Simple orthogonal and unitary compact quantum systems and the İnönü–Wigner contraction

R. J. B. Fawcett and A. J. Bracken
Department of Mathematics, University of Queensland, Brisbane, Australia

(Received 30 September 1987; accepted for publication 9 March 1988)

Compact quantum dynamics, based on the Lie algebra so(n), is known to be relevant to the description of the internal motion of the electron and neutrino. It is shown here that a compact dynamics can also be based on u(n). In both cases there exists a contraction to a noncompact algebra containing a Weyl–Heisenberg subalgebra and consequently, there also exists an approximation scheme defined entirely by the contraction process, in which a sequence of compact systems described in terms of finite square matrices of increasing size, approaches a noncompact system of the usual type, described in terms of operators on a Hilbert space. Particular boson realizations of so(n) and u(n) are shown to be especially convenient for the discussion of the contractions. Some simple illustrative examples of compact systems are treated algebraically and numerically in order to show how their properties approach those of the corresponding noncompact systems.

I. INTRODUCTION

The notion of a "compact quantum system" based on the representation theory of the Lie algebra so(n), has been developed by Barut and Bracken.[1,2] The idea arose from their reappraisal[3,4] of the *Zitterbewegung* of Dirac's electron, which was shown to be describable as a compact quantum oscillator associated with the four-dimensional spinor representation of so(5).

The essence of the idea in general is to modify the commutation relations satisfied by the Hermitian coordinate operators Q_i and momentum operators P_i ($i = 1,...,n$) of a system with n degrees of freedom so that they generate the compact Lie algebra so($n+2$) rather than the Weyl–Heisenberg algebra w_n, which can be recovered from so($n+2$) by contraction. One has in place of the canonical commutation relations ($i,j,k,l = 1,...,n$),

$$\begin{aligned}
&[Q_i,Q_j] = (i\lambda^2/\hbar)S_{ij}, \quad [P_i,P_j] = (i\hbar/\lambda^2)S_{ij}, \\
&[Q_i,P_j] = i\hbar\delta_{ij}J, \quad [Q_i,S_{jk}] = i\hbar(\delta_{ik}Q_j - \delta_{ij}Q_k), \\
&[P_i,S_{jk}] = i\hbar(\delta_{ik}P_j - \delta_{ij}P_k), \qquad (1.1)\\
&[Q_i,J] = -(i\lambda^2/\hbar)P_i, \quad [P_i,J] = (i\hbar/\lambda^2)Q_i, \\
&[J,S_{ij}] = 0, \\
&[S_{ij},S_{kl}] = i\hbar(\delta_{ik}S_{jl} + \delta_{jl}S_{ik} - \delta_{jk}S_{il} - \delta_{il}S_{jk}).
\end{aligned}$$

The J_{ij} [$= -J_{ji} = (1/\hbar)S_{ij}$] span an so($n$) subalgebra, and one can see that the algebra (1.1) as a whole is so($n+2$) by setting

$$\begin{aligned}
&J_{in+1} = -J_{n+1i} = (1/\lambda)Q_i, \\
&J_{in+2} = -J_{n+2i} = (\lambda/\hbar)P_i, \\
&J_{n+1n+2} = -J_{n+2n+1} = J.
\end{aligned}$$

The positive constant λ appearing here has dimensions of length and is, initially at least, to be regarded as arbitrary. A compact quantum system is defined by specifying λ, together with a particular Hermitian finite-dimensional representation of so($n+2$), and a particular dynamics. Because the time evolution of states in the finite-dimensional state space is assumed to be governed in the usual way by Schrödinger's equation, the dynamics of a system is governed by a choice of Hamiltonian operator H. This should be some Hermitian operator on the space [for example, the representative of some element of the enveloping algebra of so($n+2$)].

There is a contraction[1,2] of so($n+2$) into the semidirect sum k_n of w_n and so(n), the latter being the group of rotations of the n-vector coordinate and momentum operators q_i, p_i which are obtained from the Q_i, P_i by contraction, and which generate w_n in the usual way with canonical commutation relations. This indicates that a compact quantum system can be regarded, in a sense defined by the contraction process, as an analog of a corresponding quantum system of the usual (noncompact) type, described in an infinite Hilbert space.

It is the main purpose of the present work to explore this relationship between noncompact quantum systems and their compact analogs through consideration of simple but hopefully typical examples. We are interested especially in the way in which the spectrum of a particular Hermitian operator for a noncompact system is approximated by the eigenvalue spectra of corresponding operators (matrices) in a sequence of compact analogs of increasing dimension as the contraction limit is approached. In order to consider such aspects, it is evidently not sufficient to consider the contraction at the level of the abstract Lie algebra: the contraction of suitably chosen sequences of representations of so($n+2$) into appropriate Hermitian representations of k_n must be considered.

Although the concept of contraction of Lie groups and algebras is well known,[5,6] most discussions at the level of representations have been limited to special cases.[7,8] Of particular relevance to our work here is the discussion by Bacry *et al.*[9] of the contraction of SO(3)$\times\mathbb{R}$ into the oscillator group Osc(1); although we are interested in a different contraction, of so(3) into w_1, and its generalizations, their technique using the Dyson–Vilenkin realization[10] of so(3) is ideal for our purposes. A natural generalization of this realization was given by Exner *et al.*[11] for the classical Lie algebras, in terms of conjugate coordinate and momentum oper-

 0022-2488/88/071521-08$02.50

ators, though they did not discuss contractions of these realizations. Their realization of $so(n+2)$, rewritten in terms of boson operators, is used in what follows.

It has been found that, apart from $so(n+2)$, one can use $u(n+1)$ to define another type of compact quantum system with n degrees of freedom. Accordingly, in what follows, we shall refer to orthogonal or unitary compact quantum systems as appropriate. We shall also use the realization of $u(n+1)$ given (essentially) by Exner *et al.*[11] when we discuss unitary systems and the associated contraction of $u(n+1)$ to $w_n \oplus u(n)$. The commutation relations in the unitary case are $(i,j,k,l = 1,...,n)$

$$
\begin{aligned}
&[Q_i,Q_j] = (i\lambda^2/\hbar)L_{ij}, \quad [P_i,P_j] = (i\hbar/\lambda^2)L_{ij}, \\
&[Q_i,P_j] = i\hbar\delta_{ij}M - iT_{ij}, \\
&[Q_i,M] = -(2i\lambda^2/\hbar)P_i, \quad [P_i,M] = (2i\hbar/\lambda^2)Q_i, \\
&[Q_i,L_{kl}] = i\hbar(\delta_{il}Q_k - \delta_{ik}Q_l), \\
&[P_i,L_{kl}] = i\hbar(\delta_{il}P_k - \delta_{ik}P_l), \\
&[Q_i,T_{kl}] = i\lambda^2(\delta_{ik}P_l + \delta_{il}P_k), \\
&[P_i,T_{kl}] = -(i\hbar^2/\lambda^2)(\delta_{ik}Q_l + \delta_{il}Q_k), \\
&[T_{ij},T_{kl}] = i\hbar(\delta_{ik}L_{jl} + \delta_{jl}L_{ik} + \delta_{jk}L_{il} + \delta_{il}L_{jk}), \\
&[L_{ij},L_{kl}] = i\hbar(\delta_{ik}L_{jl} + \delta_{jl}L_{ik} - \delta_{il}L_{jk} - \delta_{jk}L_{il}), \\
&[T_{ij},L_{kl}] = i\hbar(\delta_{il}T_{jk} + \delta_{jl}T_{ik} - \delta_{ik}T_{jl} - \delta_{jk}T_{il}).
\end{aligned}
\tag{1.2}
$$

The $K_{ij} = (1/\hbar)L_{ij} (= -K_{ji})$ span an $so(n)$ subalgebra of the $u(n)$ subalgebra spanned by the K_{ij} and $M_{ij} = (1/\hbar)T_{ij} (= M_{ji})$, and one can see that the algebra (1.2) as a whole is $u(n+1)$ by setting

$$
\begin{aligned}
&M_{i\,n+1} = M_{n+1\,i} = (1/\lambda)Q_i, \\
&K_{i\,n+1} = -K_{n+1\,i} = -(\lambda/\hbar)P_i,
\end{aligned}
$$

and

$$
M_{n+1\,n+1} = M.
$$

The boson realizations of $so(n+2)$ and $u(n+1)$ that we use[11] contain several parameters. One is a scaling parameter, the others determine and label the representations involved. For appropriate values of these labels, finite-dimensional representations are obtained. More precisely, in the infinite-dimensional Hilbert space on which act the boson operators in one of these realizations, the invariant subspace spanned by all vectors that can be obtained by the action of the basis operators on a vacuum vector, is finite dimensional and carries a representation of $so(n+2)$ or $u(n+1)$. In the contraction processes the dimension of this subspace increases and becomes infinite in the limit.

After discussing these contractions in what follows, we go on to look at some orthogonal and unitary compact quantum systems that are analogs of well-known noncompact systems. We shall look in particular at the behavior under contraction of the spectrum of the Hamiltonian operator, and of the distribution of probability over the eigenvalues of a coordinate operator when the system is in an energy eigenstate.

One motivation for studying compact quantum systems is the possibility that such systems may be involved in the internal structure of "elementary particles." This motivation also lies at the heart of other recent approaches to finite quantum systems.[12,13] Compact quantum systems of the orthogonal type are known, through their application to the *Zitterbewegung* of the electron[3,4] and neutrino,[14] to be relevant to this purpose in some cases. Whether or not other orthogonal or unitary systems are relevant remains unknown.

There is another motivation, however, for their study. In computer studies of noncompact quantum systems, one is forced to consider finite-dimensional approximations whose effects on the spectra of the noncommuting observables may be difficult to take into account. Replacing a noncompact system by a sequence of compact analogs in spaces of increasing dimension, related by the process of group contraction, may provide a suitable framework for computer studies in situations where it is desirable to be able to characterize exactly the nature of the approximation process. In this context, the contraction parameter gives a measure of the size of the finite-dimensional approximation, and hence of the fineness of the "discretization." It is difficult to give it a more physical interpretation. The concept of an approximation scheme based on a group (or Lie algebra) contraction may well be extended to operator equations other than Schrödinger's equation, in areas other than quantum mechanics.

It is intended to carry out in later work computer studies of compact analogs of classical systems that exhibit chaotic behavior, and to see how they behave as the contraction limit is approached. Before doing that, it is necessary to understand how elementary compact quantum systems behave, and that is the major objective of the present work, as we have already indicated.

II. CONTRACTION OF LIE GROUPS

Let G be a Lie group, given locally by its generators X_i and corresponding parameters a^i $(i = 1,...,n)$. The generators satisfy commutation relations of the form

$$
[X_i,X_j] = c_{ij}^k X_k,
$$

where the c_{ij}^k are the structure constants of G. Let the X_i and a^i be subjected to a linear transformation

$$
\begin{aligned}
&Y_i = U_i^j X_j, \quad a^i = b^j U_j^i, \\
&a^i X_i = b^j U_j^i X_i = b^j Y_j.
\end{aligned}
$$

New structure constants are defined by the commutation relations of the Y_i and these are easily related back to the old structure constants.[15] The İnönü–Wigner contraction chooses the matrix U to depend linearly on a positive parameter ϵ in such a way that one obtains

$$
\begin{aligned}
&Y_a = X_a + \epsilon \sum_{b=1}^{r} v_a^b X_b \quad (a = 1,...,r), \\
&Y_a = \epsilon X_a \quad (a = r+1,...,n), \\
&a^q = b^q + \epsilon \sum_{p=1}^{r} b^p v_p^q \quad (q = 1,...,r), \\
&a^q = \epsilon b^q \quad (q = r+1,...,n).
\end{aligned}
$$

In the contraction limit $\epsilon \to 0$, a new Lie group G' with new parameters b'^q and generators Y'_q is formed. The Lie group thus obtained is called the contraction of the group G. A generalization suggested by Wigner[5] involving the use of

higher orders of ϵ is taken up in this paper. The contraction used here is quadratic in ϵ. Alternatively, this can be regarded as a two stage linear contraction.[1,5,6] An example of a contraction, quadratic in ϵ, is given in Ref. 8 with regard to de Sitter groups.

A. Contraction of the relevant Lie algebras

The Lie algebra $\mathrm{so}(n+2)$ has basis operators $J_{AB}(=-J_{BA})$, $A,B=1,\ldots,n+2$, which satisfy the commutation relations

$$[J_{AB},J_{CD}]=i(\delta_{AC}J_{BD}+\delta_{BD}J_{AC}-\delta_{AD}J_{BC}-\delta_{BC}J_{AD})\,.$$

When this Lie algebra is contracted as (with $r,s=1,\ldots,n$),

$$\begin{aligned} &J^*_{rs}=J_{rs},\quad J^*_{r\,n+1}=\epsilon J_{r\,n+1}\,,\\ &J^*_{r\,n+2}=\epsilon J_{r\,n+2},\quad J^*_{n+1\,n+2}=\epsilon^2 J_{n+1\,n+2}\,, \end{aligned} \tag{2.1}$$

in the limit, the Lie algebra $w_n \dot{+} \mathrm{so}(n)$ results[1] with the operators $J^*_{r\,n+1}, J^*_{r\,n+2}, J^*_{n+1\,n+2}$ spanning the w_n component. The limiting commutation relations of this component are

$$\begin{aligned} &[J^*_{r\,n+1},J^*_{s\,n+2}]=i\delta_{rs}J^*_{n+1\,n+2}\,,\\ &[J^*_{r\,n+1},J^*_{s\,n+1}]=[J^*_{r\,n+2},J^*_{s\,n+2}]=0\,,\\ &[J^*_{r\,n+1},J^*_{n+1\,n+2}]=[J^*_{r\,n+2},J^*_{n+1\,n+2}]=0\,, \end{aligned} \tag{2.2}$$

so one can regard $J^*_{r\,n+1}, J^*_{r\,n+2}$ as being mapped onto canonically conjugate operators q_r, p_r, although not necessarily in that order.

The Lie algebra $\mathrm{u}(n+1)$ with generators $M_{PQ}(=M_{QP})$, $K_{PQ}(=-K_{QP})$, $P,Q=1,\ldots,n+1$, which satisfy the commutation relations

$$\begin{aligned} &[M_{PQ},M_{RS}]\\ &\quad=i(\delta_{QR}K_{PS}+\delta_{QS}K_{PR}+\delta_{PR}K_{QS}+\delta_{PS}K_{QR})\,,\\ &[K_{PQ},K_{RS}]\\ &\quad=i(\delta_{PR}K_{QS}+\delta_{QS}K_{PR}-\delta_{QR}K_{PS}-\delta_{PS}K_{QR})\,,\\ &[M_{PQ},K_{RS}]\\ &\quad=i(\delta_{PS}M_{QR}+\delta_{QS}M_{PR}-\delta_{QR}M_{PS}-\delta_{PR}M_{QS})\,, \end{aligned}$$

can be contracted in a similar way; with $i,j=1,\ldots,n$, set

$$\begin{aligned} &M^*_{ij}=M_{ij},\quad K^*_{ij}=K_{ij}\,,\\ &M^*_{i\,n+1}=\epsilon M_{i\,n+1},\quad K^*_{i\,n+1}=\epsilon K_{i\,n+1}\,,\\ &M^*_{n+1\,n+1}=\epsilon^2 M_{n+1\,n+1}\,. \end{aligned} \tag{2.3}$$

The result here, in the limit $\epsilon\to 0$, is the Lie algebra $w_n \dot{+} \mathrm{u}(n)$ with the operators $M^*_{r\,n+1}, K^*_{r\,n+1}, K^*_{n+1\,n+1}$, spanning the w_n subalgebra. Note that the way in which the parameter λ appears in (1.1) and (1.2), compared with the way ϵ appears in (2.1) and (2.3), precludes the identification of λ as the contraction parameter.

B. Contraction of representations

The solution of physical problems in quantum mechanics involves the calculation of eigenvalues of various operators such as the Hamiltonian. In the present context, we therefore require a knowledge of representations of the Lie algebras $\mathrm{so}(n+2)$ and $\mathrm{u}(n+1)$ associated with the compact quantum systems under discussion, and also a knowledge of how these representations may be contracted.

