary) as $w_0\to\infty$. For given μ and K, the right-hand side is a monotonically increasing function of w_0, tending to 0^- as $w_0\to\infty$ if $\alpha+2\mu^2>0$. Whatever the value of this right-hand side, the left-hand side attains it an infinite number of times in this limit, proving the theorem.

Theorem 4.4: In terms of the parameter $\Lambda=-S^2$, there is a continuous spectrum in $\Lambda\in(0,\infty)$, and a point spectrum (which may be null) in $\Lambda\in(-\infty,0)$.

Proof: We make use of some results of Titchmarsh.[47] Starting with the canonical form (3.8) of the hypergeometric equation, let

$$w=-\sinh^2\tfrac{1}{2}\bar{z} \tag{4.8}$$

and

$$f=r(\bar{z})u, \tag{4.9}$$

where

$$r(\tilde{z}) = \{(e^{\tilde{z}}-1)/(e^{\tilde{z}}+1)\}^{1/2+(1/2)(a+b)-c} \times \sinh^{-(1/2)(a+b)}\tilde{z}. \quad (4.10)$$

Then Eq. (3.8) reduces to

$$\frac{d^2u}{d\tilde{z}^2} + \{\Lambda - q(\tilde{z})\}u = 0, \quad (4.11)$$

where

$$\Lambda = ab - \tfrac{1}{4}(a+b)^2 = -S^2 \quad (4.12)$$

and

$$q(\tilde{z}) = [2(a+b-1)(2c-1-a-b)\cosh\tilde{z} + 2(a+b)^2 - 4c(a+b) + (1-2c)^2]/(4\sinh^2\tilde{z}) \quad (4.13)$$

(we can simplify these expressions by using $c = a+b$ for this particular problem; but for complete generality at this stage we do not do so). From a theorem of Titchmarsh,[47] since $q(\tilde{z}) \to 0$ as $|\tilde{z}| \to \infty$, the result is established.

In fact, for $\Lambda \in (0,\infty)$, the continuous spectrum contributes to the expansion of some $g(\tilde{z})$ an amount

$$\frac{1}{2^{a+b+2}\pi}\int_0^\infty \left|\frac{\Gamma(a)\Gamma(c-b)}{\Gamma(c)\Gamma(a-b)}\right|^2 \times \frac{u(\tilde{z},\Lambda)d\Lambda}{\sqrt{\Lambda}}\int_0^\infty u(y,\Lambda)g(y)dy, \quad (4.14)$$

for $c > 2$ (see Theorem 4.5). There will also be a finite number of simple poles on the negative real axis when $a = -n$ or $c - b = -n$, $n = 0,1,2,\ldots$, with residues $(-1)^n/n!$. Since $a = K - S + \frac{1}{2} = c - b$, this means, for poles to exist, that $S > 0$ and hence $S^2 > 0$. (This also follows from the fact that $\Lambda < 0$ implies $S^2 > 0$.) For the boundary conditions considered in Sec. III, we know from Theorem 4.2 that there are no roots of the dispersion relation in this case, i.e., the point spectrum is null.

Theorem 4.5: For Eq. (4.11) the origin ($\tilde{z} = 0$; $w = 0$) is of limit circle type if $c < 2$ ($K < \frac{1}{2}$), otherwise it is of limit point type. The point $w = 1$ is of limit circle type if $0 < a + b - c + 1 < 2$; this is always true for the problem at hand.

Proof: As $\tilde{z} \to 0$, $r(\tilde{z}) \sim \tilde{z}^{1/2-c}$, $w \sim z^2$,

$$f(w) \to 1, \quad \text{so } u(\tilde{z}) \to \tilde{z}^{c-1/2}.$$

A second linearly independent solution corresponding to Eq. (3.8) as $w \to 0$ is $f(w) \to w^{1-c}$: thus

$$u(\tilde{z}) \to \tilde{z}^{c-1/2}(\tilde{z}^2)^{1-c} \to \tilde{z}^{3/2-c}.$$

Examining the square-integrability of these asymptotic solutions, we find

$$\int_0 \tilde{z}^{2c-1}\,d\tilde{z} \sim \frac{\tilde{z}^{2c}}{2c}$$

and

$$\int_0 \tilde{z}^{3-2c}\,d\tilde{z} \sim \frac{\tilde{z}^{4-2c}}{4-2c}.$$

These are both in $\mathscr{L}^2(0,1)$ if (i) $c > 0$ and (ii) $4 - 2c > 0$. Hence the origin is limit circle type if $0 < c < 2$. This corresponds to $K < \frac{1}{2}$, since K is non-negative by definition.

The transformation $\tilde{w} = 1 - w$ in (3.8) yields the hypergeometric equation in parameters a, b, and $\tilde{c}$, where $\tilde{c} = a + b - c + 1$; thus the above analysis holds for $\tilde{w} \to 0$, i.e., $w \to 1$. Both solutions are in $\mathscr{L}^2(0,1)$ if $0 < \tilde{c} < 2$. Since $a + b = c$ for the problem at hand, the point $w = 1$ is of the limit-circle type irrespective of the value of K. The cases in which $a + b = c$ for the hypergeometric equation give rise to logarithmic solution behavior, as is well known. Indeed, the general solution of (3.8) in the neighborhood of $w = 1^-$ ($\tilde{\omega} = 0^+$) is of the form

$$f(\tilde{w}) \sim f_1(\tilde{w})\ln\tilde{w} + \sum_{r=1}^\infty c_r\tilde{w}^r. \quad (4.15)$$

Since $f_1(\tilde{w}) \to 1$ as $\tilde{w} \to 0$,

$$\int_0 f^2(\tilde{w})d\tilde{w} \sim \int_0 \ln^2\tilde{w}\,d\tilde{w} \sim \{\tilde{w}[\ln^2\tilde{w} - 2\ln\tilde{w} + 2]\}|_0 < \infty,$$

thus exhibiting the limit-circle behavior at that point.

Finally, we consider the model equation in the light of some results of Dunford and Schwartz.[48] If we denote the hypergeometric differential operator in (3.8) by $H_w^{a,b,c}$, then the transformations

$$\alpha = \tfrac{1}{2}(c-1), \quad \beta = \tfrac{1}{2}(a+b-c), \quad t = 2w - 1, \quad (4.16)$$

and

$$f(w) = (t+1)^{-\alpha}(t-1)^{-\beta}\phi(t), \quad (4.17)$$

give

$$H_w^{a,b,c}\{f\} \to L_t^{\alpha,\beta}\{\phi\}, \quad (4.18)$$

where

$$L_t^{\alpha,\beta} \equiv -\left[\frac{d}{dt}\right](1-t^2)\left[\frac{d}{dt}\right] + \frac{2\alpha^2}{1+t} + \frac{2\beta^2}{1-t}, \quad (4.19)$$

on $(-1,1)$, $\alpha \geqslant 0$, $\beta \geqslant 0$.

Properties of operator $L_t^{\alpha,\beta}$ were discussed by Dunford and Schwartz, and we merely note those that are germane to the analysis presented here. For any λ, $\mathrm{L} - \lambda$ has regular singularities at -1, 1, and ∞ with respective exponents $\{\alpha, -\alpha\}$, $\{\beta, -\beta\}$, and $\{\frac{1}{2} + [\lambda + \frac{1}{4}]^{1/2}, \frac{1}{2} - [\lambda + \frac{1}{4}]^{1/2}\}$. There are three cases to consider.

(i) $\alpha \geqslant \frac{1}{2}, \beta \geqslant \frac{1}{2}$. Clearly $\alpha = K$ and $\beta = 0$ for the problem at hand, so this case does not apply. In this case though, the deficiency indices of L are both zero, so L gives to a unique self-adjoint operator in Hilbert space.

(ii) $\alpha \geqslant \frac{1}{2}$, $0 \leqslant \beta < \frac{1}{2}$. (The transformation $t \to -t$ interchanges α and β so there is no loss of generality.) In this case ($K \geqslant \frac{1}{2}, \beta = 0$) the deficiency indices of L are both unity: one boundary condition is required to define a unique self-adjoint operator in Hilbert space.

(iii) $0 \leqslant \alpha < \frac{1}{2}$, $0 \leqslant \beta < \frac{1}{2}$. All solutions of $(L - \lambda)\phi = 0$ are square-integrable at both end points. The deficiency indices of L are both 2, so two boundary conditions must be imposed to obtain a self-adjoint operator in Hilbert space.