In general, there are two methods of contracting representations.[5] These are (1) the method of ϵ-dependent similarity transformations; and (2) the method of the sequence of representations. Only the second method is of interest here.

In this second method, a sequence of representations of the uncontracted Lie algebra is chosen, dependent on some parameter l [such as the representation label in the sequence of symmetric representations of $\mathrm{so}(n+2)$] which, in the cases we consider, determines the finite dimension of each representation in the sequence. The parameter is then varied according to some power of ϵ. The relationship $\epsilon=1/\sqrt{l}$ has been chosen for the boson realizations presented in this paper. As $\epsilon\to 0$, the dimension of the corresponding representations approaches infinity.

The representation of the contracted Lie algebra obtained in this way may be reducible even though each representation of the contracting sequence may be irreducible, and we may need to project onto an irreducible component. The use of the boson realizations[11] as follows avoids this difficulty; projection onto the desired irreducible component is automatic. However, when we turn in Sec. IV C to numerical calculation using matrix realizations, the difficulty will appear.

III. THE REALIZATIONS AND THEIR CONTRACTIONS

Realizations for a number of classical Lie algebras [$\mathrm{gl}(n+1)$, $\mathrm{o}(2n+1)$, $\mathrm{o}(2n)$, $\mathrm{sp}(2n)$] have been given by Exner *et al.*[11] These are used in a slightly modified form for the purposes of this paper. As has been shown,[11] these realizations have a recurrent structure, above and beyond the normal chain of inclusion

$$\begin{aligned} &\mathrm{so}(n+1)>\mathrm{so}(n)>\cdots>\mathrm{so}(2)\,,\\ &\mathrm{u}(n+1)>\mathrm{u}(n)>\cdots>\mathrm{u}(1)\,, \end{aligned}$$

that allows the construction of a representation with any desired labeling.

A. The so(n+ 2) realization

Let $\bar{a}_r, a_r$, $r=1,\ldots,n$, be n boson pairs satisfying the commutation relations

$$[a_r,\bar{a}_s]=\delta_{rs},\quad [a_r,a_s]=0,\quad [\bar{a}_r,\bar{a}_s]=0\,. \tag{3.1}$$

These boson pairs (creation and annihilation operators) act on a Hilbert space of vectors constructed from a vacuum vector $|0\rangle$ for which $a_i|0\rangle=0$. The vectors

$$\bar{a}_1^k\bar{a}_2^l\ldots\bar{a}_n^m|0\rangle\quad(k,l,\ldots,m=0,1,2,\ldots)$$

form a basis for the Hilbert space. At this point it is convenient to introduce the number operators

$$\begin{aligned} &N_i=\bar{a}_i a_i\quad(\text{no } i \text{ sum},\ i=1,\ldots,n)\,,\\ &N=N_1+\cdots+N_n=\bar{a}_s a_s\quad(\text{summation convention})\,. \end{aligned}$$

A realization of the Lie algebra $\mathrm{so}(n+2)$ can be formulated as follows. Let

$$
\begin{aligned}
J_{rs} &= -J_{sr} = -i(\bar{a}_r a_s - \bar{a}_s a_r) + S_{rs}\,, \\
J_{rn+1} &= -J_{n+1r} \\
&= \tfrac{1}{2}((1/k)a_r + k(-2\bar{a}_r(N-l) + \bar{a}_s\bar{a}_s a_r)) \\
&\quad + ik\bar{a}_s S_{sr}\,, \\
J_{rn+2} &= -J_{n+2r} \\
&= (i/2)((1/k)a_r - k(-2\bar{a}_r(N-l) + \bar{a}_s\bar{a}_s a_r)) \\
&\quad + k\bar{a}_s S_{sr}\,, \\
J_{n+1n+2} &= -J_{n+2n+1} = \bar{a}_s a_s - l = N - l \\
&\qquad (r,s = 1,\ldots,n)\,.
\end{aligned} \tag{3.2}
$$

The variable k is an arbitrary real/complex scaling parameter and the $S_{rs}(=-S_{sr})$, with commutation relations

$$
\begin{aligned}
&[a_i, S_{rs}] = [\bar{a}_i, S_{rs}] = 0\,, \\
&[S_{ij}, S_{rs}] = i(\delta_{ir}S_{js} + \delta_{js}S_{ir} - \delta_{is}S_{jr} - \delta_{jr}S_{is}) \\
&(i,j,r,s = 1,\ldots,n)\,,
\end{aligned} \tag{3.3}
$$

span a representation of so(n) in a space distinct from the Hilbert space on which the creation and annihilation operators act, and are used to establish the recurrence property. The S_{rs} could be replaced by a matrix representation of so(n), or by a boson representation generated by a new set of $(n-2)$ creation and annihilation operators, which would in turn have provision for the inclusion of a further distinct representation of so($n-2$), and so on. The Casimir operator G for this representation of so($n+2$) is

$$
\begin{aligned}
G &= \tfrac{1}{2}J_{rs}J_{rs} + J_{rn+1}J_{rn+1} + J_{rn+2}J_{rn+2} + J^2_{n+1n+2} \\
&= l(l+n)I + \tfrac{1}{2}S_{rs}S_{rs}\,,
\end{aligned} \tag{3.4}
$$

where I is the identity operator.

Suppose that the representation of so(n) spanned by the S_{rs} is the irreducible one labeled[15] $(l_1,l_2,\ldots,l_m)$, where $m = \frac{1}{2}n$ if n is even, and $m = \frac{1}{2}(n-1)$ if n is odd, and the l's are all integral or all half-odd-integral, with $l_1 \geqslant l_2 \geqslant \cdots \geqslant |l_m| \geqslant 0$. (If n is odd, the modulus signs can be dropped.) Suppose also that $l - |l_1|$ is a non-negative integer. The operators J_{rs}, J_{rn+1}, J_{rn+2}, and J_{n+1n+2} span an irreducible representation of so($n+2$) labeled $(l,l_1,\ldots,l_m)$, in the space obtained by the action of these operators on the vacuum vector $|0\rangle$. If the S_{rs} vanish, we have the symmetric representation $(l,0,\ldots,0)$ of so($n+2$).

B. Contraction of the representation

The contraction of this representation is made via the substitution $\epsilon = (1/\sqrt{l})$ in Eq. (2.1). A change of variable is also required. Let

$$
b_r = (1/\sqrt{2l})a_r, \quad \bar{b}_r = \sqrt{2l}\,\bar{a}_r \quad (r = 1,\ldots,n)\,. \tag{3.5}
$$

Then in the limit of the contraction, as $l \to \infty$, $\epsilon \to 0$,

$$
\begin{aligned}
J^*_{rs} &= -i(\bar{b}_r b_s - \bar{b}_s b_r) + S_{rs}\,, \\
J^*_{rn+1} &= (1/\sqrt{2})((1/k)b_r + k\bar{b}_r)\,, \\
J^*_{rn+2} &= (i/\sqrt{2})((1/k)b_r - k\bar{b}_r)\,, \\
J^*_{n+1n+2} &= -1\,.
\end{aligned} \tag{3.6}
$$

The operators J^*_{rn+1}, J^*_{rn+2}, J^*_{n+1n+2} obviously form the basis for a Hermitian representation of w_n, of the type used in ordinary quantum mechanics.

C. The u(n+1) realization

A similar recurrent realization is easily developed for the Lie algebra u($n+1$). Using the same set of creation and annihilation operators as before, with $i,j,s = 1,\ldots,n$, let

$$
\begin{aligned}
M_{ij} &= M_{ji} = (\bar{a}_i a_j + \bar{a}_j a_i) + R_{ij}\,, \\
M_{in+1} &= M_{n+1i} = -(1/k)a_i + k\bar{a}_i(N-l) \\
&\qquad - (k/2i)\bar{a}_s(S_{is} - iR_{is})\,, \\
M_{n+1n+1} &= -2(\bar{a}_s a_s - l) = -2(N-l)\,, \\
K_{ij} &= -K_{ji} = -i(\bar{a}_i a_j - \bar{a}_j a_i) + S_{ij}\,, \\
K_{in+1} &= -K_{n+1i} = -i((1/k)a_i + k\bar{a}_i(N-l)) \\
&\qquad + (k/2)\bar{a}_s(S_{is} - iR_{is})\,.
\end{aligned} \tag{3.7}
$$

Again the k is an arbitrary real/complex scaling parameter and the $R_{ij}(=R_{ji})$ and $S_{ij}(=-S_{ji})$ appearing have a similar function to the S_{ij} in the so($n+2$) formulas. Here though, the R_{ij} and the S_{ij}, which have the commutation relations $(i,k,l,r,s = 1,\ldots,n)$,

$$
\begin{aligned}
&[a_i, R_{rs}] = [\bar{a}_i, R_{rs}] = [a_i, S_{rs}] = [\bar{a}_i, S_{rs}] = 0\,, \\
&[R_{kl}, R_{rs}] = i(\delta_{lr}S_{ks} + \delta_{ls}S_{kr} + \delta_{kr}S_{ls} + \delta_{ks}S_{lr})\,, \\
&[S_{kl}, S_{rs}] = i(\delta_{kr}S_{ls} + \delta_{ls}S_{kr} - \delta_{lr}S_{ks} - \delta_{ks}S_{lr})\,, \\
&[R_{kl}, S_{rs}] = i(\delta_{ks}R_{lr} + \delta_{ls}R_{kr} - \delta_{lr}R_{ks} - \delta_{kr}R_{ls})\,,
\end{aligned} \tag{3.8}
$$

span a representation of u(n), which may in turn contain a further distinct representation of u($n-1$), and so on.

Suppose this representation of u(n) is irreducible, labeled[15] $(l_1,l_2,\ldots,l_n)$ by n non-negative integers satisfying $l_1 \geqslant l_2 \geqslant \cdots \geqslant l_n$, and suppose l is a non-negative integer with $l \geqslant l_1$. Then the representation $(l,l_1,\ldots,l_n)$ of u($n+1$) is obtained.

D. Contraction of the representation

The proposed contraction of this representation is similar to that of the representation of so($n+2$) with the same substitution $\epsilon = 1/\sqrt{l}$ in Eq. (2.3). A change of variable is again necessary;

$$
b_s = (1/\sqrt{l})a_s, \quad \bar{b}_s = \sqrt{l}\,\bar{a}_s \quad (s = 1,\ldots,n)\,. \tag{3.9}
$$

The result of the contraction is as follows:

$$
\begin{aligned}
M^*_{ij} &= \bar{b}_i b_j + \bar{b}_j b_i + R_{ij}\,, \\
M^*_{in+1} &= -((1/k)b_i + k\bar{b}_i)\,, \\
M^*_{n+1n+1} &= 2\,, \\
K^*_{ij} &= -i(\bar{b}_i b_j - \bar{b}_j b_i) + S_{ij}\,, \\
K^*_{in+1} &= -i((1/k)b_i - k\bar{b}_i)\,.
\end{aligned} \tag{3.10}
$$

E. The so(3) and u(2) realizations

A realization of su(2) can be derived from the realization of u(2). This su(2) realization is shown to be isomorphic to the so(3) realization. Let

$$
X_1 = \tfrac{1}{4}(M_{11} - M_{22}), \quad X_2 = -\tfrac{1}{2}M_{12}, \quad X_3 = -\tfrac{1}{2}K_{12}\,.
$$

Then the X_i are su(2) generators satisfying

$$[X_i,X_j] = i\epsilon_{ijk}X_k ,$$

and

$$X_1 = J_{23}, \quad X_2 = J_{31}, \quad X_3 = J_{12} ,$$

with J_{23}, J_{31}, J_{12} being the so(3) generators, if the following substitutions are made:

$$a_{u(2)} = -ia_{so(3)}, \quad \bar{a}_{u(2)} = i\bar{a}_{so(3)} ,$$
$$R_{u(2)} = 0, \quad l_{u(2)} = 2l_{so(3)} .$$

IV. APPLICATIONS AND EXAMPLES

A number of simple compact quantum systems are examined here in detail. These include the one-dimensional harmonic oscillator (HO) and the one-dimensional finite square well problem. Each compact quantum system is modeled using an appropriate Lie algebra (and a sequence of representations) in such a way that, under the process of contraction, the analogous noncompact quantum system is described in the limit. The principle aim is to calculate the eigenvalues and eigenvectors of the Hamiltonian operator (from which the entire dynamics can be derived), and to examine the effect of group contraction on these. However, only in a few special cases is it possible to calculate these eigenvalues and eigenvectors analytically; the numerical method used in the square well problem provides a general framework for one-dimensional problems in which the position coordinate values are not constrained in any way. We expect also that, in addition to the energy spectrum, all other physical properties of a noncompact system will be approximated as the contraction limit is approached. To support this, we calculate, for the compact oscillator and square well problems in one dimension, the distribution of probability over the eigenvalues of the coordinate operator when the system is in a low-lying energy eigenstate. The resulting histograms are compared with the well-known probability density functions (in position space) for the appropriate bound states of the corresponding noncompact systems.

A. The free particle

The kinematical Lie algebra for the description in ordinary quantum mechanics of a free particle of mass m is w_1 with a pair of conjugate position and momentum operators q and p, satisfying the usual commutation relations, and a Hamiltonian operator $H = p^2/2m$.

Let a be an arbitrary fixed length scale, and let J_1, J_2, J_3 be the generators of a $(2l+1)$-dimensional irreducible matrix representation of the Lie algebra so(3), in which J_3 is diagonal. Denote by $\phi_1,\phi_2,...,\phi_{2l+1}$, the eigenvectors of J_3 corresponding to eigenvalues $-l, -l+1,...,l$, and take $\phi_1 = (1,0,...,0)$, $\phi_2 = (0,1,...,0)$, etc. A compact quantization of the problem is obtained by putting

$$Q = (a/\sqrt{l})J_1, \quad P = -(\hbar/a\sqrt{l})J_2 .$$

The contraction of so(3) to w_1 is obtained by considering a sequence of representations with $l \to \infty$. We see that

$$[Q,P] = (i\hbar/l)\,\mathrm{diag}(l,l-1,...,-l) \to i\hbar\,\mathrm{diag}(1,1,...) ,$$

provided we project in the limit onto the subspace spanned by the vectors $\phi_1,\phi_2,...$. It becomes immediately obvious that if the matrix elements and eigenvectors were labeled in a different manner, the result $[Q,P] \to -i\hbar I$ could be obtained. In fact, the representation of w_1 obtained in the limit is clearly reducible, and we have to bear that fact in mind when considering the limits approached by the spectra of the operators. In the matrix representation of so(3) for given l, the eigenvalues of J_2 are $-l,...,l$. Now let

$$H = \frac{P^2}{2m} = \frac{\hbar^2}{ma^2}\frac{1}{2l}J_2^2 .$$

Consequently the eigenvalues of H are

$$\frac{\hbar^2}{ma^2}\left[\frac{(-l)^2}{2l}, \frac{(-l+1)^2}{2l}, ..., \frac{(l-1)^2}{2l}, \frac{l^2}{2l}\right],$$

which in the case of half-integral l are

$$\frac{\hbar^2}{2ma^2}\left[\frac{1}{4l}, \frac{1}{4l}, \frac{9}{4l}, \frac{9}{4l}, \frac{25}{4l}, \frac{25}{4l}, ..., l, l\right] .$$

These approach the quasicontinuum $[0,\infty)$, in the limit $l \to \infty$, as expected. It can be seen that each point in the limiting spectrum is doubly degenerate. (In the case of integral l, the zero energy eigenvalue is the only nondegenerate eigenvalue.) This clearly reflects the fact that the representation of w_1 obtained in the limit is reducible.

B. The harmonic oscillator

The general isotropic n-dimensional HO has n conjugate position/momentum pairs q_i, p_i satisfying the standard commutation relations. If m is the mass of the oscillator, and ω the common angular frequency of oscillation, then

$$H = \tfrac{1}{2}((1/m)p_ip_i + m\omega^2 q_iq_i) \quad \text{(summ. conv.)} .$$

A compact analog of this system can be described in terms of $u(n+1)$ by replacing q_i and p_i by

$$Q_i = \sqrt{(\hbar/2lm\omega)}\,M_{i\,n+1}, \quad P_i = -\sqrt{(\hbar m\omega/2l)}\,K_{i\,n+1} ,$$

respectively, with $M_{i\,n+1}, K_{i\,n+1}$ as in (3.7) for example. The problem can also be described in terms of $so(n+2)$ by the substitutions

$$Q_i = \sqrt{(\hbar/lm\omega)}\,J_{i\,n+1}, \quad P_i = -\sqrt{(\hbar m\omega/l)}\,J_{i\,n+2} ,$$

with $J_{i\,n+1}, J_{i\,n+2}$ as in (3.2), but the calculation of the eigenvectors of H is then rather complicated for $n > 2$. This is due to the presence in the $so(n+2)$ operators of terms in $\bar{a}_s\bar{a}_s$, not present in the $u(n+1)$ case. There is a natural length scale $a = \sqrt{\hbar/m\omega}$ in both cases. Using a symmetric realization of $u(n+1)$ (one in which $R_{ij} = S_{ij} = 0$),

$$H = \tfrac{1}{2}((1/m)P_iP_i + m\omega^2 Q_iQ_i) \quad \text{(summ. conv.)}$$
$$= (\hbar\omega/4l)(M_{i\,n+1}M_{i\,n+1} + K_{i\,n+1}K_{i\,n+1})$$
$$= (\hbar\omega/4l)(-2)[(N+n)(N-l) + N(N-l-1)] ,$$

where N is the previously mentioned number operator. In the Hilbert space generated from the vacuum vector $|0\rangle$, the energy eigenvectors are of the form

$$\bar{a}_1^{m_1} \cdots \bar{a}_n^{m_n}|0\rangle, \quad m_1,...,m_n = 0,1,...,$$

where $M = m_1 + \cdots + m_n$, the total number of quanta, is never larger than l in this finite-dimensional space. The energy eigenvalues are

$$E_{m_1 \cdots m_n}$$
$$= -(\hbar\omega/2l)[(M+n)(M-l)+M(M-l-1)]$$
$$\to \hbar\omega(M+(n/2)) \quad \text{as } l\to\infty .$$

As remarked above, the realizations for so(3) and su(2) [derived from u(2)] are isomorphic. This obviously implies that the results for any problem described in terms of so(3) or u(2) will be the same. However, the results from the so(4) and u(3) descriptions of the two-dimensional oscillator are quite different. The u(3) results may be deduced directly from the general case above. The energy eigenvalues are

$$E_{n_1 n_2} = -(\hbar\omega/2l)((n_1+n_2+2)(n_1+n_2-l) + (n_1+n_2)(n_1+n_2-1-l))$$

and the eigenvectors are just $\bar{a}_1^{n_1}\bar{a}_2^{n_2}|0\rangle$. The so(4) description yields energy eigenvalues

$$E_{n_1 n_2} = (\hbar\omega/2l)(2l(n_1+n_2+1) - 2m(n_1-n_2) - 2(n_1^2+n_2^2)),$$

where m in this equation is the spin term S_{12}. The energy eigenvectors are $(\bar{a}_1 + i\bar{a}_2)^{n_1}(\bar{a}_1 - i\bar{a}_2)^{n_2}|0\rangle$. The energy eigenvalue expressions in the u(3) and so(4) approaches give the same result in the limit $l\to\infty$, viz., $E_{n_1 n_2} = \hbar\omega(n_1+n_2+1)$, whatever the value of m.

The u$(n+1)$ description for the n-dimensional isotropic HO can be generalized to the anisotropic oscillator. The Hamiltonian operator will now be of the form $H = H_1 + \cdots + H_n$, where

$$H_i = \tfrac{1}{2}((1/m_i)P_iP_i + m_i\omega_i^2 Q_iQ_i) \quad (\text{no } i \text{ sum})$$

and

$$Q_i = \sqrt{(\hbar/2lm_i\omega_i)}M_{in+1}, \quad P_i = -\sqrt{(\hbar m_i\omega_i/2l)}K_{in+1} .$$

Using the number operators N_i,

$$H_i = -(\hbar\omega_i/2l)[(N_i+n)(N-l)+N_i(N-l-1)],$$

with eigenvalues

$$-(\hbar\omega_i/2l)[(m_i+n)(M-l)+m_i(M-l-1)];$$

the vectors $\bar{a}_1^{m_1}\cdots\bar{a}_n^{m_n}|0\rangle$, $m_1,\ldots,m_n = 0,1,2,\ldots$, $M = m_1+\cdots+m_n \leqslant l$, are common eigenvectors. In the limit $l\to\infty$,

$$E_{m_1\cdots m_n} = \hbar\omega_1(m_1+\tfrac{1}{2})+\cdots+\hbar\omega_n(m_n+\tfrac{1}{2}) .$$

C. A numerical example: The finite square well

As remarked earlier, very few problems can be solved analytically in these compact descriptions. Therefore, in most cases, it will be necessary to resort to numerical computation on matrix representations. The problem below shows how a particular class of problems can be treated. The only restriction on the potential function $V(q)$ is that it be defined for all values of the position coordinate q.

Consider a particle of mass M with Hamiltonian operator

$$H = \frac{p^2}{2M} + V(q), \text{ where } V(q) = \begin{cases} 0, & \text{for } |q| > a/2, \\ -V_0, & \text{for } |q| < a/2 . \end{cases}$$

As in the case of the free particle, an irreducible $(2l+1)$-dimensional matrix representation of so(3) (again with J_3 diagonal) is used, with the further requirement that the matrices J_1, J_2, J_3 be Hermitian. The Casimir operator is

$$\mathbf{J}^2 = J_1^2 + J_2^2 + J_3^2 = l(l+1)I_{2l+1} ,$$

where I is the identity matrix. There is an orthonormal basis of column vectors $|lm\rangle$ in which

$$\mathbf{J}^2|lm\rangle = l(l+1)|lm\rangle, \quad J_3|lm\rangle = m|lm\rangle .$$

The ith element of $|lm\rangle$ is $\delta_{i,l+m+1}$ (using the Kronecker delta). The unitary matrix

$$S = e^{i(\pi/2)J_1}e^{i\pi J_3}e^{i(\pi/2)J_2} = e^{-i(2\pi/3)(J_1+J_2+J_3)/\sqrt{3}}$$

rotates the labels of the generators; i.e.,

$$S^\dagger J_1 S = J_3, \quad S^\dagger J_2 S = J_1, \quad S^\dagger J_3 S = J_2 .$$

1. so(3) quantization

Let

$$Q = (a/\sqrt{l})J_1, \quad P = -(\hbar/a\sqrt{l})J_2 .$$

Thus

$$H = \frac{\hbar^2}{Ma^2}\frac{1}{2l}J_2^2 + V\left(\frac{a}{\sqrt{l}}J_1\right).$$

Now $V(Q) = V((a/\sqrt{l})J_1)$ is most readily calculated in a basis in which Q is diagonal, for then

$$V(Q) = \text{diag}(V(q_{-l}),\ldots,V(q_l)),$$

where $q_{-l},\ldots,q_l$ are the eigenvalues of Q. Since Q is not diagonal in the basis $|lm\rangle$, we observe that

$$S^\dagger QS = (a/\sqrt{l})J_3, \quad Q(S|lm\rangle) = (am/\sqrt{l})(S|lm\rangle),$$

hence

$$V(Q) = SV(aJ_3/\sqrt{l})S^\dagger = S\,\text{diag}(V(q_{-l}),\ldots,V(q_l))S^\dagger,$$

where

$$q_m = \frac{m}{\sqrt{l}} \text{ and } V(q_m) = \begin{cases} 0, & \text{if } |m/\sqrt{l}| > \tfrac{1}{2}, \\ -V_0, & \text{if } |m/\sqrt{l}| \leqslant \tfrac{1}{2}. \end{cases}$$

Introducing the dimensionless parameter

$$k = V_0 Ma^2/\hbar^2,$$

the problem can be rewritten as

$$S^\dagger HS = \frac{\hbar^2}{Ma^2}\left[\frac{1}{2l}J_1^2 + f\left(\frac{1}{\sqrt{l}}J_3\right)\right],$$

where f is diagonal with eigenvalues $f_m = (k/V_0)V(q_m)$. The eigenvalues and eigenvectors of this transformed Hamiltonian are readily calculated numerically.

2. Position probabilities

If ϕ_E is a normalized eigenvector of H corresponding to energy eigenvalue E, then let ϕ_E be expanded in terms of the basis vectors $|lm\rangle$;

$$\phi_E = \sum_{m=-l}^{l} c_m|lm\rangle, \quad \text{where} \quad \sum_{m=-l}^{l} c_m^* c_m = 1 .$$

Alternatively, ϕ_E can be expanded in terms of the orthonormal position eigenvectors $S|lm\rangle$;

$$\phi_E = \sum_{m=-l}^{l} C_m S|lm\rangle, \quad \text{where} \quad \sum_{m=-l}^{l} C_m^* C_m = 1 .$$

For the energy eigenstate ϕ_E, $C_m^* C_m$ is the probability that the position coordinate (which takes on values from the set of eigenvalues of Q) of the system is $am/\sqrt{I}$.

The eigenvectors of $S^\dagger HS$ are then

$$S^\dagger \phi_E = \sum_{m=-l}^{l} C_m |lm\rangle ,$$

corresponding to the same energy eigenvalues E, and the coefficients C_m can be read off directly.

3. Numerical calculations

If spinor representations $(\frac{1}{2})$, $(\frac{3}{2})$, $(\frac{5}{2})$,..., of so(3) are considered, then an even potential $V(q)$ [i.e., $V(q) = V(-q)$ for all q] will always lead to a double degeneracy of all the eigenvalues of the Hamiltonian matrix H. In what follows, these eigenvalues will only be listed once, and scaled in units of $\hbar^2/Ma^2$.

For the finite square well (of depth parameter $k = 10$) lying between $-\frac{1}{2} \leqslant (q/a) \leqslant \frac{1}{2}$ in terms of the dimensionless position coordinate q/a, the first few scaled dimensionless energy eigenvalues are shown (to four decimal places) in Table I.

The ordinary (noncompact) quantum finite square well (with depth parameter $k = 10$) has only two bound states with energy values -7.7050 and -1.8628 (calculated numerically by solving Schrödinger's time-independent equation in one dimension), and there is a continuum of unbound energy states with non-negative energy.

Figures 1 and 2 show histograms of probability distribution over coordinate eigenvalues corresponding to the bound energy eigenstates in a compact approximation ($d = 2l + 1 = 100$, $k = 10$). The bars of the histogram represent the probabilities associated with the (scaled) position eigenvalues q_m/a of Q/a which run from $q_{-l}/a = -\sqrt{I} \approx -7.0356$ to $q_l/a = \sqrt{I} \approx 7.0356$. For comparison, the dotted lines show the probability density functions $\rho(x) = \psi(x)^* \psi(x)$ for the bound states of the noncompact problem found by solving Schrödinger's equation. The histograms (calculated so that the sum of the heights is 1) have been sketched rescaled so that the area under each histogram is the same as the area under the corresponding curve. The probability density functions are given by

TABLE I. Energy eigenvalues for the finite square well in four compact approximations.

$d = 2l + 1$	1st eig.	2nd eig.	3rd eig.	4th eig.	5th eig.
100	−8.2400	−3.5907	0.0110	0.0429	0.0991
200	−7.8324	−2.2367	0.0053	0.0216	0.0480
400	−7.7412	−1.9639	0.0026	0.0105	0.0236
800	−7.7385	−1.9527	0.0012	0.0051	0.0116

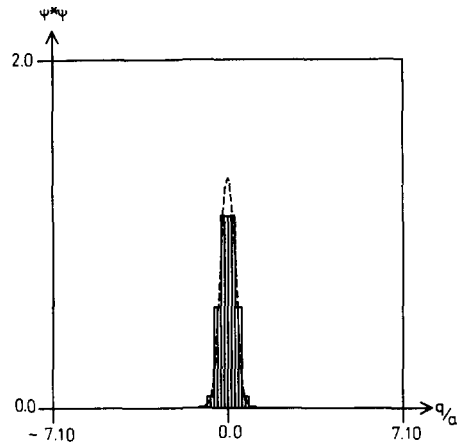


FIG. 1. Distribution of probability over position values for the first energy eigenstate of the compact finite square well problem.

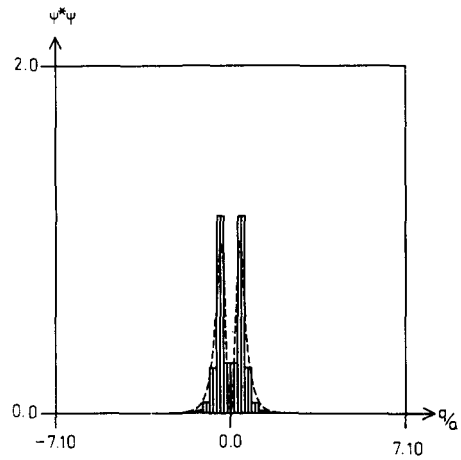


FIG. 2. Distribution of probability for the second energy eigenstate, square well problem.

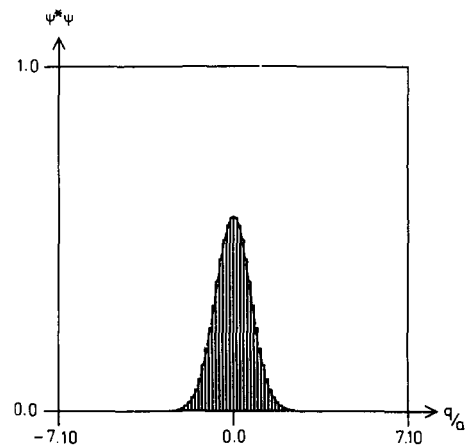


FIG. 3. Distribution of probability over position values for the first energy eigenstate of the compact simple harmonic oscillator problem.

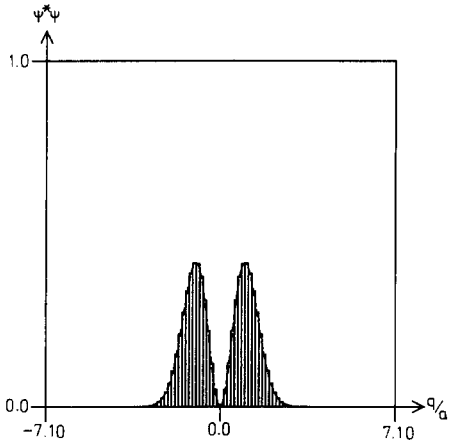


FIG. 4. Distribution of probability for the second energy eigenstate, oscillator problem.

$$\rho(x) = Ae^{2\sqrt{-2\epsilon}\,x}, \quad x = q/a < -\tfrac{1}{2},$$

$$\rho(x) = B\begin{bmatrix}\sin^2\\ \cos^2\end{bmatrix}\sqrt{2(\epsilon+k)}\,x, \quad -\frac{1}{2} \leqslant x \leqslant \frac{1}{2},$$

$$\rho(x) = Ae^{-2\sqrt{-2\epsilon}\,x}, \quad x > \tfrac{1}{2}.$$

The values for the coefficients A, B and the energy eigenvalue ϵ are determined in each case in the usual way by continuity of $\rho(x)$ and $d\rho/dx$ at $x = \pm\frac{1}{2}$, and the normalization condition $\int_{-\infty}^{\infty} \rho(x)dx = 1$. The "$\cos^2$" term is chosen for the first bound state and the "$\sin^2$" term for the second. Each bar of the histogram is $1/\sqrt{l}$ wide (the distance between position eigenvalues) so the probability density functions are plotted from $x = -(l+\frac{1}{2})/\sqrt{l} \approx -7.10$ to $x = (l+\frac{1}{2})/\sqrt{l} \approx 7.10$.

Figures 3 and 4 show a similar treatment for the first two eigenvalues 0.5 and 1.4898 of the simple HO in a compact approximation ($d = 2l + 1 = 100$). The dotted lines represent the normalized probability density functions ρ for the energy eigenvalues 0.5 and 1.5, which are given by

$$\rho_{1/2}(x) = (1/\sqrt{\pi})e^{-x^2} \text{ and } \rho_{3/2}(x) = (2/\sqrt{\pi})x^2e^{-x^2}.$$

Again, the histograms have been scaled to give equal areas with the density functions.

ACKNOWLEDGMENTS

The authors wish to thank Mr. D. D. McGregor for advice concerning the numerical treatment of the finite square well problem.