V. A BOUNDED MAGNETOATMOSPHERE

Consider the domain $0 \leqslant z \leqslant d < \infty$ for the system discussed in Sec. III. The boundary conditions

$\phi(z=0)=0=\phi(z=d)$ are chosen, corresponding to $w\in[w_0,\alpha w_0]$, where $\alpha=e^{-d/H}<1$. If $|w_0|<1$, the boundary conditions yield, for constants D and E not both zero,

$$D\,{}_2F_1(a,b;c;w_0)+Ew_0^{1-c}{}_2F_1(a-c+1,b-c+1;2-c;w_0)=0 \tag{5.1}$$

and

$$D_2F_1(a,b;c;\alpha w_0)+E(\alpha w_0)^{1-c}{}_2F_1(a-c+1,b-c+1;2-c;\alpha w_0). \tag{5.2}$$

The necessary and sufficient condition for nontrivial solutions yields the following dispersion relation for the discrete spectrum:

$$w_0^{1-c}\{\alpha^{1-c}F_1(w_0)F_2(\alpha w_0)-F_1(\alpha w_0)F_2(w_0)\}=0, \tag{5.3}$$

where F_1 and F_2 denote the first and second hypergeometric functions in each of the expressions (5.1) and (5.2).

The general solution for $\phi(z)$ valid for $|w|>1$ is the linear combination

$$Gw^{-a}{}_2F_1(a,a-c+1;a-b+1;w^{-1})+Hw^{-b}{}_2F_1(b,b-c+1;b-a+1;w^{-1}) \tag{5.4}$$

or

$$Gw^{-a}F_1(w^{-1})+Hw^{-b}F_2(w^{-1}). \tag{5.5}$$

If $|\alpha w_0|$ and hence $|w_0|$ is greater than unity, the portions of discrete spectrum residing outside any continua are described in terms of solutions of

$$w_0^{-(a+b)}[\alpha^{-b}F_1(w_0^{-1})F_2(\alpha^{-1}w_0^{-1})-\alpha^{-a}F_1(\alpha^{-1}w_0^{-1})F_2(w_0^{-1})]=0. \tag{5.6}$$

It is apparent that, in the light of the analysis carried out in Sec. IV B for $S=i\mu$, expressions (5.3) and (5.6) indicate that the singular point $w_0=\infty$ is not a point of accumulation for that part of the spectrum. The case with $|\alpha w_0|<1$ and $|w_0|>1$ is apparently different, containing as it does the major case of interest in Sec. III ($\alpha=0$). Consider Eq. (3.8) for $|w|=|\alpha w_0|<1$, and the appropriate analytic continuation of this equation for $|w|=|w_0|>1$, subject to the boundary conditions $\phi(0)=\phi(d)=0$:

$$DF_1(\alpha w_0)+E(\alpha w_0)^{1-c}F_2(\alpha w_0)=0 \tag{5.7}$$

and

$$D\lambda_1(w_0^{-1})+E(w_0)^{1-c}\lambda_2(w_0^{-1})=0, \tag{5.8}$$

where

$$\lambda_1=\frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}\times(-w_0)^{-a}{}_2F_1(a,1-c+a;1-b+a;w_0^{-1})+\frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}\times(-w_0)^{-b}{}_2F_1(b,1-c+b;1-a+b;w_0^{-1}), \tag{5.9}$$

with a similar expression for λ_2. From these two equations we have the condition

$$w_0^{1-c}\{F_1(\alpha w_0)\lambda_2-\alpha^{1-c}F_2(\alpha w_0)\lambda_1\}=0. \tag{5.10}$$

The expressions above for λ_1 and λ_2 both contain, in particular, a term $w_0^{i\mu}=\exp(i\mu\ln w_0)$. This implies, as before, the existence of a sequence of eigenvalues accumulating at $w_0=\infty$, *but*, for this representation to be valid, we require that $|\alpha w_0|<1$, which constrains α, and hence the position of the upper boundary, to change accordingly. This rapidly degenerates to the case of the semi-infinite magnetoatmosphere discussed previously.

VI. THE CONSTANT PARAMETER MAGNETOATMOSPHERE

In order for the Alfvén speed to be uniform, it is necessary that $(d/dz)[|\mathbf{B}_0|^2\rho_0^{-1}]=0$. This condition, together with the equation of magnetohydrostatic equilibrium in the form

$$\frac{d}{dz}\left[p_0+\frac{\mathbf{B}_0\cdot\mathbf{B}_0}{2}\right]=-\rho_0 g \tag{6.1}$$

and the perfect gas equation, yield for constant sound speed a density scale height $H=(2c^2+\gamma a^2)/2\gamma g$. For simplicity, we let $B_y=0$, so that the wave vector $\mathbf{k}$ lies in the plane of the magnetic field and gravity vectors. The coefficients A (2.13) and B (2.14) are independent of z, and a Fourier analysis of Eq. (2.12) in the z direction yields the following expression:

$$[\omega^2-a^2k^2][\omega^4-[c^2+a^2]K^2\omega^2+c^2a^2k^2K^2+N^2c^2k^2]=0. \tag{6.2}$$

The first term represents the decoupled Alfvén mode, while the other term represents the fast and slow magnetoacoustic modes modified by gravity, as is well known.[4] In these expressions,

$$K^2=k^2+k_z^2+1/4H^2, \tag{6.3}$$

k_z being the vertical component of the wavenumber $\mathbf{k}=(k,0,k_z)$, and $N^2=g(1/H-g/c^2)$ being the square of the Brunt–Vaisala frequency, when positive. The fast and slow modes correspond to the plus and minus roots $\omega_\pm^2$ defined by

$$\omega_\pm^2=\tfrac{1}{2}[c^2+a^2]K^2\times\left[1\pm\left\{1-\frac{4c^2k^2(N^2+a^2)K^2}{(c^2+a^2)^2K^4}\right\}^{1/2}\right] \tag{6.4}$$

(see Fig. 3). It is a straightforward matter to show that

$$\omega_2^2=c^2a^2k^2/(c^2+a^2)\leqslant\omega_-^2\leqslant\omega_+^2\leqslant K^2(a^2+c^2). \tag{6.5}$$

In the absence of gravity ($N=0$), it is always the case that $\omega_-^2\leqslant\omega_1^2=a^2k^2\leqslant\omega_+^2$ also. When $N\neq0$, these inequalities still hold provided $k_z^2\geqslant K_z^2=N^2c^2a^{-4}-(4H^2)^{-1}$. For k_z in the range $(0,K_z)$, ω_-^2 is greater than $\omega_1^2=a^2k^2$. However, $\omega_2^2<\max\{\omega_1^2,\omega_-^2\}$ always.

A related matter is that concerning the points of accumulation of the discrete spectra corresponding to the above "+ and −" modes. In a magnetoatmospheric "slab" bounded by the rigid infinite planes $z=0$ and $z=d$, we may identify the wavenumber k_z as $n\pi/d$, $n-1$ being the number of interior nodes of the eigenfunction $\phi(z)$. Thus increasing k_z is heuristically equivalent to increasing the node number n for the specified value of d. Retaining terms of $O(k_z^{-2})$ as $k_z \to \infty$, k being fixed, it transpires from (6.4) that[11]

$$\omega_+^2 \approx (c^2 + a^2)k_z^2 \to \omega_f^2 = \infty, \tag{6.6}$$

so that ∞ is an accumulation point of the fast mode discrete spectrum, and

$$\omega_-^2 \to c^2a^2k^2/(c^2 + a^2) = \omega_2^2. \tag{6.7}$$

The slow mode discrete spectrum accumulates at the point ω_2^2 (see Fig. 4). We may notice the interesting fact that if $N^2 \geqslant 0$, $\omega_-^2 \downarrow$ as $k_z \uparrow$ such that $\omega_-^2 \to \omega_2^{2+}$. This corresponds to an *anti-Sturmian* discrete subspectrum. If, however, $N^2 < 0$, $\omega_-^2 \uparrow$ as $k_z \uparrow$. Thus an unstable gravitational equilibrium gives rise to a *Sturmian* discrete subspectrum. Furthermore, if in this case a^2 is sufficiently small, ω^2 may become negative for some range of finite k_z. This is related to the interchange or hydromagnetic Rayleigh–Taylor instability,[9] stabilized here for sufficiently large k_z.

The behavior of the (point) spectrum for the constant parameter magnetoatmosphere lays the basis for an appreciation of what the spectral theory discussed earlier in Sec. IV means in physical terms. A piecewise uniform model will be used in the next section.