[1] A. O. Barut and A. J. Bracken, J. Math. Phys. **26**, 2515 (1985).
[2] A. O. Barut, in *Conformal Groups and Related Symmetries, Lecture Notes in Physics*, Vol. 261, edited by A. O. Barut and H. D. Doebner (Springer, New York, 1986).
[3] A. O. Barut and A. J. Bracken, Phys. Rev. D **23**, 2454 (1981).
[4] A. O. Barut and A. J. Bracken, in *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 135, edited by K. B. Wolf (Springer, New York, 1980), p. 206.
[5] E. İnönü and E. P. Wigner, Proc. Natl. Acad. Sci. **40**, 119 (1954); E. İnönü, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964); E. J. Saletan, J. Math. Phys. **2**, 1 (1960).
[6] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974); A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific, Warsaw, 1977).
[7] J. Voisin, J. Math. Phys. **8**, 611 (1967); C. W. Conatser, *ibid.* **13**, 196 (1972); R. H. Brennich, Rep. Math. Phys. **8**, 153 (1975); P. L. Huddleston, J. Math. Phys. **19**, 1645 (1978); D. Arnal and J-C. Cortet, *ibid.* **20**, 556 (1979); Y. Alhassid, F. Gürsey, and F. Iachello, Ann. Phys. (NY) **167**, 181 (1986).
[8] J. Mickelsson and J. Niederle, Commun. Math. Phys. **27**, 167 (1972).
[9] H. Bacry, A. Grossman, and J. Zak, in *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 50, edited by A. Janner, T. Janssen, and M. Boon (Springer, New York, 1976), p. 249.
[10] F. J. Dyson, Phys. Rev. **102**, 1217 (1956); N. J. Vilenkin, *Special Functions and The Theory of Group Representations*, translated by V. N. Singh (Am. Math. Soc., Providence, RI, 1968).
[11] P. Exner, M. Havlicek, and W. Lassner, Czech. J. Phys. B **26**, 1213 (1976).
[12] A. H. Kamupingene, T. D. Palev, and S. P. Tsaneva, J. Math. Phys. **27**, 2067 (1986).
[13] S. Gudder and V. Naroditsky, Int. J. Theor. Phys. **20**, 619 (1981); R. Jagannathan and T. S. Santhanam, *ibid.* **21**, 351 (1982); P. Stovicek and J. Tolar, Rep. Math. Phys. **20**, 157 (1984).
[14] A. O. Barut, A. J. Bracken, and W. D. Thacker, Lett. Math. Phys. **8**, 172 (1984).
[15] Barut and Razcka, Ref. 6, pp. 286 and 278.

Multivariable continuous Hahn polynomials

M. V. Tratnik
CNLS and T7, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 26 October 1987; accepted for publication 16 March 1988)

A multivariable generalization of the continuous Hahn polynomials is presented; it is a $(4p+4)$-parameter family, where p is the number of variables. It is shown that they are orthogonal with respect to subspaces of equal degree and biorthogonal within a given subspace. In the simplest case the multivariable weight function takes the form $\operatorname{sech}[\pi(x_1+x_2+\cdots+x_p)]\operatorname{sech}(\pi x_1)\operatorname{sech}(\pi x_2)\cdots\operatorname{sech}(\pi x_p)$.

I. INTRODUCTION

The classical orthogonal polynomials of a discrete variable occupy an important place among the special functions of mathematical physics. These are the difference analogs of the Jacobi, Laguerre, and Hermite polynomials and include the Hahn, Meixner, Krawtchouk, and Charlier[1] polynomials among others. They appear in a number of problems of theoretical and mathematical physics, group representation theory, and computational physics, as well as other fields.

Quite recently (1985) Atakishiyev and Suslov[2] constructed continuous analogs of the Hahn and Meixner polynomials by analytically continuing the discrete variable and parameter into the complex plane. The previously unknown continuous Hahn polynomials were shown to be real and orthogonal with respect to a real weight function. It was also demonstrated that they are closely related to unitary irreducible representations of the Lorentz group SO(3,1). Soon after Askey[3] presented a complex generalization of these polynomials and their weight function. More recently Bender, Mead, and Pinsky[4] have considered a specific class of the continuous Hahn polynomials and derived a number of interesting properties including a close connection with the Heisenberg algebra; they also note several additional applications.

In this paper we present a multivariable generalization of the complex continuous Hahn polynomials. These are a $(4p+4)$-parameter family, where p denotes the number of variables. It is first shown that these polynomials are orthogonal with respect to subspaces of equal degree. That is, the inner product of any two polynomials not of the same degree vanishes. Then we demonstrate that within a given subspace of equal degree these polynomials are biorthogonal. We also discuss some specific examples where the weight function takes a relatively simple form.

We begin with the multivariable conformal polynomials

$$C^{\mu_1\cdots\mu_{p+1}}_{n_1n_2\cdots n_p}(x_1,x_2,\ldots,x_p) = \sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\frac{\Gamma(n_k+\mu_k)}{\Gamma(n_k+\mu_k-j_k)}x_k^{n_k-j_k}\right] \times\frac{\Gamma(N+\mu_{p+1})}{\Gamma(J+\mu_{p+1})}(X-1)^J, \tag{1.1}$$

and the dual conformal polynomials

$$D^{\mu_1\cdots\mu_{p+1}}_{n_1n_2\cdots n_p}(x_1,x_2,\ldots,x_p) = (2N+\mu-1)\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\frac{x_k^{n_k-j_k}}{n_k!\,\Gamma(n_k+\mu_k-j_k)}\right] \times\frac{\Gamma(2N+\mu-J-1)}{\Gamma(N+\mu_{p+1})}(-1)^J, \tag{1.2}$$

which were introduced by Lam and Tratnik[5] in the context of SU(1,1) group representations. These are polynomials in the p variables $x_1,x_2,\ldots,x_p$ over the p-simplex $x_k>0$, $k=1,2,\ldots,p$, $x_1+x_2+\cdots+x_p<1$, which are associated with the following weight function:

$$w^{\mu_1\cdots\mu_{p+1}}(x_1,x_2,\ldots,x_p) = \left[\prod_{k=1}^{p}x_k^{\mu_k-1}\right](1-X)^{\mu_{p+1}-1}. \tag{1.3}$$

The $p+1$ parameters $\mu_1,\mu_2,\ldots,\mu_{p+1}$ identify a particular family of polynomials and their weight function whereas the set of p non-negative integers $n_1,n_2,\ldots,n_p$ label the members of a given family. The $\{j_k\}$ denotes summation indices $j_1,j_2,\ldots,j_p$, which run over non-negative integers from $j_k=0$ to $j_k=n_k$, $k=1,2,\ldots,p$, the shorthand notation

$$X\equiv\sum_{k=1}^{p}x_k,\quad N\equiv\sum_{k=1}^{p}n_k,\quad J\equiv\sum_{k=1}^{p}j_k,\quad \mu\equiv\sum_{k=1}^{p+1}\mu_k \tag{1.4}$$

is being employed, and the degree of a polynomial is simply given by N. The conformal and dual conformal polynomials are each orthogonal with respect to subspaces of equal degree, that is,

$$\int dx_1\cdots\int dx_p\; C^{\mu}_{n}(x)C^{\mu}_{n'}(x)w^{\mu}(x)=0,\ \text{if } N\neq N', \tag{1.5}$$

and

$$\int dx_1\cdots\int dx_p\; D^{\mu}_{n}(x)D^{\mu}_{n'}(x)w^{\mu}(x)=0,\ \text{if } N\neq N', \tag{1.6}$$

where we are using a condensed notation and the integration is over the p-simplex, while in general these two families are biorthogonal:

$$\int dx_1\cdots\int dx_p\; C^{\mu}_{n}(x)D^{\mu}_{n'}(x)w^{\mu}(x)=\prod_{k=1}^{p}\delta_{n_kn'_k}. \tag{1.7}$$

In a manner analogous to the single variable case[2,3] the conformal and dual conformal polynomials are discretized

 0022-2488/88/071529-06$02.50

and then analytically continued to complex values of the discrete variables and parameters. In this manner we obtain the multivariable continuous Hahn polynomials in the p variables $x_1, x_2, \ldots, x_p$,

$$P^{a,b,c,d}_{n_1 n_2 \cdots n_p}(x_1,x_2,\ldots,x_p) = i^N \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} \frac{\Gamma(n_k+a_k+b_k)}{n_k!\,\Gamma(j_k+a_k+b_k)} \times \frac{\Gamma(j_k+a_k+ix_k)}{\Gamma(a_k+ix_k)} \right] \frac{\Gamma(N+A+d)}{\Gamma(J+A+d)} \times \frac{\Gamma(N+J+A+B+c+d-1)}{\Gamma(N+A+B+c+d-1)} (-1)^J, \quad (1.8)$$

and their biorthogonal counterparts

$$Q^{a,b,c,d}_{n_1 n_2 \cdots n_p}(x_1,x_2,\ldots,x_p) = (-i)^N \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} \frac{\Gamma(n_k+a_k+b_k)}{n_k!\,\Gamma(j_k+a_k+b_k)} \times \frac{\Gamma(j_k+b_k-ix_k)}{\Gamma(b_k-ix_k)} \right] \times \frac{\Gamma(N+c+d)}{\Gamma(N-J+c+d)} \frac{\Gamma(N-J+c+iX)}{\Gamma(c+iX)} (-1)^J, \quad (1.9)$$

as well as the weight function

$$w^{a,b,c,d}(x_1,x_2,\ldots,x_p) = \left[\prod_{k=1}^{p} \Gamma(a_k+ix_k)\Gamma(b_k-ix_k) \right] \times \Gamma(c+iX)\Gamma(d-iX). \quad (1.10)$$

The $(2p+2)$ complex parameters $a_k, b_k, c, d, k = 1,2,\ldots,p$, label the different families and we have defined

$$A \equiv \sum_{k=1}^{p} a_k, \quad B \equiv \sum_{k=1}^{p} b_k. \quad (1.11)$$

As usual i denotes the square root of minus one and the remaining notation is as we discussed previously. This multivariable extension of the Hahn polynomials is nontrivial in that the polynomials and weight function do not factor with respect to the independent variables. When no confusion arises we simply write $P_n(x)$, $Q_n(x)$, and $w(x)$ for the polynomials and weight function, respectively.

This $P_n(x)$ is of degree n_k in the variable x_k, $Q_n(x)$ is of degree N in x_k, and both are of total degree N. In the next section we demonstrate that each of these families is orthogonal with respect to the degree N, analogous to (1.5) and (1.6).

II. ORTHOGONALITY PROPERTIES

We will have frequent use of the following multiple integral, where n_k and m_k are non-negative integers:

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \left[\prod_{k=1}^{p} \Gamma(n_k+a_k+ix_k) \times \Gamma(m_k+b_k-ix_k) \right] \Gamma(c+iX)\Gamma(d-iX). \quad (2.1)$$

The p integration contours are simply the p real axes if $\mathrm{Re}(a_k), \mathrm{Re}(b_k), \mathrm{Re}(c), \mathrm{Re}(d) > 0$, $k = 1,2,\ldots,p$. In the more general case the contours are deformed to separate the increasing sequences of poles of the integrand from the decreasing sequences, which is possible whenever the increasing sequences do not have any poles in common with the decreasing sequences. In the remainder of this manuscript the contours will simply be written as the real axes but should be interpreted as above if necessary.

The above integral can be calculated by repeated use of the single integral formula[6]

$$\int_{-\infty}^{\infty} dx\, \Gamma(\alpha+ix)\Gamma(\beta+ix)\Gamma(\delta-ix)\Gamma(\gamma-ix) = (2\pi) \frac{\Gamma(\alpha+\delta)\Gamma(\alpha+\gamma)\Gamma(\beta+\delta)\Gamma(\beta+\gamma)}{\Gamma(\alpha+\beta+\delta+\gamma)}, \quad (2.2)$$

and in this manner we obtain

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \left[\prod_{k=1}^{p} \Gamma(n_k+a_k+ix_k)\Gamma(m_k+b_k-ix_k) \right] \Gamma(c+iX)\Gamma(d-iX) = (2\pi)^p \frac{\Gamma(c+d)\Gamma(N+A+d)\Gamma(M+B+c)\prod_{k=1}^{p}\Gamma(n_k+m_k+a_k+b_k)}{\Gamma(N+M+A+B+c+d)}, \quad (2.3)$$

which can be verified by induction on p.

Now we discuss some properties of the polynomials $P_n(x)$; in particular, we consider the following inner product:

$$I \equiv \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p\, P_n(x) w(x) \times \prod_{k=1}^{p} \frac{\Gamma(m_k+b_k-ix_k)}{\Gamma(b_k-ix_k)}, \quad (2.4)$$

where again the m_k are non-negative integers ($M \equiv \sum_{k=1}^{p} m_k$). We demonstrate below that if $N > M$ then this integral vanishes.

Substituting expressions (1.8) and (1.10) into (2.4) and then using the integral formula (2.3) we obtain

$$I \equiv i^N (2\pi)^p \frac{\Gamma(c+d)\Gamma(M+B+c)\Gamma(N+A+d)}{\Gamma(N+A+B+c+d-1)} \times \left[\prod_{k=1}^{p} \frac{\Gamma(n_k+a_k+b_k)}{n_k!} \right] \times \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} \frac{\Gamma(m_k+j_k+a_k+b_k)}{\Gamma(j_k+a_k+b_k)} \right] \times \frac{\Gamma(N+J+A+B+c+d-1)}{\Gamma(M+J+A+B+c+d)} (-1)^J. \quad (2.5)$$

We ignore the multiplicative constant outside the summation and denote the remainder as I'. Then consider first the special case $N = M + 1$, in which case

$$I' = \sum_{\{j_k\}} \prod_{k=1}^{p} \binom{n_k}{j_k} \frac{\Gamma(m_k + j_k + a_k + b_k)}{\Gamma(j_k + a_k + b_k)} (-1)^{j_k}. \quad (2.6)$$

Introducing a set of real variables z_k, $k = 1,2,...,p$, this summation can be written as follows:

$$\begin{aligned} I' &= \sum_{\{j_k\}} \prod_{k=1}^{p} \binom{n_k}{j_k} \left(\frac{\partial}{\partial z_k}\right)^{m_k}_{z_k=1} z_k^{m_k + j_k + a_k + b_k - 1} (-1)^{j_k} \\ &= \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial z_k}\right)^{m_k}_{z_k=1} z_k^{m_k + a_k + b_k - 1}\right] \\ &\quad \times \sum_{\{j_k\}} \prod_{k=1}^{p} \binom{n_k}{j_k} (-z_k)^{j_k} \\ &= \prod_{k=1}^{p} \left(\frac{\partial}{\partial z_k}\right)^{m_k}_{z_k=1} z_k^{m_k + a_k + b_k - 1} (1 - z_k)^{n_k}. \end{aligned} \quad (2.7)$$

If $N = M + 1$, as we supposed, then $n_k > m_k$ for at least one of the k values. In this case at least one factor of $(1 - z_k)$ will survive the differentiation, and then will vanish upon setting $z_k = 1$.

Now consider the general case $N > M$, which is treated in a similar manner. First we write

$$\frac{\Gamma(N + J + A + B + c + d - 1)}{\Gamma(M + J + A + B + c + d)} = \sum_{l=0}^{N-M-1} \alpha_l J^l, \quad (2.8)$$

where the α_l are some constants independent of J. Then from Eqs. (2.5) and (2.8) we have

$$\begin{aligned} I' = \sum_{\{j_k\}} \sum_{l=0}^{N-M-1} \alpha_l J^l \prod_{k=1}^{p} \binom{n_k}{j_k} \\ \times \frac{\Gamma(m_k + j_k + a_k + b_k)}{\Gamma(j_k + a_k + b_k)} (-1)^{j_k}, \end{aligned} \quad (2.9)$$

and if we again introduce the variables z_k, $k = 1,2,...,p$,

$$\begin{aligned} I' = \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial z_k}\right)^{m_k}_{z_k=1} z_k^{m_k + a_k + b_k - 1}\right] \\ \times \sum_{\{j_k\}} \sum_{l=0}^{N-M-1} \alpha_l J^l \prod_{k=1}^{p} \binom{n_k}{j_k} (-z_k)^{j_k}, \end{aligned} \quad (2.10)$$

and then use the identity

$$\begin{aligned} J^l \prod_{k=1}^{p} \binom{n_k}{j_k} (-z_k)^{j_k} \\ = \left(-\sum_{k=1}^{p} z_k \frac{\partial}{\partial z_k}\right)^l \prod_{k=1}^{p} \binom{n_k}{j_k} (-z_k)^{j_k}, \end{aligned} \quad (2.11)$$

we obtain

$$\begin{aligned} I' = \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial z_k}\right)^{m_k}_{z_k=1} z_k^{m_k + a_k + b_k - 1}\right] \\ \times \sum_{l=0}^{N-M-1} \alpha_l \left(-\sum_{k=1}^{p} z_k \frac{\partial}{\partial z_k}\right)^l \prod_{k=1}^{p} (1 - z_k)^{n_k}. \end{aligned} \quad (2.12)$$

One can easily verify that the highest derivative acting on the product $\Pi_{k=1}^{p} (1 - z_k)^{n_k}$ is of order $N - 1$. So as before, at least one factor of $(1 - z_k)$, for some k, will survive the differentiation, and then will vanish upon setting $z_k = 1$. Thus we have shown

$$\begin{aligned} \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, P_n(x) w(x) \\ \times \prod_{k=1}^{p} \frac{\Gamma(m_k + b_k - i x_k)}{\Gamma(b_k - i x_k)} = 0, \quad \text{if } N > M. \end{aligned} \quad (2.13)$$

An arbitrary polynomial can be written as a linear combination of products of the form $\Pi_{k=1}^{p} \Gamma(m_k + b_k - i x_k) / \Gamma(b_k - i x_k)$ for different values of m_k. In particular we can write

$$P_m(x) = \sum_{\{j_k\}} \alpha(j_1, j_2, ..., j_p) \prod_{k=1}^{p} \frac{\Gamma(j_k + b_k - i x_k)}{\Gamma(b_k - i x_k)}, \quad (2.14)$$

where $\alpha(j_1, j_2, ..., j_p)$ are some constants. Then we immediately deduce from (2.13) that

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, P_n(x) P_m(x) w(x) = 0, \quad \text{if } N \neq M, \quad (2.15)$$

and thus the $P_n(x)$ are orthogonal with respect to subspaces labeled by N, which, however, says nothing about the different polynomials of the same degree.

Now we demonstrate the analogous result for $Q_n(x)$. Consider the inner product

$$\begin{aligned} \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, Q_n(x) w(x) \\ \times \prod_{k=1}^{p} \frac{\Gamma(m_k + a_k + i x_k)}{\Gamma(a_k + i x_k)}, \end{aligned} \quad (2.16)$$

which upon substituting (1.9) and (1.10) and the use of (2.3) yields, apart from a multiplicative constant,

$$\sum_{\{j_k\}} \prod_{k=1}^{p} \binom{n_k}{j_k} \frac{\Gamma(m_k + j_k + a_k + b_k)}{\Gamma(j_k + a_k + b_k)} (-1)^{j_k}. \quad (2.17)$$

This is identical to (2.6) which in turn is equal to (2.7) which is clearly zero for $N > M$. Thus

$$\begin{aligned} \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, Q_n(x) w(x) \\ \times \prod_{k=1}^{p} \frac{\Gamma(m_k + a_k + i x_k)}{\Gamma(a_k + i x_k)} = 0, \quad \text{if } N > M, \end{aligned} \quad (2.18)$$

and if we write

$$Q_m(x) = \sum_{\{j_k\}} \alpha'(j_1, j_2, ..., j_p) \prod_{k=1}^{p} \frac{\Gamma(j_k + a_k + i x_k)}{\Gamma(a_k + i x_k)}, \quad (2.19)$$

where $\alpha'(j_1, j_2, ..., j_p)$ are some constants, then we immediately deduce that

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, Q_n(x) Q_m(x) w(x) = 0, \quad \text{if } N \neq M. \quad (2.20)$$

Thus the $Q_n(x)$ are also orthogonal with respect to subspaces labeled by N.

Although the polynomial families $P_n(x)$ and $Q_n(x)$ are each orthogonal among themselves with respect to degree N

they are not so for different polynomials of the same degree. In Sec. III we demonstrate that these two families form a biorthogonal system, that is, they are orthogonal to each other with respect to degree and within a subspace of equal degree.

III. BIORTHOGONALITY OF $P_n(x)$ AND $Q_n(x)$

We begin by presenting several identities that will be of use. The first,

$$\frac{\Gamma(\alpha+1)}{\Gamma(\alpha\pm j+1)}=(-1)^j\frac{\Gamma(-\alpha\mp j)}{\Gamma(-\alpha)}, \tag{3.1}$$

where α is some constant, can be verified by inspection. The second identity is the following:

$$\sum_{j=0}^{n}\binom{n}{j}\frac{\Gamma(\alpha)}{\Gamma(\alpha-j)}\frac{\Gamma(\beta)}{\Gamma(\beta-n+j)}$$
$$=\sum_{j=0}^{n}\binom{n}{j}\left[\left(\frac{\partial}{\partial z}\right)_{z=1}^{j}z^{\alpha-1}\right]\left[\left(\frac{\partial}{\partial z}\right)_{z=1}^{n-j}z^{\beta-1}\right]$$
$$=\left(\frac{\partial}{\partial z}\right)_{z=1}^{n}z^{\alpha+\beta-2}=\frac{\Gamma(\alpha+\beta-1)}{\Gamma(\alpha+\beta-n-1)}, \tag{3.2}$$

where β is also some constant. This can then be used repeatedly to derive the formula[5]

$$\sum_{\{j_k\}}^{\{n_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\right]\frac{\Gamma(\alpha)}{\Gamma(\alpha-J)}\frac{\Gamma(\beta)}{\Gamma(\beta-N+J)}$$
$$=\frac{\Gamma(\alpha+\beta-1)}{\Gamma(\alpha+\beta-N-1)}, \tag{3.3}$$

which can be verified by induction on p.

We turn now to calculating the inner product

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\,P_n(x)Q_m(x)w(x), \tag{3.4}$$

which, upon substituting (1.8)–(1.10), and then using the integral formula (2.3), becomes

$$i^N(-i)^M(2\pi)^p\frac{\Gamma(N+A+d)\Gamma(M+B+c)\Gamma(M+c+d)}{\Gamma(N+A+B+c+d-1)}\left[\prod_{k=1}^{p}\frac{\Gamma(n_k+a_k+b_k)}{n_k!\,m_k!}\right]$$
$$\times\sum_{\{j_k\}}\sum_{\{l_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\binom{m_k}{l_k}\frac{\Gamma(j_k+l_k+a_k+b_k)\Gamma(m_k+a_k+b_k)}{\Gamma(j_k+a_k+b_k)\Gamma(l_k+a_k+b_k)}\right]\frac{\Gamma(N+J+A+B+c+d-1)}{\Gamma(M+J+A+B+c+d)}(-1)^{J+L}. \tag{3.5}$$

The subsequent manipulations of (3.5) essentially follow those of Lam and Tratnik[5] for the conformal polynomials.

We use identity (3.1) to write

$$(-1)^{l_k}\Gamma(j_k+l_k+a_k+b_k)=\frac{\Gamma(j_k+a_k+b_k)\Gamma(-j_k-a_k-b_k+1)}{\Gamma(-j_k-l_k-a_k-b_k+1)}, \tag{3.6}$$

which, when substituted into (3.5), gives

$$i^N(-i)^M(2\pi)^p\frac{\Gamma(N+A+d)\Gamma(M+B+c)\Gamma(M+c+d)}{\Gamma(N+A+B+c+d-1)}\left[\prod_{k=1}^{p}\frac{\Gamma(n_k+a_k+b_k)}{n_k!\,m_k!}\right]$$
$$\times\sum_{\{j_k\}}\sum_{\{l_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\binom{m_k}{l_k}\frac{\Gamma(-j_k-a_k-b_k+1)\Gamma(m_k+a_k+b_k)}{\Gamma(-j_k-l_k-a_k-b_k+1)\Gamma(l_k+a_k+b_k)}\right]\frac{\Gamma(N+J+A+B+c+d-1)}{\Gamma(M+J+A+B+c+d)}(-1)^J. \tag{3.7}$$

The $\{l_k\}$ sums can now be performed by using identity (3.2),

$$\sum_{\{l_k\}}\prod_{k=1}^{p}\binom{m_k}{l_k}\frac{\Gamma(-j_k-a_k-b_k+1)}{\Gamma(-j_k-a_k-b_k+1-l_k)}\frac{\Gamma(m_k+a_k+b_k)}{\Gamma(m_k+a_k+b_k-m_k+l_k)}$$
$$=\prod_{k=1}^{p}\frac{\Gamma(m_k-j_k)}{\Gamma(-j_k)}=\prod_{k=1}^{p}(-1)^{m_k}\frac{j_k!}{(j_k-m_k)!}, \tag{3.8}$$

where the second step follows from identity (3.1). Substituting this result into (3.7) we obtain the somewhat simplified expression

$$i^{N+M}(2\pi)^p\frac{\Gamma(N+A+d)\Gamma(M+B+c)\Gamma(M+c+d)}{\Gamma(N+A+B+c+d-1)}\left[\prod_{k=1}^{p}\frac{\Gamma(n_k+a_k+b_k)}{n_k!\,m_k!}\right]$$
$$\times\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\frac{n_k!}{(n_k-j_k)!(j_k-m_k)!}\right]\frac{\Gamma(N+J+A+B+c+d-1)}{\Gamma(M+J+A+B+c+d)}(-1)^J. \tag{3.9}$$

If we reverse the order of the sums $j_k\to n_k-j_k$ and use identity (3.1) to write

$$(-1)^J\Gamma(2N-J+A+B+c+d-1)=\frac{\Gamma(-2N-A-B-c-d+2)\Gamma(2N+A+B+c+d-1)}{\Gamma(-2N-A-B-c-d+2+J)}, \tag{3.10}$$

then (3.9) becomes

$$(-i)^N i^M (2\pi)^p \frac{\Gamma(N+A+d)\Gamma(M+B+c)\Gamma(M+c+d)}{\Gamma(N+A+B+c+d-1)} \left[\prod_{k=1}^{p} \frac{\Gamma(n_k+a_k+b_k)}{m_k!(n_k-m_k)!}\right]$$

$$\times \frac{\Gamma(2N+A+B+c+d-1)}{\Gamma(N+M+A+B+c+d)} \frac{\Gamma(-2N-A-B-c-d+2)}{\Gamma(-N-M-A-B-c-d+2)}$$

$$\times \sum_{\{j_k\}}^{\{n_k-m_k\}} \binom{n_k-m_k}{j_k} \frac{\Gamma(N+M+A+B+c+d)}{\Gamma(N+M+A+B+c+d-J)} \frac{\Gamma(-N-M-A-B-c-d+2)}{\Gamma(-N-M-A-B-c-d+2-N+M+J)}. \quad (3.11)$$

We can now use identity (3.3) to perform the $\{j_k\}$ summations, and we can also use identity (3.1) once more to write

$$\frac{\Gamma(-2N-A-B-c-d+2)}{\Gamma(-N-M-A-B-c-d+2)} \frac{\Gamma(2N+A+B+c+d-1)}{\Gamma(N+M+A+B+c+d)} = \frac{(-1)^{N-M}}{(N+M+A+B+c+d-1)}, \quad (3.12)$$

we have finally, for the inner product (3.4),

$$i^N(-i)^M(2\pi)^p \frac{\Gamma(N+A+d)\Gamma(M+B+c)\Gamma(M+c+d)}{(N+M+A+B+c+d-1)\Gamma(N+A+B+c+d-1)} \left[\prod_{k=1}^{p} \frac{\Gamma(n_k+a_k+b_k)}{m_k!(n_k-m_k)!}\right] \frac{1}{(M-N)!}, \quad (3.13)$$

which is clearly zero unless $n_k = m_k$ for every k. That is,

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \, P_n(x) Q_m(x) w(x) = h_n \prod_{k=1}^{p} \delta_{n_k m_k}, \quad (3.14)$$

where the normalization constant is given by

$$h_n = (2\pi)^p \frac{\Gamma(N+A+d)\Gamma(N+B+c)\Gamma(N+c+d)}{(2N+A+B+c+d-1)\Gamma(N+A+B+c+d-1)} \prod_{k=1}^{p} \frac{\Gamma(n_k+a_k+b_k)}{n_k!}. \quad (3.15)$$

IV. MISCELLANEOUS RESULTS

It is interesting to note that the conformal polynomials can be recovered from the multivariable Hahn polynomials by a limit transition. If we put $a_k = b_k = \mu_k/2$, $c = \mu_{p+1}/2 - i\lambda$, $d = \mu_{p+1}/2 + i\lambda$, replace x_k by λx_k, multiply the polynomials by λ^{-N}, and then take the limit $\lambda \to \infty$, one finds

$$\lim_{\lambda\to\infty} \lambda^{-N} P_n(\lambda x) = \frac{\Gamma(N+\mu_{p+1})}{(2N+\mu-1)\Gamma(N+\mu-1)} \left[\prod_{k=1}^{p} \Gamma(n_k+\mu_k)\right] D_n^{\mu}(x),$$

$$\lim_{\lambda\to\infty} \lambda^{-N} Q_n(\lambda x) = \left[\prod_{k=1}^{p} \frac{1}{n_k!}\right] C_n^{\mu}(x), \quad (4.1)$$

which elucidates the close connection between these two families.

Now let us consider a few specific examples where the weight function takes on a more familiar form. If $a_k = b_k^*$, $k = 1,2,...,p$, and $c = d^*$ and the real parts are greater than zero then one can easily verify that the weight function is pure real. More specifically if $a_k = b_k = t_k + 1$, $c = d = t_{p+1} + 1$, where $t_1, t_2, ..., t_{p+1}$ are non-negative integers, then using known properties of the gamma function[1] we can write (apart from a multiplicative constant)

$$w(x) = \left[\prod_{k=1}^{p} \frac{(\pi x_k)}{\sinh(\pi x_k)} \prod_{j_k=1}^{t_k} \left(1 + \frac{x_k^2}{j_k^2}\right)\right] \frac{(\pi X)}{\sinh(\pi X)} \prod_{j_{p+1}=1}^{t_{p+1}} \left(1 + \frac{X^2}{j_{p+1}^2}\right), \quad (4.2)$$

which in the simplest case $t_k = 0$, $k = 1, 2, ..., p+1$, reduces to

$$w(x) = \left[\prod_{k=1}^{p} \frac{(\pi x_k)}{\sinh(\pi x_k)}\right] \frac{(\pi X)}{\sinh(\pi X)}. \quad (4.3)$$

Another interesting family is that for the parameter values $a_k = b_k = t_k + \frac{1}{2}$, $c = d = t_{p+1} + \frac{1}{2}$, where again $t_1, t_2, ..., t_{p+1}$ denote non-negative integers. In this case we can write[1] for the weight function (apart from a multiplicative constant)

$$w(x) = \left[\prod_{k=1}^{p} \operatorname{sech}(\pi x_k) \prod_{j_k=1}^{t_k} \left\{1 + \frac{x_k^2}{(j_k - \frac{1}{2})^2}\right\}\right] \operatorname{sech}(\pi X) \prod_{j_{p+1}=1}^{t_{p+1}} \left\{1 + \frac{X^2}{(j_{p+1} - \frac{1}{2})^2}\right\}, \quad (4.4)$$

which in the simplest case $t_k = 0$, $k = 1,2,...,p+1$, reduces to just a product of hyperbolic secants:

$$w(x) = \left[\prod_{k=1}^{p} \operatorname{sech}(\pi x_k)\right] \operatorname{sech}(\pi X). \quad (4.5)$$

We display below a few of the two variable polynomials for the latter case:

$$
\begin{aligned}
&P_{00}(x_1,x_2) = 1, && Q_{00}(x_1,x_2) = 1,\\
&P_{10}(x_1,x_2) = 3x_1, && Q_{10}(x_1,x_2) = 2x_1 + x_2,\\
&P_{01}(x_1,x_2) = 3x_2, && Q_{01}(x_1,x_2) = x_1 + 2x_2,\\
&P_{20}(x_1,x_2) = -\tfrac{5}{4} + 10x_1^2, && Q_{20}(x_1,x_2) = -\tfrac{1}{2} + 6x_1^2 + 6x_1x_2 + x_2^2,\\
&P_{11}(x_1,x_2) = \tfrac{5}{4} + 20x_1x_2, && Q_{11}(x_1,x_2) = -\tfrac{1}{4} + 3x_1^2 + 8x_1x_2 + 3x_2^2,\\
&P_{02}(x_1,x_2) = -\tfrac{5}{4} + 10x_2^2, && Q_{02}(x_1,x_2) = -\tfrac{1}{2} + 6x_2^2 + 6x_1x_2 + x_1^2.
\end{aligned}
$$

Notice that these polynomials have all real coefficients, which is not obvious from the defining equations (1.8) and (1.9). In fact they are real whenever $a_k = b_k^*$, $k = 1,2,\ldots,p$, and $c = d^*$, and if in addition the real parts of these parameters are greater than zero then the weight function is also real. This point will be discussed in a future publication where we also consider symmetry properties, Rodrigues-type formulas, and the discrete analog of these polynomials.