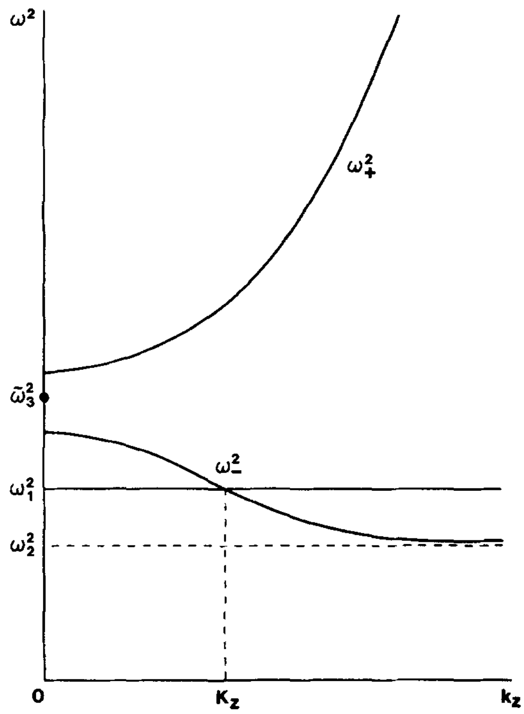


FIG. 4. The $\omega^2 - k_z$ diagram for the constant-parameter magnetoatmosphere (Sec. VI). For the upper branch (the "fast" mode modified by gravity), $\partial\omega^2/\partial k_z > 0$ indicating a Sturmian discrete spectrum accumulating at infinity. For the lower branch (the "slow" mode modified by gravity), $\partial\omega^2/\partial k_z < 0$, indicating an anti-Sturmian discrete spectrum accumulating at $\omega^2 = \omega_2^2$. Note that this point of accumulation $\to 0$ as $a \to 0$, so compressibility-modified gravity waves accumulate at $\omega^2 = 0$. The horizontal line $\omega^2 = \omega_1^2$ represents the infinitely degenerate Alfvén mode (decoupled here for $k_y = 0$).

VII. SPECTRUM OF THE EXPONENTIAL MAGNETOATMOSPHERE: DISCUSSION

It has already been noted that the appropriate dispersion or eigenvalue relation for the discrete spectrum is Eq. (3.18), i.e.,

$${}_2F_1(a,b;c;w_0) = 0,$$

provided $|w_0| < 1$. If $|w_0| > 1$ then the analytic continuation (4.3) is used, being set equal to zero. In what follows, these equations will be referred to as EI and EII, respectively. Since it is precisely these equations that have been used in a model of flare-induced coronal waves, the notation of Ref. 5 will be adopted in the remainder of this section for ease of identification. Thus if

$$\Omega = \omega H/c, \tag{7.1}$$

then

$$w_0 = \Omega^2/\beta^2(K^2 - \Omega^2). \tag{7.2}$$

For EI, the condition $|w_0| < 1$ yields the conditions

$$0 < \Omega^2 < \beta^2K^2(1 + \beta^2)^{-1} < K^2 \tag{7.3}$$

and if $\beta^2 > 1$ (which is permitted in this model for which a is not constant)

$$\Omega^2 > \beta^2K^2(\beta^2 - 1)^{-1} > K^2. \tag{7.4}$$

For EII, the conditions are

$$K^2 > \Omega^2 > \beta^2K^2(1 + \beta^2)^{-1} \tag{7.5}$$

and if $\beta^2 > 1$

$$K^2 < \Omega^2 < \beta^2K^2(\beta^2 - 1)^{-1}. \tag{7.6}$$

These regions are illustrated in Fig. 5. In terms of the original variables, the four inequalities (7.3)–(7.6) become, respectively,

$$0 < \omega^2 < \frac{a^2(0)c^2k^2}{a^2(0) + c^2} = \omega_2^2(0) < c^2k^2, \tag{7.7}$$

$$a^2(0) > c^2, \quad \omega^2 > \frac{a^2(0)c^2k^2}{a^2(0) - c^2} > c^2k^2, \tag{7.8}$$

$$c^2k^2 > \omega^2 > \frac{a^2(0)c^2k^2}{a^2(0) + c^2}, \tag{7.9}$$

and if

$$a^2(0) > c^2, \quad c^2k^2 < \omega^2 < \frac{a^2(0)c^2k^2}{a^2(0) - c^2}. \tag{7.10}$$

The union of these four regions [or two of them if $a^2(0) < c^2$] represents the domain of the discrete spectrum outside the continuum region. This continuum arises through the term $(\omega^2 - \omega_2^2(z))$ in Eqs. (2.13) and (2.14); in the more general, case of variable sound speed $c(z)$ and $B_y \neq 0$ there will, in general, be four continua, which in weakly inhomogeneous media may be separated from each other by portions of the discrete spectrum and regions of spectral nonmonotonicity,[49] e.g., $\{\omega_3^2(z)\}$ and $\{\omega_4^2(z)\}$.

The expression

$$\omega_2^2(z) = a^2(0)c^2e^{z/H}k^2/[a^2(0)e^{z/H} + c^2]$$

is a monotonically increasing function in $z \in [0,\infty)$, $\omega_2^2(0) \leqslant \omega_2^2(z) \leqslant \omega_2^2(\infty) = c^2k^2$ (see Fig. 6). Hence this may be inverted to give a monotonically increasing profile

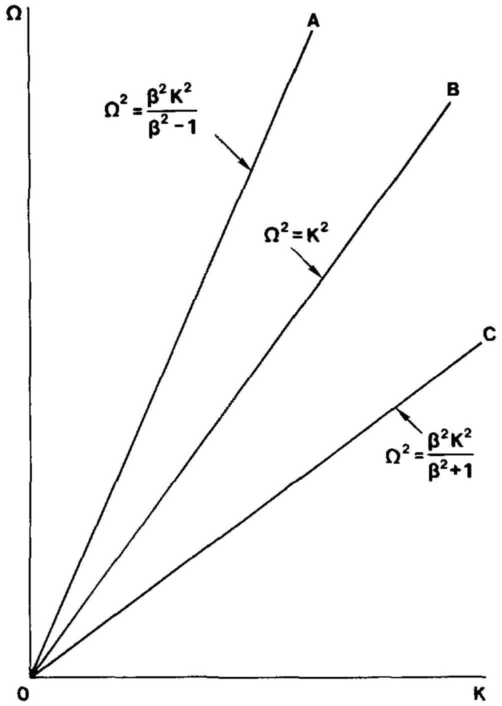


FIG. 5. The Ω-K plane (or dimensionless ω-k plane) illustrating the domains of validity (7.3)–(7.6) of the dispersion relations (3.18) (EI) and (4.3) (EII). For EI and $\beta^2 < 1$, the sector KOC is the appropriate domain; for $\beta^2 > 1$, it is sector ΩOA. For EII and $\beta^2 < 1$, the domain COB is appropriate; for $\beta^2 > 1$ it is sector AOB. Reference to Fig. 9 shows that the domain EII ($\beta^2 < 1$) coincides with the continuous spectral region $\{\omega_2^2\}$: if any discrete spectral values are embedded here, they are in the point-continuous spectrum, and constitute a previously unnoticed set of eigenvalues.

$z_2 = z_2(\omega^2)$. The set of frequencies $\omega^2 \in \{\omega_2^2(z)\,|\,0 \leqslant z \leqslant \infty\}$ constitutes the continuous spectrum: the set of improper eigenvalues of the operator $\rho^{-1}\mathbf{F}$ (see also Fig. 7).

It has been shown[49] that although systems of the type

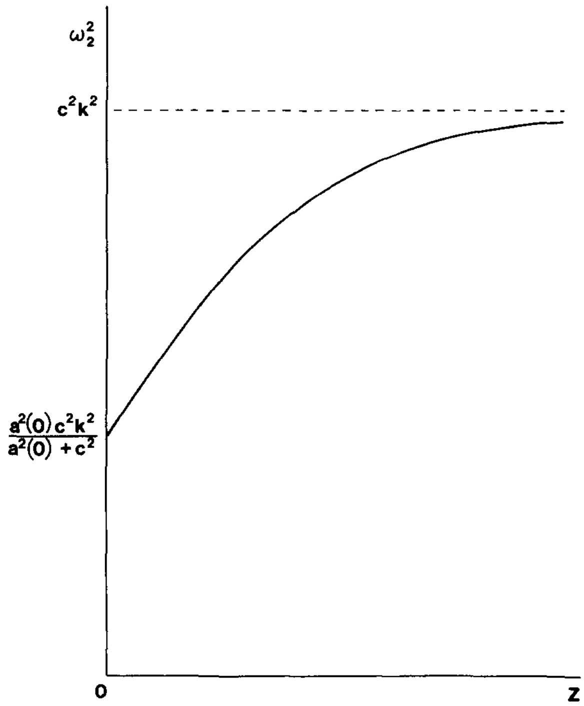


FIG. 6. Schematic behavior of the function $\omega_2^2(z) = a^2(z)c^2k^2 \times (a^2(z) + c^2)^{-1}$ for the exponential magnetoatmosphere (Secs. III and IV).