We are also currently investigating analogous multivariable extensions of other families of hypergeometric orthogonal polynomials.

ACKNOWLEDGMENTS

I would like to thank the referee for pointing out the limit relation between the Hahn and conformal polynomials and for other useful comments.

This work was supported by the Natural Sciences and Engineering Research Council of Canada and the United States Department of Energy.

[1] H. Bateman, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vols. 1–3.
[2] N. M. Atakishiyev and S. K. Suslov, J. Phys. A **18**, 1583 (1985).
[3] R. Askey, J. Phys. A **18**, L1017 (1985).
[4] C. M. Bender, L. R. Mead, and S. S. Pinsky, J. Math. Phys. **28**, 509 (1987).
[5] C. S. Lam and M. V. Tratnik, Can. J. Phys. **63**, 1427 (1985).
[6] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).

A remark concerning canonical commutation relations

Helmut Reeh
Institut für Theoretische Physik, Universität Göttingen, Bunsenstrasse 9, D-3400 Göttingen, West Germany

(Received 29 September 1987; accepted for publication 16 March 1988)

An elementary quantum mechanical example of canonical commutation relations not equivalent to the Schrödinger representation is discussed. It demonstrates that the "Nelson phenomenon" occurs in a natural situation.

In this short paper I should like to present an elementary example demonstrating that in quantum mechanics, for systems with a finite number of degrees of freedom, representations of the canonical commutation relations occur that are not equivalent to the Schrödinger representation and that cannot be integrated to the Weyl form (thus avoiding the applicability of the uniqueness theorem of von Neumann[1]). The example is related to an idealized description of the Aharonov–Bohm effect. I do not claim originality for this example, on the contrary, I am convinced that there are quite a number of experts aware of the situation (see, e.g., Ref. 2). However, the example is so simple that it should be of general interest, in particular since it demonstrates that the "Nelson phenomenon" occurs quite naturally.

In a theoretical description of the Aharonov–Bohm effect, a charged particle—e.g., an electron—is considered outside a cylinder extending from $-\infty$ to $+\infty$, say along the z axis containing in its interior a magnetic flux parallel to its axis. The electron is (approximately) confined to the outside of the cylinder where there is (approximately) no magnetic field. Then the state space for the electron is $\mathscr{L}^2$ ($\mathbb{R}^3$ cylinder). In a limiting case (rather difficult to realize in experiment by very low energy electrons), the radius of the cylinder may be considered to tend to zero, if the flux inside is kept fixed. Disregarding the degree of freedom parallel to the z axis, we are left with only two degrees of freedom and the canonical operators in Schrödinger representation heuristically may be chosen to be

$$\mathbf{p} = \frac{1}{i}\frac{\partial}{\partial \mathbf{x}} + e\mathbf{A}, \quad \mathbf{q} = \mathbf{x}, \tag{1}$$

$\mathbf{x} := (x,y)$ is the multiplication operators, $\mathbf{A}$ is the vector potential, $\mathbf{A} = (\alpha/\sqrt{x^2+y^2})\mathbf{e}$ (or is gauge equivalent to this), $\mathbf{e} = (-y/\sqrt{x^2+y^2}, x/\sqrt{x^2+y^2})$ is the unit vector in tangential direction. The charge e will be absorbed into the real flux α.

For giving (1) a definite meaning as operators on $\mathscr{L}^2(\mathbb{R}^2-0) = \mathscr{L}^2(\mathbb{R}^2)$, we have to specify the domains in particular of p_x and p_y. Because of the singularity of $\mathbf{A}$ at 0, we take for both domains the dense set

$$\vartheta = \{\varphi \in \mathscr{D}(\mathbb{R}^2) : \operatorname{supp}\varphi \not\ni (0,0)\},$$

the C^∞ functions of compact support vanishing in neighborhoods of the origin. This ϑ is invariant under $\mathbf{p}$ and under the multiplication operators x and y, and on ϑ the canonical commutation relations are fulfilled. Among these relations are

$$[p_x, p_y]\varphi = 0, \quad \forall \varphi \in \vartheta,$$

because $\operatorname{rot}\mathbf{A} = 0$ outside (0,0). To be specific, we now fix the gauge as specified.

Consider, p_x. Apparently, it is Hermitian on ϑ and

$$p_x^* \varphi = \frac{1}{i}\frac{\partial}{\partial x}\varphi - \alpha\frac{y}{x^2+y^2}\varphi,$$

with the derivative as distribution at least for φ vanishing at the origin. If the solutions of the differential equation

$$p_x^* \psi = \pm i\psi$$

are considered,

$$\psi = c(y)e^{\mp x}e^{i\alpha \arctan(x/y)}$$

($-\pi/2 < \arctan s < \pi/2$, for $-\infty < s < \infty$), it follows that p_x has vanishing deficiency indices[3]; hence, from a theorem of von Neumann, the (unique) closure $\bar{p}_x$ of p_x is self-adjoint.

The eigenvalue equation

$$\left(\frac{\partial}{\partial x} - i\alpha\frac{y}{x^2+y^2}\right)\psi = ik\psi$$

has the solutions

$$\psi_k = c(y)e^{ikx}e^{i\alpha \arctan(x/y)} \tag{2}$$

$$= \hat{c}(y)e^{ikx}e^{i\alpha(\arctan(x/y) - \pi\Theta(-y))} \tag{2'}$$

[$\Theta(s) := 1$, for $s>0$, $\Theta(s) := 0$, for $s<0$]. Observe that the last exponent of (2) is smooth in x; however, it is discontinuous in y at $y = 0$ with jump π, whereas the corresponding bracket in (2') is discontinuous in y only at the positive x axis with a jump of 2π! For every fixed y, the $\{\psi_k\}_k$ is complete for $\mathscr{L}^2(\mathbb{R}^1)$.

Similarly, p_y has eigenfunctions

$$\psi_l' = d(x)e^{ily}e^{-i\alpha \arctan(y/x)} \tag{3}$$

$$= \hat{d}(x)e^{ily}e^{-i\alpha(\arctan(y/x) + \pi\Theta(-x))} \tag{3'}$$

(with jumps in x by π at $x = 0$ and by 2π at the negative y axis).

Because of $\arctan s + \arctan(1/s) = (\pi/2)\epsilon(s)$ [with $\epsilon(s) = 1$, for $s>0$, $\epsilon(s) = -1$, for $s<0$], it is easy to see that a simultaneous complete set of eigenfunctions of p_x, p_y exists in the case where $\alpha = 0, \pm 1, \pm 2, \ldots$. It does not in the other cases!

It also follows from (2) [or (3)] that $\bar{p}_x$ (or $\bar{p}_y$) is unitarily equivalent to $(1/i)(\partial/\partial x)$ [or $(1/i)(\partial/\partial y)$] on a suitable domain; however, the unitary intertwiner in general depends on y (or x)! Likewise, for the exponentiated unitaries,

 0022-2488/88/071535-02$02.50

$$e^{ia\bar{p}_x} = e^{-i\alpha \arctan(x/y)} e^{ia(1/i)\partial_x} e^{i\alpha \arctan(x/y)}$$

$$= e^{-i\alpha \arctan(x/y)} e^{i\alpha \arctan[(x+a)/y]} e^{ia(1/i)\partial_x}, \quad (4)$$

since $e^{ia(1/i)\partial_x}$ implements the shift in the x direction on $\mathscr{L}^2(\mathbb{R}^2)$, and

$$e^{ib\bar{p}_y} = e^{i\alpha \arctan(y/x)} e^{ib(1/i)\partial_y} e^{-i\alpha \arctan(y/x)}$$

$$= e^{i\alpha \arctan(y/x)} e^{-i\alpha \arctan[(y+b)/x]} e^{ib(1/i)\partial_y}. \quad (5)$$

Hence the canonical operators for each separate degree of freedom (p_x, q_x) and (p_y, q_y) are unitarily equivalent to the Schrödinger representation. However, in general, this is not the case for both degrees of freedom together [and, in general, there is no vector in $\mathscr{L}^2(\mathbb{R}^2)$ annihilated by both a_x and a_y, $a := (1/\sqrt{2})(q + ip)$, i.e., the representation is "non-Fock"]. In fact, a simple calculation leads to

$$e^{-ib\bar{p}_y} e^{ia\bar{p}_x} e^{ib\bar{p}_y} = e^{i\alpha[(\epsilon(x) - \epsilon(x+a))(\epsilon(y) - \epsilon(y-b))](\pi/2)} e^{ia\bar{p}_x}.$$

Observe that $[\cdots]$ in the exponent assumes values from $\{0, \pm 4\}$. Hence $e^{ia\bar{p}_x}$ and $e^{ib\bar{p}_y}$ commute if and only if α is integer.

So much for the interesting aspects of the example in sufficient detail and precision [a mathematically more complete treatment would just directly start from the unitary equivalences (4) and (5) and the infinitesimal generators on suitable domains]. An essential aspect of p_x, p_y is that they provide an example of Nelson's type[4] for two operators essentially self-adjoint on ϑ, commuting there, leaving ϑ invariant, but with noncommuting spectral measures. Its nice feature is that it is directly related to a physical situation. Perhaps it should also be mentioned that $p_x^2 + p_y^2$ is not essentially self-adjoint on ϑ (for integer as well as for noninteger α).[5]

[1] J. von Neumann, Math. Ann. **104**, 570 (1931).

[2] K. Sekine, *Proceedings of the 2nd International Symposium on Foundations of Quantum Mechanics*, Tokyo, 1986, p. 127.

[3] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II, Chap X.

[4] E. Nelson, Ann. Math. **70**, 572 (1959); see, also, Ref. 3, Vol. I, Chap. VIII 5, pp. 237 and 306.

[5] S. N. M. Ruijsenaars, Ann. Phys. (NY) **146**, 1 (1983); see, also, Ref. 3, Vol. II, p. 161.

Elliptical billiards and Poncelet's theorem

Shau-Jin Chang and Richard Friedberg[a)]
Department of Physics, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, Illinois 61801

(Received 14 August 1987; accepted for publication 2 March 1988)

Two- and three-dimensional billiard systems with elliptical and ellipsoidal boundaries, respectively, are studied. It is known that the trajectory of such a two-dimensional system generates a caustic conic curve. Many properties of the elliptical billiard system in the language of projective geometry can be described. One of these properties is that if a trajectory is closed after p bounces, then all trajectories sharing the same caustic conic are also closed after p bounces. In projective geometry, this is known as Poncelet's theorem. Many of the two-dimensional results are extended to three dimensions. In particular, it is shown that all straight segments of a trajectory in a three-dimensional ellipsoidal billiard system are always tangent to two caustic quadrics. If a trajectory is closed after p bounces, then all trajectories sharing the same two caustic quadrics are also closed after p bounces. A generalized Poncelet's theorem in three dimensions is thus established. On the basis of numerical studies, it is conjectured that this generalized Poncelet's theorem is also valid for an arbitrary finite number field. The implications of the results are discussed and their extension to n dimensions is outlined.

"On a cloth untrue,
With a twisted cue,
And elliptical billiard balls."
—W. S. Gilbert, *The Mikado*

I. INTRODUCTION

Birkhoff introduced the notion of a billiard system to describe the motion of a free particle inside a closed curve. The particle reflects perfectly at the boundary and follows a straight-line path between the bounces.[1] He found that the billiard system has an invariant measure, and by the use of Poincare's method of the return map,[2] it can be cast as a discrete area-preserving mapping problem.

It has been long known that the trajectory of a billiard system with an elliptical boundary always generates a caustic curve. This curve is a conic section confocal to the original ellipse. Keller and Rubinow[3] and Sinai[4] are among the earliest investigators of such a system. Sinai gave a simple geometrical proof of the existence of a caustic conic which we shall review in Sec. II.

Many properties of the elliptical billiard system were known in different forms to nineteenth century mathematicians. Given the existence of the inner caustic conic, we can describe the trajectory of a billiard as a collection of straight-line segments joining on the outer conic, and tangent to the inner conic. Since the concepts of "straight line," "conic," "intersection," and "tangent" are projectively invariant, we can translate many of the physical properties of our billiard system into the language of projective geometry. One of these properties is that if a trajectory is closed after p bounces, then all trajectories sharing the same caustic conic are also closed after p bounces. In projective geometry, this is known as Poncelet's theorem.[5] This theorem was studied and extended by many nineteenth century mathematicians. Jacobi recognized certain algebraic relations between the geometrical construction of Poncelet's theorem and the addition formula for elliptical functions.[6,7] Hart gave a purely projective geometrical proof that relies entirely on finite operations.[8] Hart's result showed that Poncelet's theorem can be extended to any number field system including all the finite number fields.

In this paper, we extend many of the results to three and higher dimensions. Making use of elliptical coordinates,[9] we can separate the Hamilton–Jacobi equation. From the equations of motion in this new coordinate system, we can show that the trajectory of a particle in a three-dimensional ellipsoid gives rise to two caustic quadric surfaces that are confocal to the original ellipsoid. We have established a three-dimensional extension of Poncelet's theorem in real space. (A quite different three-dimensional version of Poncelet's theorem has been proved by Griffiths and Harris.[10] We are indebted to Dr. Harris for guiding us to the earlier literature on the subject.) In addition, we have numerical evidence that this extension is valid over finite number fields as well. If this is true, it has several important implications. In algebra, this theorem suggests that there exists multidimensional addition theorems for a set of generalized elliptical functions. In geometry, it suggests the existence of a purely projective geometrical proof for our three-dimensional Poncelet's theorem. We have not quite been able to construct such a proof, but we suspect that one exists. In physics, this extension (which we have proved in real space) tells us that if any trajectory in a three-dimensional ellipsoid is closed after p bounces, then all the trajectories which share the same caustic quadric surfaces are also closed after p bounces independent of the starting point. In addition, all these trajectories

[a)] Visiting from Barnard College and Columbia University, New York, New York 10027.

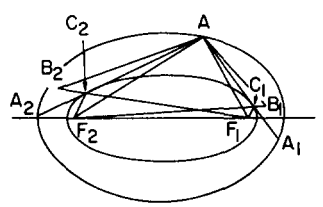
 0022-2488/88/071537-14$02.50

have the same period in time and the same total path length in space. The latter statement depends on the special properties of the confocal quadric surfaces and not on the invariant properties of projective transformations. In quantum physics, the above facts have important implications for the stability and nature of WKB wave functions.[11]

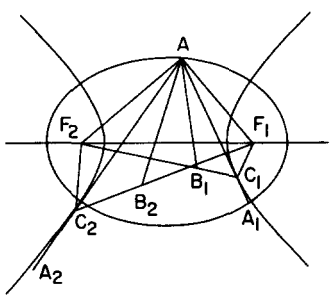
II. THE TWO-DIMENSIONAL MODEL

Keller and Rubinow[3] and Sinai[4] discussed the motion of a free point particle (called a billiard) inside an ellipse. The billiard moves along a straight line between bounces, and reflects at the elliptical boundary according to the law of reflection for a light ray. Every segment or its extension of the trajectory is tangent to a unique conic confocal with the original ellipse. This conic forms a caustic to the trajectory, and may be either an ellipse or a hyperbola. In this section, we shall study the motion of these trajectories explicitly in the two-dimensional model.