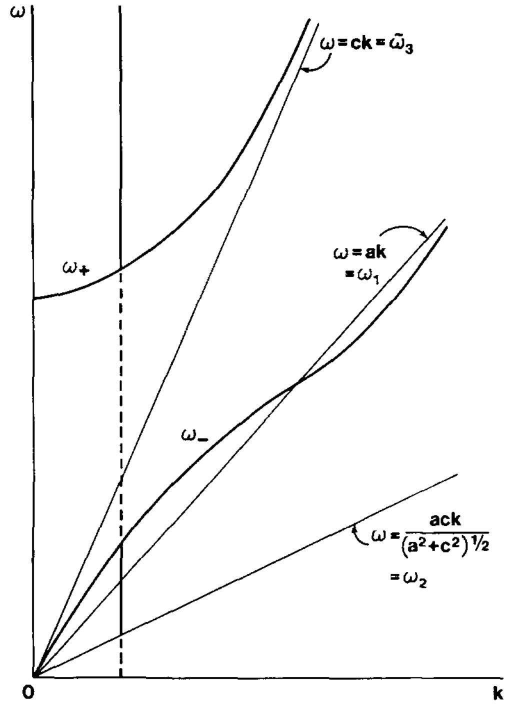


FIG. 7. The propagation curves for the constant parameter magnetoatmosphere (Sec. VI). This is similar to Fig. 3, but with the ω_1, ω_2, and $\tilde{\omega}_3$ lines present, illustrating, respectively, the relative positions of the Alfvén, "limiting" slow, and Lamb modes. The Alfvén mode is the only true mode: the other two are limits related to the accumulation points of the discrete spectra. The solid vertical line segments correspond to branch-line integrals in the related initial-value problem.[54]

(2.12)–(2.14) are highly nonlinear and hence non-Sturm–Liouville in the eigenvalue ω^2, partial monotonicity of the discrete spectrum can nevertheless be established outside the regions $\{\omega_i^2(z)\}$, $i = 1,2,3,4$. Specifically, if in Eq. (2.12) the expression for A is positive, then the discrete spectrum is *Sturmian* outside $\cup_i \{\omega_i^2(z)\}$, meaning that the node number (or number of oscillations) of the eigenfunction $\phi(z)$ increases with increasing eigenvalue ω^2. If, on the other hand, $A < 0$, the discrete spectrum is anti-Sturmian, so the node number increases as the eigenvalue decreases (this concept was discussed in terms of the constant parameter atmosphere in Sec. VI). Clearly, if the regions $\{\omega_i^2\}$ are disjoint (by virtue of weak inhomogeneity in the system, for example), then we may have the discrete spectrum changing from Sturmian to anti-Sturmian or vice versa every time one of the factors $[\omega^2 - \omega_i^2(z)]$ changes sign. In the case of the exponential atmosphere of Sec. III, A is defined by

$$A(z,\omega) = \rho_0(a^2 + c^2)(\omega^2 - \omega_2^2(z))/(\omega^2 - c^2k^2). \tag{7.11}$$

For the model at hand,

$$\sup_z \omega_2^2(z) = c^2k^2, \tag{7.12}$$

so $A > 0$ when

$$\text{(i) } \omega^2 < \inf_z \omega_2^2(z) = a^2(0)c^2k^2/[a^2(0) + c^2] \tag{7.13}$$

and when

$$\text{(ii) } \omega^2 > c^2k^2. \tag{7.14}$$

In these regions, the discrete spectrum, if it exists, is Sturmian in nature. There is no region, outside the continuum $\{\omega_2^2(z)\}$, for which the coefficient A is negative, by virtue of (7.12). In the case of a *finite* upper boundary, however,

$$\inf_z \omega_2^2(z) = \frac{a^2(0)c^2k^2}{a^2(0)+c^2} \leqslant \omega_2^2(z) \leqslant \sup_z \omega_2^2(z)$$
$$= \frac{a^2(0)c^2k^2}{a^2(0)+c^2\alpha} < c^2k^2, \qquad (7.15)$$

where $\alpha = e^{-d/H}$ (as in Sec. V) for an upper boundary at $z=d$. Then the spectrum, if it exists, is Sturmian in the regions denoted by (7.13) and (7.14), and anti-Sturmian when

$$a^2(0)c^2k^2/[a^2(0)+c^2\alpha] < \omega^2 < c^2k^2. \qquad (7.16)$$

Returning to the case of a semi-infinite medium ($\alpha=0$), considerable insight is afforded into the nature of the spectrum by examining a piecewise uniform atmosphere with the same basic features (e.g., increasing Alfvén speed with increasing values of z), To this end, we refer to Eqs. (6.4) in Sec. VI and Fig. 4 [see also Eqs. (6.6) and (6.7)]. First, consider the purely acoustic-gravity case for which $a^2 \equiv 0$. Under these circumstances it is easy to show that the upper branch $\omega_+^2(k_z) \to \infty$ as k_z increases, whereas $\omega_-^2(k_z) \to 0$ as k_z increases. Thus the gravity-modified acoustic spectrum accumulates at ∞ (Sturmian spectrum), whereas the compressibility-modified gravity spectrum accumulates at 0^+; the spectrum is anti-Sturmian[50] (see Fig. 4). As the Alfvén speed is increased from zero, the "line" of accumulation moves upward from $\omega^2=0$ to $\omega^2 \approx a^2c^2k^2/(a^2+c^2)$ as the magnetic term gradually dominates the gravitational term. (This can happen for sufficiently large values of k no matter how small a is.)

In Fig. 8 the ω^2-k_z diagram for the $\omega_\pm$ modes is presented schematically: for "weak" stratification the plus and minus modes may emanate, respectively, from above and below their values in the absence of gravity.

Figures 8(a)–8(c) show the progressive behavior of the modes (for a prescribed value of k) as the Alfvén speed increases. This is loosely equivalent to moving upward in the continuously varying exponential atmosphere. As a^2 increases, the line $\omega^2 = (\omega_2^2)_i$ moves progressively closer to $\omega^2 = \tilde{\omega}_3^2 = k^2c^2$. Thus, for the slow mode, the ω^2-k_z curve gradually flattens out and moves slowly upward until it collapses onto the line $\omega^2 = k^2c^2$ (the Lamb mode). It has already been noted that the discrete spectrum for the constant parameter atmosphere is anti-Sturmian for the slow mode, accumulating at $\omega^2 = \omega_2^2$. As a^2 increases, these lines, $\omega^2 = (\omega_2^2)_i$ accumulate from below to $\omega^2 = k^2c^2$, and the spectrum becomes infinitely degenerate in the limit $a^2 \to \infty$. Referring to the exponential model discussed in Secs. III and IV, it can be seen that as $z\to\infty$, $a^2\to\infty$; thus, for the slow mode, $\omega^2 \to \omega_3^{2-} = (k^2c^2)^-$. This is the result found in Theorem 4.3, namely that $w_0=\infty$ is a cluster point or accumulation point of the eigenvalues of the system: physically the slow mode accumulates to the Lamb mode eigenvalue ($\omega^2 = c^2k^2$) from below. The fast mode accumulates at $\omega^2=\infty$ as noted earlier: this corresponds to $w_0 \to (-\beta^{-2})^-$ (see also Fig. 1). Figure 9 illustrates the relationship between real values of ω^2 and w_0, again in a schematic fashion.

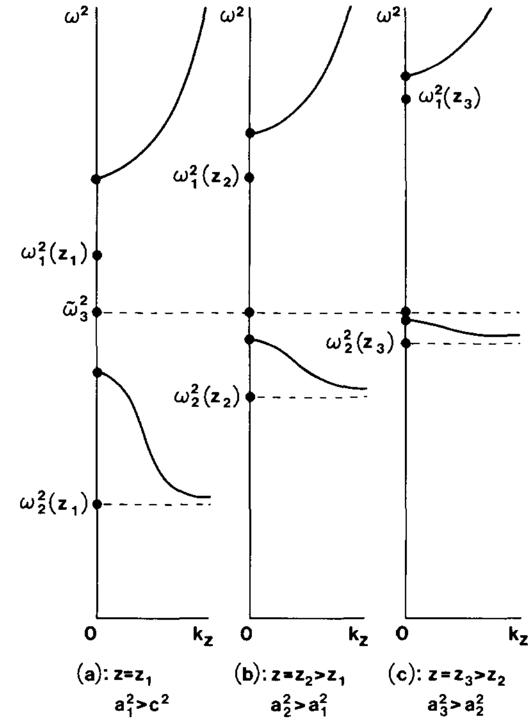


FIG. 8. (a), (b), (c) A schematic sequence of ω^2-k_z diagrams for a constant parameter medium (for simplicity, $g=0$ here, without much loss of generality). The sequency (a)–(c) may be regarded as mimicking the evolution of a local ω^2-$k_z(z)$ diagram as the Alfvén speed increases with altitude z. The diagrams are drawn for $a>c$ at all locations z; for $c>a$ any *given* diagram is schematically very similar, the only difference being that the positions of ω_1^2 and $\tilde{\omega}_3^2$ are interchanged. Note how the anti-Sturmian slow-mode is collapsing (from below!) onto the Lamb mode $\omega^2 = \tilde{\omega}_3^2$ as a increases.

It is clear that the restriction of S^2 to negative values for the analysis in Sec. III and parts of Sec. IV will place some

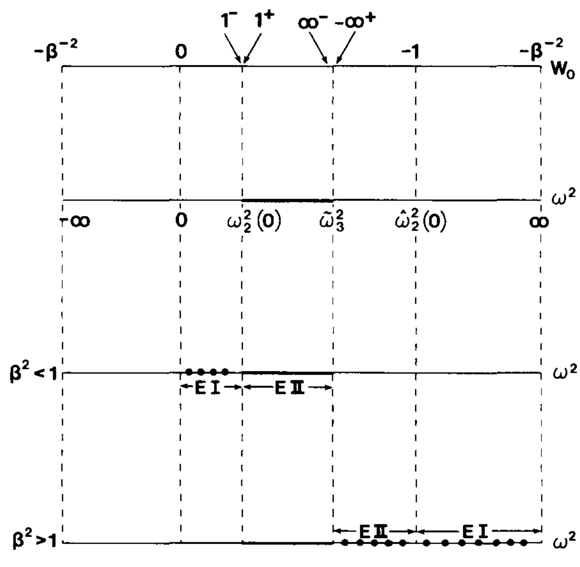


FIG. 9. A schematic representation of the relationship between w_0 and ω^2, as defined by the mapping (3.7b). The quantity $\hat{\omega}_2^2(0) = a^2(0)c^2k^2(a^2(0)-c^2)^{-1}$ exists only for $\beta^2>1$. Also illustrated are the subspectra (both discrete and continuous) which are permitted by the various dispersion relations EI (3.18) and EII (4.3).

restriction on w_0 also. To examine the implications of $R > \frac{1}{4}$ (i.e., $S^2 < 0$) for w_0, we have that, from (3.7a),

$$\omega^2 = a^2(0)c^2k^2w_0/[c^2 + a^2(0)w_0] \,. \tag{7.17}$$

Note that $\omega^2 < 0$ for $-c^2/a^2(0) < w_0 < 0$, corresponding to convective instability. On substituting for ω^2 in the expression (3.13) for R, the condition $R > \frac{1}{4}$ becomes, after some manipulation,

$$\beta^4\left[\frac{4(\gamma-1)}{\gamma^2} - 1\right]w_0^2 + \beta^2\left[\frac{8(\gamma-1)}{\gamma^2} - (1+4K^2)\right]w_0 + \frac{4(\gamma-1)}{\gamma^2} > 0 \,. \tag{7.18}$$

Denoting $1 + 4K^2$ by α, it follows that

$$\beta^2 w_0 > W_+ \quad \text{or} \quad \beta^2 w_0 < W_- \,, \tag{7.19}$$

where

$$W_\pm = \frac{-(\alpha\gamma^2 - 8\gamma + 8) \pm \{(\alpha\gamma^2 - 8\gamma + 8)^2 + 16(\gamma-1)(\gamma-2)^2\}^{1/2}}{2(\gamma-2)^2} \,. \tag{7.20}$$

Since

$$\beta^2 w_0 = \omega^2(c^2k^2 - \omega^2)^{-1} \,, \tag{7.21}$$

these constraints on w_0 correspond to

$$\omega^2 > \frac{c^2k^2W_+}{1+W_+} \quad \text{or} \quad \omega^2 < \frac{c^2k^2W_-}{1+W_-} \,. \tag{7.22}$$

For realistic magnetoatmospheres $1 < \gamma < 2$.

In particular, if $\alpha = 2$, i.e., $kH = 1/2$, and $\gamma = 5/3$ then $W_+ = 4$ and $W_- = -6$. Reference to Fig. 1 yields the following inequalities in ω^2:

(i) $0.8c^2k^2 < \omega^2 < c^2k^2$ (subsonic branch)

(ii) $c^2k^2 < \omega^2 < 1.2c^2k^2$ (supersonic branch)

APPENDIX: HYDRODYNAMIC SHEAR FLOW: A COMPARISON

Miles examined the shear flow given by[51]

$$U(z) = V(1 - e^{-z/H}), \quad 0 \leqslant z \leqslant \infty \,, \tag{A1}$$

with density profile

$$\ln(\rho(0)/\rho(z)) = \sigma(1 - e^{-z/H}), \quad 0 < \sigma \ll 1 \,, \tag{A2}$$

for an inviscid incompressible fluid bounded below by a rigid horizontal plane $z = 0$, filling the half-space $z > 0$. The governing Fourier-transformed ordinary differential equation and boundary conditions are (in terms of the perturbation stream function ϕ)

$$\phi'' + \left\{\frac{N^2}{(U-c)^2} - \frac{U''}{U-c} - k^2\right\}\phi = 0 \tag{A3}$$

and

$$(U-c)^{-1}\phi = 0 \quad (z = 0) \,, \tag{A4}$$

$$\phi = 0 \quad (z = \infty) \,, \tag{A5}$$

where $c = \omega/k$ is the (possibly complex) horizontal phase speed. As in Sec. III, a set of transformations of dependent and independent variables yields the hypergeometric differential equation. Thus if

$$\phi(z) = e^{-kz}f(w) \,, \tag{A6}$$

$$w = w_0 e^{-z/H} \,, \tag{A7}$$

where

$$w_0 = V/(V-c) \,, \tag{A8}$$

$$f = P\begin{Bmatrix} 0 & \infty & 1 & \\ 0 & K - \sqrt{1+K^2} & \frac{1}{2}(1+\nu) & w \\ -2K & K + \sqrt{1+K^2} & \frac{1}{2}(1-\nu) & \end{Bmatrix}, \tag{A9}$$

where $K = kH$, $\nu = (1 - 4Jw_0)^{1/2} \equiv i\mu$ if $1 < 4Jw_0$, and $J = \sigma g H V^{-2}$ is the Richardson number for the flow. The above boundary conditions yield the following eigenvalue equation:

$$(1 - w_0)^{1/2(-1+\nu)}\, {}_2F_1(a,b;1 + 2K;w_0) = 0 \,, \tag{A10}$$

where

$$a = \tfrac{1}{2}(1+\nu) + K - (1+K^2)^{1/2} \tag{A11}$$

and

$$b = \tfrac{1}{2}(1+\nu) + K + (1+K^2)^{1/2} \,. \tag{A12}$$

Of physical interest in a system such as this, governed by an equation of the type (A3), are those values of z (if any) for which $U(z) = c$—the so-called critical layer singularities. It is easily seen that this corresponds to the regular singular point $w = 1$. Using techniques similar to those used here, Miles establishes the following results, which we state in the form of a theorem.

Theorem A.1: For $Jw_0 < \frac{1}{4}$, ${}_2F_1(a,b;2K + 1;w_0)$ has

(i) one zero in $w_0 \in (0,1)$ iff $0 < \nu < \nu_0(K)$;

(ii) no zeros in $w_0 \in (1,\infty)$;

(iii) n zeros in $w_0 \in (-\infty,0)$ iff $\nu_n < \nu < \nu_{n+1}$,

where $\nu_0(K)$ and $\nu_n(K,n)$ are specified expressions. Miles established that there are no unstable modes for any K and J.