We follow Sinai[4] to give a simple geometrical proof for the existence of the caustics. In Fig. 1(a), we consider the trajectory of a billiard as A_1A, AA_2 where the billiard has a bounce at A. Let F_1 and F_2 be the foci, and let B_1 and B_2 be the reflections of F_1 and F_2 relative to A_1A and AA_2, respectively. Draw F_2B_1 and it intersects A_1A at C_1. Similarly, draw F_1B_2 and it intersects AA_2 at C_2. It is a simple matter to show that $F_1C_1 + F_2C_1 = F_2B_1$ and it is the smallest sum of distances from F_1 and F_2 to any point on A_1A. Similarly, $F_1C_2 + F_2C_2 = F_1B_2$ and it is the smallest sum of distances from the foci to points on AA_2. The law of reflection implies that $\angle A_1AF_1 = \angle F_2AA_2$ which leads to congruent triangles AF_2B_1 and AB_2F_1. This implies $F_2B_1 = F_1B_2$ and consequently $F_1C_1 + F_2C_1 = F_1C_2 + F_2C_2$. Thus an ellipse defined by the foci at F_1 and F_2 and with the sum of distances given above will be tangent to A_1A and AA_2 at C_1 and C_2. Further iterations of the trajectory indicate that they should all be tangent to the same inner ellipse as shown. If the initial trajectory falls between F_1F_2, the caustic curve becomes a hyperbola. One can construct easily a similar geometrical proof in this case as well. [See Fig. 1(b).]



(a)



(b)

FIG. 1. (a) When a particle travels outside the foci F_1 and F_2, different segments of the trajectory, here A_1A and AA_2, are tangent to the same inner ellipse. Points B_1 and B_2 are the images of F_1 and F_2 with respect to lines AA_1 and AA_2, respectively. Points C_1 and C_2 are the intersection of F_2B_1 and AA_1, and of F_1B_2 and AA_2, respectively. (b) When the particle travels between the foci, the caustic curve is a hyperbola. The construction is similar to case (a). One can establish easily the following equalities: $F_2C_1 - F_1C_1 = F_2B_1$, $F_1C_2 - F_2C_2 = F_1B_2$, and $F_2B_1 = F_1B_2$.

III. TWO-DIMENSIONAL HAMILTON–JACOBI EQUATION

The Lagrangian for the billiard system is

$$L = \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2) - V(x,y), \tag{3.1}$$

where $V = 0$ inside and $V = \infty$ outside the ellipse

$$x^2/a^2 + y^2/b^2 = 1. \tag{3.2}$$

We introduce the elliptical coordinates (ρ,θ) by Ref. 9

$$x = c\cosh\rho\cos\theta, \tag{3.3a}$$

$$y = c\sinh\rho\sin\theta, \tag{3.3b}$$

where

$$c \equiv \sqrt{a^2 - b^2} \tag{3.4}$$

is $\frac{1}{2}$ the distance between the foci. It is straightforward to see that the $\rho =$ const curves are ellipses and $\theta =$ const curves are hyperbolas. Both families of curves share the same foci at $(\pm c,0)$ as shown in Fig. 2.

In terms of the elliptical coordinates, the Lagrangian and Hamiltonian of the billiard system become

$$L = (c/2)(\sinh^2\rho + \sin^2\theta)(\dot{\rho}^2 + \dot{\theta}^2) - V(\rho,\theta) \tag{3.5}$$

and

$$H = [2c(\sinh^2\rho + \sin^2\theta)]^{-1}(p_\rho^2 + p_\theta^2) + V(\rho,\theta), \tag{3.6}$$

respectively, where for simplicity we have chosen the mass of the billiard to be 1.

To facilitate the transition to three dimensions, we shall introduce the Jacobi variables (λ_1,λ_2) here, and put $A = a^2$ and $B = b^2$. We express the ellipse associated with $\rho =$ const as

$$x^2/(A - \lambda_1) + y^2/(B - \lambda_1) = 1 \tag{3.7}$$

with

$$A - \lambda_1 = c^2\cosh^2\rho, \tag{3.8a}$$

$$B - \lambda_1 = c^2\sinh^2\rho. \tag{3.8b}$$

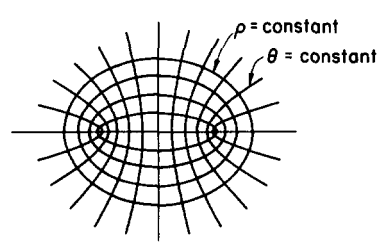


FIG. 2. Two-dimensional elliptical coordinates.

These expressions define λ_1. Similarly, we can express the hyperbola $\theta = \text{const}$ as

$$x^2/(A-\lambda_2) + y^2/(B-\lambda_2) = 1 \tag{3.9}$$

with

$$A - \lambda_2 = c^2 \cos^2\theta, \tag{3.10a}$$

$$B - \lambda_2 = -c^2 \sin^2\theta. \tag{3.10b}$$

The ranges of parameters for λ_1 and λ_2 are

$$-\infty < \lambda_1 < B < \lambda_2 < A. \tag{3.11}$$

Even though the transformation from (x,y) to λ_1, λ_2 is single valued, the inverse is not unique. There are four points in the xy plane which give the same (λ_1,λ_2). We can remove this ambiguity by assigning a discrete index to denote in which quadrant the point resides. However, this ambiguity does not affect any of our conclusions. Since λ_1 and λ_2 are single-variable functions of ρ and θ, respectively, we can obtain the Lagrangian and Hamiltonian in terms of λ_1 and λ_2 straightforwardly,

$$L = \tfrac{1}{2}(g_1\dot{\lambda}_1^2 + g_2\dot{\lambda}_2^2) - V(\lambda_1,\lambda_2), \tag{3.12}$$

$$H = \tfrac{1}{2}(p_1^2/g_1 + p_2^2/g_2) + V(\lambda_1,\lambda_2), \tag{3.13}$$

with

$$g_1 \equiv (\lambda_2 - \lambda_1)/[4(A-\lambda_1)(B-\lambda_1)], \tag{3.14a}$$

$$g_2 \equiv (\lambda_1 - \lambda_2)/[4(A-\lambda_2)(B-\lambda_2)], \tag{3.14b}$$

where

$$V = 0, \quad \text{for } \lambda_1 > 0, \quad V = \infty, \quad \text{for } \lambda_1 < 0. \tag{3.15}$$

To achieve a separation of variables while treating V as a proper limit of a finite V, we choose

$$V(\lambda_1\lambda_2) = [1/(\lambda_2 - \lambda_1)]V_\epsilon(\lambda_1), \tag{3.16}$$

where

$$\begin{aligned} &V_\epsilon(\lambda_1) = 0, \quad \text{for } \lambda_1 > \epsilon, \\ &V_\epsilon(\lambda_1) = V_0/\epsilon, \quad V_0 > 0, \quad \text{for } \lambda_1 < -\epsilon, \end{aligned} \tag{3.17}$$

and V_ϵ is continuous in the region $(-\epsilon,\epsilon)$. The Hamilton–Jacobi equation is

$$\begin{aligned} &2\left[\frac{(A-\lambda_1)(B-\lambda_1)}{\lambda_2-\lambda_1}\left(\frac{\partial W}{\partial\lambda_1}\right)^2 \right. \\ &\left. + \frac{(A-\lambda_2)(B-\lambda_2)}{\lambda_1-\lambda_2}\left(\frac{\partial W}{\partial\lambda_2}\right)^2\right] \\ &+ \frac{1}{\lambda_2-\lambda_1}V_\epsilon(\lambda_1) = \alpha, \quad \text{the energy.} \end{aligned} \tag{3.18}$$

To achieve the separation of variables, we postulate

$$W = W_1(\lambda_1) + W_2(\lambda_2) \tag{3.19}$$

and then multiply (3.18) by $(\lambda_2 - \lambda_1)$ obtaining

$$\begin{aligned} &2(A-\lambda_1)(B-\lambda_1)\left(\frac{dW_1}{d\lambda_1}\right)^2 \\ &- 2(A-\lambda_2)(B-\lambda_2)\left(\frac{dW_2}{d\lambda_2}\right)^2 + V_\epsilon(\lambda_1) \\ &= \alpha(\lambda_2 - \lambda_1). \end{aligned} \tag{3.20}$$

We then have the separation equation,

$$\begin{aligned} &2(A-\lambda_1)(B-\lambda_1)\left(\frac{dW_1}{d\lambda_1}\right)^2 + V_\epsilon(\lambda_1) + \alpha\lambda_1 \\ &= 2(A-\lambda_2)(B-\lambda_2)\left(\frac{dW_2}{d\lambda_2}\right)^2 + \alpha\lambda_2 \\ &= \alpha\alpha', \quad \text{separation constant.} \end{aligned} \tag{3.21}$$

The momenta associated with λ_1 and λ_2 are

$$p_1 \equiv \frac{dW_1}{d\lambda_1} = \left(\frac{\alpha(\alpha'-\lambda_1) - V_\epsilon(\lambda_1)}{2(A-\lambda_1)(B-\lambda_1)}\right)^{1/2}, \tag{3.22a}$$

$$p_2 \equiv \frac{dW_2}{d\lambda_2} = \left(\frac{\alpha(\alpha'-\lambda_2)}{2(A-\lambda_2)(B-\lambda_2)}\right)^{1/2}, \tag{3.22b}$$

where the signs of the square roots are determined by requiring $p_i\dot{\lambda}_i$ to be positive. From the reality of p_1 and p_2, we find

$$\lambda_1 < \alpha' < \lambda_2. \tag{3.23}$$

Depending on the sign of $\alpha' - B$, we encounter two different physical situations:

$$\text{case 1: } 0 < \lambda_1 < \alpha' < B < \lambda_2, \tag{3.24a}$$

$$\text{case 2: } 0 < \lambda_1 < B < \alpha' < \lambda_2. \tag{3.24b}$$

In case 1, variable λ_1 oscillates between 0 and α'. The trajectory bounces at the outer ellipse at $\lambda_1 = 0$ (because of V_ϵ) and comes to touch (i.e., is tangent to) the inner ellipse at $\lambda_1 = \alpha'$, as shown in Fig. 3(a). Since this confocal conic depends only on the separation constant α', it is the same conic for all segments of the trajectory and hence forms a caustic curve. Variable λ_2 is the one associated with angle θ and it oscillates between B and A. Every oscillation leads to an increase of θ by 2π. In this case, the trajectory moves outside the foci.

In case 2, variable λ_1 oscillates between 0 and B. The trajectory passes through the x axis between the foci (at $\lambda_1 = B$) and bounces at the outer ellipse (at $\lambda_1 = 0$). Variable λ_2 now oscillates in a restricted region $\alpha' < \lambda_2 < A$ which corresponds to the region between two branches of the hyperbola $\lambda_2 = \alpha'$ [see Fig. 3(b)]. The turn at the upper limit corresponds to the crossing of the vertical axis which has no special meaning. The turn at the lower limit $\lambda_2 = \alpha'$ corresponds to the ray being tangent to the hyperbola $\lambda_2 = \alpha'$. In this case, we have the hyperbola as the caustic conic.

The energy α and the separation constant α' are constants of motion. We choose them as the generalized momenta. The coordinates conjugate to α and α' have trivial time dependence,

$$\begin{aligned} Q = \frac{\partial W}{\partial\alpha} = \frac{1}{2}\left(\frac{1}{2\alpha}\right)^{1/2}\left[\int d\lambda_1\left(\frac{\alpha'-\lambda_1}{(A-\lambda_1)(B-\lambda_1)}\right)^{1/2}\right. \\ \left. + \int d\lambda_2\left(\frac{\lambda_2-\alpha'}{(A-\lambda_2)(\lambda_2-B)}\right)^{1/2}\right] \\ = t + \text{const}, \end{aligned} \tag{3.25a}$$

$$\begin{aligned} Q' = \frac{\partial W}{\partial\alpha'} = \frac{1}{2}\left(\frac{\alpha}{2}\right)^{1/2} \\ \times\left[\int d\lambda_1 \frac{1}{\sqrt{(\alpha'-\lambda_1)(A-\lambda_1)(B-\lambda_1)}}\right. \\ \left. - \int d\lambda_2 \frac{1}{\sqrt{(\lambda_2-\alpha')(A-\lambda_2)(\lambda_2-B)}}\right] = \text{const}, \end{aligned} \tag{3.25b}$$

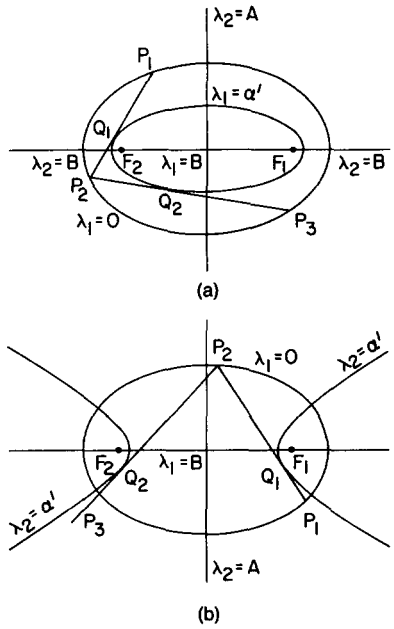


FIG. 3. The separation constant α' determines the caustic conic. (a) For $\alpha' < B$, the particle starts at P_1 on the outer ellipse ($\lambda_1 = 0$), touches the inner ellipse at Q_1 ($\lambda_1 = \alpha'$), and returns to the outer ellipse ($\lambda_1 = 0$) at P_2. Thus, the value of λ_1 oscillates one period between 0 and α' for every bounce. (b) For $\alpha' > B$, the value of λ_1 starts from 0 at P_1, increases monotonically to B as the particle crosses F_1F_2, and returns to 0 at P_2. Thus λ_1 oscillates between 0 and B. The end point at $\lambda_1 = B$ does not correspond to any caustic curve. The caustic hyperbola $\lambda_2 = \alpha'$ corresponds to the end point of an oscillation in variable λ_2.

where we have ignored V_ϵ in the expressions, and replaced it by a restriction on the integration limits. Since the square root in the ith term of Q has the same sign as that of $d\lambda_i$, the integrals follow contours that half-encircle a branch point at each turn. However, the omission of V_ϵ makes it appear that the contour for λ_1 jumps over the branch cut.

Equation (3.25b) determines λ_2 implicitly as a function of λ_1. We can express (3.25b) in terms of elliptic integrals. In particular, in case 1, we have (see Ref. 7, p. 597)

$$\int_{\lambda_1}^{\alpha'} \frac{d\lambda}{\sqrt{(\alpha'-\lambda)(A-\lambda)(B-\lambda)}} = \frac{2}{\sqrt{A-\alpha'}} F(\phi_1(\lambda_1)|k^2), \tag{3.26a}$$

$$\int_{\lambda_2}^{A} \frac{d\lambda}{\sqrt{(\lambda-\alpha')(A-\lambda)(\lambda-B)}} = \frac{2}{\sqrt{A-\alpha'}} F(\phi_2(\lambda_2)|k^2), \tag{3.26b}$$

where

$$F(\phi|k^2) \equiv \int_0^{\phi} d\theta (1-k^2\sin^2\theta)^{-1/2}, \tag{3.27}$$

$$k^2 = (A-B)/(A-\alpha'), \tag{3.28}$$

$$\cos^2\phi_1(\lambda) = (B-\alpha')/(B-\lambda), \tag{3.29a}$$

$$\cos^2\phi_2(\lambda) = (\lambda-B)/(A-B). \tag{3.29b}$$

Note that k is the eccentricity of the inner (caustic) ellipse, and ϕ_1 and ϕ_2 are functions only of λ_1 and λ_2, respectively. Within a straight segment, we have from (3.25b)

$$F(\phi_1(\lambda_1)|k^2) - F(\phi_2(\lambda_2)|k^2) = \text{const.} \tag{3.30}$$

This equation determines the trajectory of the billiard. [The minus sign in the second term comes from the factor $(\alpha'-\lambda_2)^{-1}$ in (3.25b) relative to (3.25a).]