When $Jw_0 > \frac{1}{4}$, the relation (A10) can be written as

$$\mathrm{Re}\{A^*(1 - w_0)^{1/2(-1+i\mu)}\, {}_2F_1(a,b;1 + i\mu;1 - w_0)\} = 0 \,, \tag{A13}$$

where

$$A = \Gamma(1+2K)\Gamma(i\mu)/\Gamma(a)\Gamma(b) \,. \tag{A14}$$

On examining this as $w_0 \to 1^-$, we obtain the limiting eigenvalue relation

$$-(J - \tfrac{1}{4})^{1/2}\cot\{(J - \tfrac{1}{4})^{1/2}\ln(1 - w_0)\} = \tfrac{1}{2}\nu_0(K) \,, \tag{A15}$$

and hence it is proved, as in Sec. IV:

Theorem A.2: There exists an infinite number of zeros in $w_0 \in (0,1)$ for $Jw_0 > \frac{1}{4}$, and they accumulate at $w_0 = 1^-$

$(c=0^-)$. This result corresponds physically to the existence of an infinite number of stable, propagating internal gravity waves with speeds that lie outside the range of flow speeds $(0,V)$.

Recall that in the magnetoatmospheric problem discussed in Sec. VII, the point $w_0=\infty$ $(c^2=\hat{c}^{2-})$ was the corresponding point of accumulation for propagating waves. It is natural to ask, in view of the formal similarity of the singular differential equations (A3) and (3.3), what is the nature of the regular singularity $w_0=1$ in this case? Since the usual transformation formula for ${}_2F_1(a,b;c;w_0)$ has a pole when $c=a+b$, the appropriate formula, exhibiting a logarithmic singularity, is

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(n!)^2}\times[2\psi(n+1)-\psi(a+n)-\psi(b+n)-\ln|1-w_0|]\times(1-w_0)^n=0, \qquad \text{(A16)}$$

for $|\arg(1-w_0)|<\pi$, $|1-w_0|<1$, $R>\frac{1}{4}$, where $(a)_n=\Gamma(a+n)/\Gamma(a)$ is the usual Pochammer symbol, and $\psi(\alpha)$ is the digamma function.

A careful examination of expression (A16) precludes the existence of an accumulation point at $w_0=1^-$ [or $\omega^2=(\omega_2^2)^-$]. By contrast, for the hydrodynamic stability problem a standard Frobenius expansion in the neighborhood of the singularity z_c, where $c=U(z_c)$ $(w_0=1)$ yields complex roots of the indicial equation (in fact they are complex conjugates). There is no logarithmic singularity, and the behavior of the solution (and hence wave energy flux) is altogether different (see the detailed analysis of hydrodynamic critical layers by Booker and Bretherton, and the corresponding discussions of the MHD "critical layer").[24,32,52,53]

[1]C. P. Yu, Phys. Fluids **8**, 650 (1965).
[2]J. H. Thomas, Annu. Rev. Fluid Mech. **15**, 321 (1983).
[3]J. A. Adam and J. H. Thomas, Astrophys. Space Sci. **106**, 125 (1984).
[4]A. H. Nye and J. H. Thomas, Solar Phys. **38**, 399 (1974).
[5]A. H. Nye and J. H. Thomas, Astrophys. J. **204**, 573 (1976).
[6]B. Roberts, Solar Phys. **87**, 77 (1983).
[7]H. C. Spruit and B. Roberts, Nature (London) **304**, 401 (1983).
[8]J. M. Davila, Astrophys. J. **317**, 514 (1987).
[9]E. R. Priest, *Solar Magnetohydrodynamics* (Reidel, Dordrecht, 1982).
[10]J. P. Goedbloed, Phys. Fluids **18**, 1258 (1975).
[11]J. P. Goedbloed, "Lecture notes on ideal MHD," Rijnhuizen Report 83-145, Nieuwegein, 1983.
[12]J. Grossmann and W. C. Tataronis, Z. Phys. **261**, 217 (1973).
[13]M. Kruskal and M. Schwarzschild, Proc. R. Soc. London Ser. A **223**, 348 (1954).
[14]R. J. Tayler, J. Nucl. Fusion C 3, 266 (1961).
[15]L. M. B. C. Campos, Rev. Mod. Phys. **59**, 363 (1987).
[16]Yu. D. Zhugzhda and N. S. Dzhalilov, Geophys. Astrophys. Fluid Dynam. **35**, 131 (1986).
[17]J. A. Adam, Q. J. Mech. Appl. Math. **31**, 77 (1978).
[18]J. A. Adam. J. Phys. A: Math. Gen. **13**, 3325 (1980).
[19]J. A. Adam, Astrophys. Space Sci. **78**, 293 (1981).
[20]J. A. Adam, J. Plasma Phys. **19**, 77 (1978).
[21]P. S. Cally, Geophys. Astrophys. Fluid Dynam. **23**, 43 (1983).
[22]P. S. Cally and J. A. Adam, Geophys. Astrophys. Fluid Dynam. **23** 57 (1983).
[23]P. G. Drazin and W. H. Reid, *Hydrodynamic Stability* (Cambridge U.P., Cambridge, 1981).
[24]J. A. Adam, Phys. Rep. **142**, 263 (1986).
[25]K. M. Case, Phys. Fluids **3**, 143 (1960).
[26]K. M. Case, Phys. Fluids **3**, 149 (1960).
[27]E. Balbinski, Mon. Not. R. Astron. Soc. **209**, 145 (1984).
[28]J. A. Adam, Astrophys. Space Sci. **114**, 249 (1985).
[29]H. Ando, Geophys. Astrophys. Fluid Dynam. **35**, 71 (1986).
[30]J. W. Miles, J. Fluid Mech. **10**, 496 (1961).
[31]Z. Sedlacek, J. Plasma Phys. **5**, 239 (1971).
[32]J. A. Adam, Astrophys, Space Sci. **105**, 401 (1984).
[33]J. Boyd, J. Math. Phys. **22**, 1575 (1981).
[34]E. M. Barston, Ann. Phys. (NY) **29**, 282 (1964).
[35]L. N. Howard, J. Fluid Mech. **10**, 509 (1961).
[36]E. Kreyszig, *Introductory Functional Analysis with Applications* (Wiley, New York, 1978).
[37]N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Pitman, Boston, 1980).
[38]J. Chaudhuri and W. N. Everitt, Proc. R. Soc. Edinburgh A **68**, 95 (1968).
[39]I. B. Bernstein, E. A. Frieman, M. D. Kruskal, and R. M. Kulsrud, Proc. R. Soc. London Ser. A **244**, 17 (1958).
[40]J. P. Goedbloed, Physica **53**, 42 (1971).
[41]J. A. Adam, Sol. Phys. **52**, 293 (1977).
[42]J. Tataronis and W. C. Grossmann, Z. Phys. **261**, 203 (1973).
[43]H. Grad, Proc. Natl. Acad. Sci. (USA) **70**, 3277 (1973).
[44]E. R. Van Vleck, Trans. Am. Math. Soc. **3**, 110 (1902).
[45]J. A. Adam, unpublished report, Department of Mathematics and Statistics, Old Dominion University.
[46]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), Chap. 15.
[47]E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second-Order Ordinary Differential Equations* (Clarendon, Oxford, 1964), Vol. I.
[48]N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience, New York, 1958), Vol. II.
[49]J. P. Goedbloed and P. H. Sakanaka, Phys. Fluids **70**, 908 (1974).
[50]J. P. Goedbloed, Physica D **12**, 107 (1984).
[51]J. W. Miles, J. Fluid Mech. **28**, 305 (1967).
[52]J. R. Booker and F. P. Bretherton, J. Fluid Mech. **27**, 513 (1967).
[53]J. A. Adam, Phys. Rep. **86**, 217 (1982).
[54]J. A. Adam (to be published).

Spatially homogeneous cosmological models of Bianchi type III

Shri Ram
Department of Applied Mathematics, Institute of Technology, Banaras Hindu University, Varanasi 221005, India

(Received 12 August 1988; accepted for publication 19 October 1988)

Some analytic solutions to Einstein's equations with perfect fluids are presented that are of Bianchi type III. The geometrical and physical properties of the solutions are studied.

I. INTRODUCTION

In recent years there has been considerable interest in spatially homogeneous space-times. These space-times belong to Bianchi type I–IX classes and are interpreted as cosmological models. The field equations of general relativity for spatially homogeneous Bianchi type III have been investigated by many authors. Cahen and Defrise[1] obtained the general solution for the vacuum Bianchi type III model with nonzero cosmological constant. Lorentz[2] presented exact solutions of this class of metrics with dust and a cosmological constant and obtained solutions given by Cahen and Defrise,[1] Stewart and Ellis,[3] and Moussiaux *et al.*[4] in special cases. Further, Lorentz[5] obtained exact solutions of Einstein–Maxwell equations with dust and stiff matter. Recently, Bayin and Krisch[6] presented some analytic solutions of Bianchi type III with perfect fluids and studied the associated fluid parameters.

In Sec. II, we obtain three solutions to Einstein's field equations of Bianchi type III. The energy-momentum tensor is of perfect fluid type. The physical and kinematical properties of solutions are discussed. In Sec. III, following the technique of Hajj-Boutros,[7–9] we derive an algorithm for generating exact solutions of Bianchi type III with perfect fluids. In Sec. IV the solutions obtained in Sec. II are used to generate two more exact solutions. These new classes of solutions can be added to the rare solutions not satisfying the usual equation of state.

II. FIELD EQUATIONS AND SOLUTIONS

The metric of the Bianchi type III class of model is taken of the form[10]

$$ds^2 = -dt^2 + A^2\,dx^2 + B^2 e^{2x}\,dy^2 + C^2\,dz^2, \tag{1}$$

where $A(t)$, $B(t)$, and $C(t)$ are cosmic scale functions. For perfect fluids the field equations to be solved are

$$(A'B'/AB) + (B'C'/BC) + (A'C'/AC) - (1/A^2) = \rho, \tag{2}$$

$$(B''/B) + (C''/C) + (B'C'/BC) = -p, \tag{3}$$

$$(A''/A) + (C''/C) + (A'C'/AC) = -p, \tag{4}$$

$$(A''/A) + (B''/B) + (A'B'/AB) - (1/A^2) = -p, \tag{5}$$

$$(A'/A) - (B'/B) = 0, \tag{6}$$

where ρ and p are, respectively, the energy density of matter and pressure of the fluid. A prime denotes derivation with respect to time t.

From Eq. (6), we get

$$A = \mu B, \tag{7}$$

where μ is an integration constant. Without loss of any generality we take $\mu = 1$. Equations (4), (5), and (7) lead to

$$(A''/A) - (C''/C) + (A'/A)^{-2}(A'C'/AC) - (1/A^2) = 0, \tag{8}$$

which is a single equation in two unknowns. The method we shall use to generate solutions to (8) is to assume a solution for A and then solve for C. We choose

$$A = at + b, \tag{9}$$

where a and b are arbitrary constants. Inserting (9) into (8), we obtain

$$(at+b)^2 C'' + a(at+b)C' - (a^2-1)C = 0, \tag{10}$$

which is a Legendre differential equation. To obtain a solution of (10), we consider the following cases.

Case 1: When $|a| > 1$, the general solution of (10) is

$$C = c_1(at+b)^{\gamma} + c_2(at+b)^{-\gamma}, \tag{11}$$

where c_1 and c_2 are integration constants and

$$\gamma = (a^2-1)^{1/2}/a. \tag{12}$$

The metric of the solution is therefore

$$ds^2 = -dt^2 + (at+b)^2(dx^2 + e^{2x}\,dy^2) + \{c_1(at+b)^{\gamma} + c_2(at+b)^{-\gamma}\}^2\,dz^2. \tag{13}$$

The distribution of pressure p, density ρ, expansion Θ, and shear σ (see Refs. 11 and 12) for the metric (13) are given as follows:

$$p = \frac{a^2-1}{(at+b)^2}, \tag{14}$$

$$\rho = \frac{a^2-1}{(at+b)^2} + \frac{2a^2\gamma}{(at+b)^2} \times \left[\frac{c_1(at+b)^{\gamma} - c_2(at+b)^{-\gamma}}{c_1(at+b)^{\gamma} + c_2(at+b)^{-\gamma}}\right], \tag{15}$$

$$\Theta = \frac{2a}{at+b} + \frac{a\gamma}{(at+b)}\left[\frac{c_1(at+b)^{\gamma} - c_2(at+b)^{-\gamma}}{c_1(at+b)^{\gamma} + c_2(at+b)^{-\gamma}}\right], \tag{16}$$

$$\sigma = \frac{1}{\sqrt{3}}\left[\frac{a}{at+b} - \frac{a\gamma}{at+b} \times \left\{\frac{c_1(at+b)^{\gamma} - c_2(at+b)^{-\gamma}}{c_1(at+b)^{\gamma} + c_2 at+b)^{-\gamma}}\right\}\right]. \tag{17}$$

Case 2: When $|a| < 1$, Eq. (10) has the solution

$$C = c_3\cos\{\beta\log(at+b)\} + c_4\sin\{\beta\log(at+b)\}, \tag{18}$$

where c_3 and c_4 are integration constants and β is given by

$$\beta = (1-a^2)^{1/2}/a. \tag{19}$$

The metric of the solution is

0022-2488/89/030757-03$02.50

$$ds^2 - dt^2 + (at+b)^2\{dx^2 + e^{2x}\,dy^2\} + [c_3\cos\{\beta\log(at+b)\}c_4\sin\{\beta\log(at+b)\}]^2\,dz^2. \tag{20}$$

For this space-time the matter distribution and the kinematical parameters are as follows:

$$p = \frac{1-a^2}{(at+b)^2}, \tag{21}$$

$$\rho = \frac{a^2-1}{(at+b)^2} + \frac{2a^2\beta}{(at+b)^2}\left[\frac{-c_3\sin\{\beta\log(at+b)\} + c_4\cos\{\beta\log(at+b)\}}{c_3\cos\{\beta\log(at+b)\} + c_4\sin\{\beta\log(at+b)\}}\right], \tag{22}$$

$$\Theta = \frac{2a}{at+b} + \frac{a\beta}{at+b}\left[\frac{-c_3\sin\{\beta\log(at+b)\} + c_4\cos\{\beta\log(at+b)\}}{c_3\cos\{\beta\log(at+b)\} + c_4\sin\{\beta\log(at+b)\}}\right], \tag{23}$$

$$\sigma = \frac{1}{\sqrt{3}}\left[\frac{a}{at+b} - \frac{a\beta}{at+b}\,\frac{-c_3\sin\{\beta\log(at+b)\} + c_4\cos\{\beta\log(at+b)\}}{c_3\cos\{\beta\log(at+b)\} + c_4\sin\{\beta\log(at+b)\}}\right]. \tag{24}$$

Case 3: When $|a| = 1$, Eq. (10) has the solution

$$C = c_5 + c_6\log(at+b), \tag{25}$$

where c_5 and c_6 are arbitrary constants. The metric of the solution is

$$ds^2 = -dt^2 + (at+b)^2(dx^2 + e^{2x}\,dy^2) + \{c_5 + c_6\log(at+b)\}^2\,dz^2. \tag{26}$$

The distributions of matter and kinematical parameters for the metric (26) are given as follows:

$$p = 0, \tag{27}$$

$$\rho = 2c_6/(at+b)^2\{c_5 + c_6\log(at+b)\}, \tag{28}$$

$$\Theta = 2a/(at+b) + 2ac_6/(at+b)\{c_5 + c_6\log(at+b)\}, \tag{29}$$

$$\sigma = (1/\sqrt{3})[a/(at+b) - ac_6/(at+b)\{c_5 + c_6\log(at+b)\}]. \tag{30}$$

As far as we know, these models are new and can be added to the rare perfect fluid solutions not satisfying the equation of state:

$$p = (\epsilon - 1)\rho. \tag{31}$$

The models start expanding from a singular state at $t = -b/a$ and attain infinite volume as $t\to\infty$. In this limit p, ρ, Θ, and σ all vanish. We also find that the ratio (σ/Θ) tends to a finite limit as $t\to\infty$, which means that the shear scalar does not tend to zero faster than the expansion. These anisotropic models are acceleration- and rotation-free.

III. GENERATING TECHNIQUE

We now derive an algorithm for constructing spatially homogeneous perfect fluid solutions of Einstein's field equations that are of Bianchi type III and are locally rotationally symmetric. Let us recall Eq. (8). To treat this we introduce new functions R and S given by

$$R = A'/A, \quad S = C'/C. \tag{32}$$

Introducing (32) into (8), we obtain

$$R' + 2R^2 - RS - S' - S^2 - 1/A^2 = 0, \tag{33}$$

which can be regarded as a Riccati equation in R or S. Treating it as a Riccati equation in R, we linearize it by the change of function

$$R = R_0 + 1/Y, \tag{34}$$

where R_0 is the particular solution of (33). Inserting (34) into Eq. (33) we obtain the first-order linear differential equation

$$Y^1 + Y(S - 4R_0) = 2, \tag{35}$$

which has the solution

$$Y = (A_0^4 C)\left[\int \frac{2C}{A_0^4}\,dt + k_1\right]. \tag{36}$$

Equations (34) and (36) yield

$$A = A_0\exp\left[\int \frac{dt}{(A_0^4/C)\{\int(2C/A_0^4)dt + k_1\}} + k_2\right], \tag{37}$$

where k_1 and k_2 are integration constants. Thus from the couple $[A_0, C]$ we can generate new functions $[A, C]$, where C remains invariable.

If we treat Eq. (33) as a Riccati equation in s, a straightforward calculation provides

$$C = C_0\exp\left[\int \frac{dt}{(AC_0^2)\{\int dt/AC_0^2 + k_3\}} + k_4\right], \tag{38}$$

where k_3 and k_4 are integration constants. Thus from the couple $[A, C_0]$ we can obtain the new couple $[A, C]$, where A stays invariable.

IV. GENERATED SOLUTIONS

We apply generating formulas (37) and (38) to solutions obtained in Sec. II. Applying formula (38) to the metrics (13), (20), and (26) we arrive at the same metrics with different parameters. To apply formula (37) to metric (13), we take

$$A_0 = at + b, \quad C = c_1(at+b)^\gamma + c_2(at+b)^{-\gamma}. \tag{39}$$

A straightforward calculation yields

$$A = l\,[M(at+b)^{\gamma-1} + N(at+b)^{-\gamma-1} + Q(at+b)^2]^{1/2}, \tag{40}$$

where l is an arbitrary constant and M, N, and Q are given by

$$M = 2c_1/a(\gamma - 3), \quad N = -2c_2/a(\gamma+3), \quad Q = k_1.$$

The pressure and density are

$$p = [(\gamma^2 - 4\gamma + 3)M^2a^2(at+b)^{2\gamma-4} + (\gamma^2+4\gamma+3)N^2a^2(at+b)^{-2\gamma-4}$$
$$+ 2(\gamma^2 - 5\gamma + 6)MQa^2(at+b)^{\gamma-1} + 2(\gamma^2+5\gamma+6)NQa^2(at+b)^{-\gamma-1}$$
$$+ 6MNQ^2(\gamma^2+1)(at+b)^{-4}]/4[M(at+b)^{-\gamma-1} + N(at+b)^{-\gamma-1} + Q(at+b)^2]^2$$
$$+ \frac{a^2\gamma}{(at+b)^2}\left[\frac{c_1(\gamma-1)(at+b)^\gamma + c_2(\gamma+1)(at+b)^{-\gamma}}{c_1(at+b)^\gamma + c_2(at+b)^{-\gamma}}\right]$$
$$+ \frac{1}{2}\gamma a^2\left[\frac{M(\gamma-1)(at+b)^{\gamma-2} - N(\gamma+1)(at+b)^{-\gamma-2} + 2Q(at+b)}{M(at+b)^{\gamma-1} + N(at+b)^{-\gamma-1} + Q(at+b)^2}\right]$$
$$\times\left[\frac{c_1(at+b)^{\gamma-1} - c_2(at+b)^{-\gamma-1}}{c_1(at+b)^\gamma + c_2(at+b)^{-\gamma}}\right], \tag{41}$$
$$\rho = \frac{a}{4}\left[\frac{M(\gamma-1)(at+b)^{\gamma-2} - N(\gamma+1)(at+b)^{-\gamma-2} + 2Q(at+b)}{M(at+b)^{\gamma-1} + N(at+b)^{-\gamma-1} + Q(at+b)^2}\right]^2$$
$$+ a^2\gamma\left[\frac{M(\gamma-1)(at+b)^{\gamma-2} - N(\gamma+1)(at+b)^{-\gamma-2} + 2Q(at+b)}{M(at+b)^{\gamma-1} + N(at+b)^{-\gamma-1} + Q(at+b)^2}\right]$$
$$\times\left[\frac{c_1(at+b)^{\gamma-1} - c_2(at+b)^{-\gamma-1}}{c_1(at+b)^\gamma + c_2(at+b)^{-\gamma}}\right] - \frac{1}{M(at+b)^{\gamma-1} + N(at+b)^{-\gamma-1} + Q(at+b)^2}. \tag{42}$$

For this metric all of the fluids are acceleration- and rotation-free, but they do have expansion and shear. The kinematical properties are similar to that of the metric (13).

Let us apply the formula (37) to the metric (26). Evaluating

$$I_1 \int \frac{2C}{A_0^4}\,dt + k_1, \tag{43}$$

which is in formula Eq. (37), we obtain

$$I_1 = -(at+b)^{-3}\{P + Q\log(at+b)\} + k_1, \tag{44}$$

where

$$P = (6c_5 + 2c_6)/9, \quad Q = 2c_6/3$$

Further integration in Eq. (37) is rather difficult with nonzero k_1. Setting $k_1 = 0$ and $q = \log k_2$, we obtain

$$\int \frac{dt}{(A_0^4/C\{\int(2C/A_0^4)dt\}} + k_2$$
$$= \log q(at+b)^{-3/2}\{P + Q\log(at+b)\}^{1/2}. \tag{45}$$

Substitution of Eq. (45) into Eq. (37) yields

$$A = q(at+b)^{-1/2}\{P + Q\log(at+b)\}^{1/2}. \tag{46}$$

The metric of the solution is

$$ds^2 = -dt^2 + \left\{\frac{P + Q\log(at+b)}{at+b}\right\}(dx^2 + e^{2x}dy^2)$$
$$+ \{c_5 + c_6\log(at+b)\}^2 dz^2. \tag{47}$$

The distributions of matter and kinematical parameters for the metric Eq. (47) are

$$p = Q^2/4(at+b)^2\{P + Q\log(at+b)\}^2$$
$$+ 5Q/2(at+b)^2\{P + Q\log(at+b)\}$$
$$- 7/4(at+b)^2 + (at+b)/\{P + Q\log(at+b)\}, \tag{48}$$
$$\rho = Q^2/4(at+b)^2\{P + Q\log(at+b)\}^2$$
$$- Q/2(at+b)^2\{P + Q\log(at+b)\}$$
$$+ 2Qc_6/(at+b)^2\{P + Q\log(at+b)\}$$
$$\times\{c_5 + c_6\log(at+b)\}$$
$$- c_6/(at+b)^2\{c_5 + c_6\log(at+b)\}$$
$$- (at+b)/\{P + Q\log(at+b)\}, \tag{49}$$
$$\Theta = Qa/(at+b)\{P + Q\log(at+b)\}$$
$$+ ac_6/(at+b)\{P + Q\log(at+b)\} - a/(at+b), \tag{50}$$
$$\sigma = (1/\sqrt{3})[Qa/(at+b)\{P + Q\log(at+b)\}$$
$$- a/(at+b) - ac_6/(at+b)\{c_5 + c_6\log(at+b)\}]. \tag{51}$$

It represents a contracting perfect fluid model. The model is initially in the form of an infinite disk at the start of contraction. At $t \to \infty$, the spatial volume tends to zero and ρ and p assume infinite values. At this limit of zero volume, the expansion scalar and shear become vanishing small.

ACKNOWLEDGMENT

The author is extremely grateful to Professor K. P. Singh for valuable suggestions.

[1]M. Cahen and L. Defrise, Commun. Math. Phys. **11**, 56 (1968).
[2]D. Lorentz, J. Phys. A: Math. Gen. **15**, 2809 (1982).
[3]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[4]A. Moussiaux, P. Tombal, and J. Demret, J. Phys. A: Math. Gen. **14**, L277 (1982).
[5]D. Lorentz, J. Phys. A: Math. Gen. **16**, 575 (1983)
[6]S. S. Bayin and J. P. Krisch, J. Math. Phys. **27**, 262 (1986).
[7]J. Hajj-Boutros, *Gravitation, Geometry and Relativity Physics*, Lecture Notes in Physics (Springer, Berlin, 1984), Vol. 212, p.5.
[8]J. Hajj-Boutros, J. Math. Phys. **26**, 2297 (1985).
[9]J. Hajj-Boutros, J. Math. Phys. **27**, 1592 (1986).
[10]M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, NJ, 1975).
[11]G. F. R. Ellis, *Relativity and Cosmology* (Academic, New York, 1971), p. 104.
[12]C. P. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).