In one bounce, the first term of (3.30) changes by $2F(\phi_1(0)|k^2)$, and therefore so does the second. Thus the initial and final λ_2 of a bounce are related by

$$\Delta F(\phi_2(\lambda_2)|k^2) = 2F(\phi_1(0)|k^2). \tag{3.31}$$

We can use Eq. (3.31) to determine the condition for the trajectory to form a closed orbit (i.e., a cycle). For each of these cycles, we introduce a pair of integers (p,q), called the winding numbers, where p represents the number of bounces and q the number of 2π rotations in θ (circuits about the inner ellipse) within each period.

In going around the caustic once, $\phi_2(\lambda_2)$ changes by 2π and hence $F(\phi_2(\lambda_2)|k^2)$ changes by $4K$ where $K = F(\pi/2|k^2)$. Therefore (3.31) gives

$$2pF(\phi_1(0)|k^2) = 4Kq \tag{3.32}$$

on summing over p bounces.

Note that $b = \sqrt{B}$ is the semiminor axis of the outer ellipse, and $b' \equiv \sqrt{B-\alpha'}$ is the semiminor axis of the inner (caustic) ellipse. Hence $\cos\phi_1(0) = b/b'$, and we can write (3.32) as

$$b'/b = \mathrm{cn}(2qK/p). \tag{3.33}$$

It is important to note that all cycles with the same winding numbers share the same caustic, independent of the starting points. From (3.25a) we can also show that the total time (if the energy is fixed), or more generally the total traveling length, of periodic orbits of the same winding number is independent of the starting points.

IV. THREE-DIMENSIONAL MODEL

We can generalize the billiard problem to three dimensions by considering the motion inside an ellipsoidal surface,

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1. \tag{4.1}$$

In the following, we shall denote a^2, b^2, c^2 by A, B, C, respectively. The three-dimensional generalization of a conic confocal to (4.1) is

$$\frac{x^2}{A-\lambda} + \frac{y^2}{B-\lambda} + \frac{z^2}{C-\lambda} - 1 = 0. \tag{4.2}$$

Depending on the value of λ, we can have either an ellipsoid, a hyperboloid of one sheet, or a hyperboloid of two sheets. See Figs. 4(a)–4(f) for these surfaces.

At any given point in space, we can find three mutually orthogonal confocal quadric surfaces passing through it. The λ's associated with these surfaces are the Jacobi variables describing the point. This will be our new coordinate system. Locally, the coordinate system determines the Cartesian coordinates of a point unambiguously. Globally, the λ's do not determine the point uniquely. The eight points with the same $|x|$, $|y|$, and $|z|$ have the same λ's. However, this ambiguity is not important to our application, and can

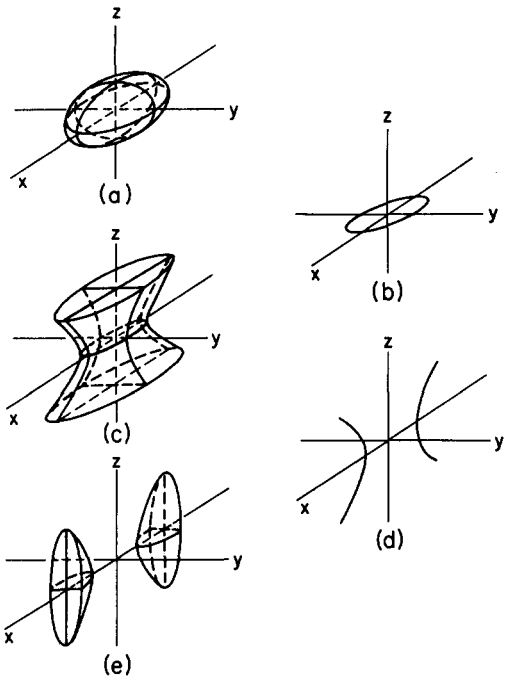


FIG. 4. Confocal quadric surfaces associated with different values of λ. (a) $\lambda < C$. The quadric is an ellipsoid. (b) $\lambda = C$. As λ approaches C from below ($\lambda = C - 0$), the ellipsoid flattens into an ellipse in the xy plane. At $\lambda = C + 0$, the quadric surface becomes the whole xy plane with the ellipse removed. (c) $C < \lambda < B$. The quadric is a hyperboloid of one sheet. (d) $\lambda = B$. At $\lambda = B - 0$, the hyperboloid flattens into the region between the branches of a hyperbola in the xz plane. At $\lambda = B + 0$, the quadric becomes the outside regions of the same hyperbola in the xz plane. (e) $B < \lambda < A$. The quadric becomes a hyperboloid of two sheets. At $\lambda = A$, the quadric becomes the whole zy plane. No real quadric exists for $\lambda > A$.

be removed simply by assigning some signature labels to these points.

To find (x^2,y^2,z^2) in terms of our new coordinates $(\lambda_1,\lambda_2,\lambda_3)$, we first obtain a cubic polynomial in λ by multiplying the left-hand side of (4.2) by $(A-\lambda)\times(B-\lambda)(C-\lambda)$, giving

$$\begin{aligned} F(\lambda) &\equiv (\lambda - A)(\lambda - B)(\lambda - C) + x^2(\lambda - B)(\lambda - C) \\ &\quad + y^2(\lambda - A)(\lambda - C) + z^2(\lambda - A)(\lambda - B) \\ &= (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3). \end{aligned} \tag{4.3}$$

The last expression follows from the fact that the root of $F(\lambda)$ are λ_1, λ_2, and λ_3. We can evaluate x^2 in (4.3) by

$$x^2 = \lim_{\lambda \to A} [F(\lambda)/(\lambda - B)(\lambda - C)]$$

giving

$$x^2 = \frac{(A-\lambda_1)(A-\lambda_2)(A-\lambda_3)}{(A-B)(A-C)}. \tag{4.4a}$$

Similarly, we have

$$y^2 = \frac{(B-\lambda_1)(B-\lambda_2)(B-\lambda_3)}{(B-A)(B-C)}, \tag{4.4b}$$

$$z^2 = \frac{(C-\lambda_1)(C-\lambda_2)(C-\lambda_3)}{(C-A)(C-B)}. \tag{4.4c}$$

We can determine the range of λ's as

$$-\infty < \lambda_1 < C < \lambda_2 < B < \lambda_3 < A. \tag{4.5}$$

From (4.4) and after some algebra, we can work out the first fundamental form as

$$\begin{aligned} ds^2 &= (dx)^2 + (dy)^2 + (dz)^2 \\ &= g_1 d\lambda_1^2 + g_2 d\lambda_2^2 + g_3 d\lambda_3^2, \end{aligned} \tag{4.6}$$

where

$$g_1 = \frac{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)}{4(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}, \tag{4.7a}$$

$$g_2 = \frac{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2)}{4(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)}, \tag{4.7b}$$

$$g_3 = \frac{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)}{4(A-\lambda_3)(B-\lambda_3)(C-\lambda_3)}. \tag{4.7c}$$

It is pleasing to see that the λ's do form an orthogonal basis as promised.

In the new coordinate system, the Lagrangian and Hamiltonian for a free particle moving inside an ellipsoid are

$$\begin{aligned} L &= \tfrac{1}{2}(\dot{x} + \dot{y} + \dot{z}) - V(x,y,z) \\ &= \tfrac{1}{2}(g_1\dot{\lambda}_1^2 + g_2\dot{\lambda}_2^2 + g_3\dot{\lambda}_3^2) - V(\lambda_1,\lambda_2,\lambda_3) \end{aligned} \tag{4.8}$$

and

$$\begin{aligned} H &= \tfrac{1}{2}(p_1^2/g_1 + p_2^2/g_2 + p_3^2/g_3) + V(\lambda_1,\lambda_2,\lambda_3) \\ &= 2\left[\frac{(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)p_1^2}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}\right. \\ &\quad \left. + \text{cyclic perms. of subscripts}\right] + V(\lambda_1,\lambda_2,\lambda_3), \end{aligned} \tag{4.9}$$

respectively, with

$$p_1 = g_1\dot{\lambda}_1, \quad p_2 = g_2\dot{\lambda}_2, \quad p_3 = g_2\dot{\lambda}_3.$$

The potential V is chosen to be zero inside the ellipsoid $\lambda_1 = 0$ and ∞ outside the same ellipsoid. The Hamilton–Jacobi equation is

$$\begin{aligned} &\frac{2(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}\left(\frac{\partial W}{\partial \lambda_1}\right)^2 \\ &\quad + \text{cyclic perms. of } \lambda\text{'s} + V(\lambda_1,\lambda_2,\lambda_3) \\ &\quad = \alpha, \quad \text{the energy.} \end{aligned} \tag{4.10}$$

Just as in the two-dimensional case, we need to replace this singular $V(\lambda_1,\lambda_2,\lambda_3)$ by a smooth function which reproduces V in the appropriate limit. A possible choice of such a function is

$$V(\lambda_1,\lambda_2,\lambda_3) = V_\epsilon(\lambda_1)/[(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)], \tag{4.11}$$

where

$$\begin{aligned} V_\epsilon(\lambda) &= 0, \quad \text{for } \lambda_1 > \epsilon, \\ &= V_0/\epsilon, V_0 > 0, \quad \text{for } \lambda_1 < -\epsilon, \end{aligned} \tag{4.12}$$

and V_ϵ is continuous in region $(-\epsilon,\epsilon)$. As $\epsilon \to 0$, we recover $V(x,y,z)$. We can solve the Hamilton–Jacobi equation by choosing

$$W = W_1(\lambda_1) + W_2(\lambda_2) + W_3(\lambda_3) \tag{4.13}$$

and by introducing the separation constants α', α'' via

$$2(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)\left(\frac{dW_1(\lambda_1)}{d\lambda_1}\right)^2+V_\epsilon(\lambda_1)$$
$$=\alpha(\lambda_1-\alpha')(\lambda_1-\alpha''), \tag{4.14a}$$
$$2(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)\left(\frac{dW_2(\lambda_2)}{d\lambda_2}\right)^2$$
$$=\alpha(\lambda_2-\alpha')(\lambda_2-\alpha''), \tag{4.14b}$$
$$2(A-\lambda_3)(B-\lambda_3)(C-\lambda_3)\left(\frac{dW_3(\lambda_3)}{d\lambda_3}\right)^2$$
$$=\alpha(\lambda_3-\alpha')(\lambda_3-\alpha''). \tag{4.14c}$$

By the use of the identity

$$\frac{(\lambda_1-\alpha')(\lambda_1-\alpha'')}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)}+\frac{(\lambda_2-\alpha')(\lambda_2-\alpha'')}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)}$$
$$+\frac{(\lambda_3-\alpha')(\lambda_3-\alpha'')}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)}=1, \tag{4.15}$$

which is valid for all α', α'', we can show that the Hamilton–Jacobi equation (4.10) is indeed separated and satisfied. From (4.14), we obtain the momentum with respect to the λ_i as

$$p_1=\frac{dW_1}{d\lambda_1}=\left(\frac{\alpha(\lambda_1-\alpha')(\lambda_1-\alpha'')-V_\epsilon(\lambda_1)}{2(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}\right)^{1/2}, \tag{4.16a}$$

$$p_2=\frac{dW_2}{d\lambda_2}=\left(\frac{\alpha(\lambda_2-\alpha')(\lambda_2-\alpha'')}{2(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)}\right)^{1/2}, \tag{4.16b}$$

$$p_3=\frac{dW_3}{d\lambda_3}=\left(\frac{\alpha(\lambda_3-\alpha')(\lambda_3-\alpha'')}{2(A-\lambda_3)(B-\lambda_3)(C-\lambda_3)}\right)^{1/2}. \tag{4.16c}$$

In (4.16) and hereafter, the signs of the square roots are chosen such that p_i and $\dot\lambda_i$ shall have the same sign. From (4.16), we can integrate to obtain

$$W_1(\lambda_1)=\int d\lambda_1\left(\frac{\alpha(\lambda_1-\alpha')(\lambda_1-\alpha'')-V_\epsilon(\lambda_1)}{2(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}\right)^{1/2}, \tag{4.17}$$

$$W_2(\lambda_2)=\int d\lambda_2\left(\frac{\alpha(\lambda_2-\alpha')(\lambda_2-\alpha'')}{2(A-\lambda_2)(B-\lambda_2)(C-\lambda_2)}\right)^{1/2}, \tag{4.18}$$

and a similar equation for $W_3(\lambda_3)$. The presence of $V_\epsilon(\lambda_1)$ implies that the particle bounces at $\lambda_1=0$. We can suppress the V_ϵ term in (4.17) after we impose an integration limit at $\lambda_1=0$. The quantities α, α', α'' are identified as the energy and conserved momenta. The equations of motion in terms of the new coordinates are

$$Q\equiv\frac{\partial W}{\partial\alpha}=\frac{1}{\sqrt{8\alpha}}\left[\int d\lambda_1\left(\frac{(\lambda_1-\alpha')(\lambda_1-\alpha'')}{(A-\lambda_1)(B-\lambda_1)(C-\lambda_1)}\right)^{1/2}\right.$$
$$\left.+\text{cyclic perms. of }\lambda\text{'s}\right]$$
$$=t+\beta,\quad \beta=\text{const}, \tag{4.19a}$$

$$Q'\equiv\frac{\partial W}{\partial\alpha'}=\beta',\quad \text{const}, \tag{4.19b}$$

$$Q''\equiv\frac{\partial W}{\partial\alpha''}=\beta'',\quad \text{const}, \tag{4.19c}$$

where the explicit integral representation of Q' and Q'' will be given in (4.21) below after we fix the order of α's. Equations (4.19b) and (4.19c) describe the particle trajectory. By evaluating Q' and Q'' at each bounce ($\lambda_1=0$) we can find out how λ_2 and λ_3 vary at subsequent bounces. We can use (4.19a) to evaluate the time between bounces.

As we trace a particle along a straight-line segment and its extension from $-\infty$ to ∞, the λ's change continuously and return to their original values. Thus each of the λ's must reach at least a maximum and a minimum. We need at least six turning points, and they occur at 0, A, B, C, α', and α''. To achieve a real motion (i.e., all p's are real), we require, in addition to

$$\lambda_1<C<\lambda_2<B<\lambda_3<A, \tag{4.5'}$$

the following restriction on α' and α'':

$$\lambda_1<\alpha'<\lambda_2<\alpha''<\lambda_3. \tag{4.20}$$

In (4.20), we have chosen without losing any generality $\alpha'<\alpha''$. How a particular λ_i behaves depends on the signs of $\alpha'-C$ and $\alpha''-B$. This leads to the following four possible ranges of λ's:

(1) $0<\lambda_1<\alpha'<C<\lambda_2<\alpha''<B<\lambda_3<A$,

(2) $0<\lambda_1<\alpha'<C<\lambda_2<B<\alpha''<\lambda_3<A$,

(3) $0<\lambda_1<C<\alpha'<\lambda_2<\alpha''<B<\lambda_3<A$,

(4) $0<\lambda_1<C<\alpha'<\lambda_2<B<\alpha''<\lambda_3<A$.

In case (1), we have one ellipsoid ($\lambda_1=\alpha'$) and one hyperboloid with one sheet ($\lambda_2=\alpha''$) as caustic surfaces. The turning points at A, B, and C are associated with the crossings of the xy, yz, and zx planes and have no physical significance. The turning point $\lambda_1=0$ corresponds to the bounce at the outer ellipsoid. Only the turning points at α',α'' are associated with the vanishing of p_i, and lead to the tangent of the trajectory to the quadric surfaces. Similarly, we have in case (2) one ellipsoid ($\lambda_1=\alpha'$) and one hyperboloid with two sheets ($\lambda_3=\alpha''$) as caustic surfaces. In case (3), we have two hyperboloids with one sheet ($\lambda_2=\alpha'$, $\lambda_2=\alpha''$) and in case (4) we have a hyperboloid with one sheet and a hyperboloid with two sheets ($\lambda_2=\alpha'$, $\lambda_3=\alpha''$) as caustic surfaces. Since the separation constants α' and α'' are constants of motion, all line segments of a given trajectory share the same caustic surfaces, as promised.

It is known in two dimensions that trajectories showing the same caustic conic always have the same period. In the following, we shall generalize this result to three dimensions. We first evaluate the derivatives of W's with respect to α's explicitly in (4.19b) and (4.19c). Depending on whether $\alpha'>\lambda$ or $\alpha'<\lambda$, the differentiation of $\sqrt{\text{const}(\alpha'-\lambda)}$ gives either a positive or a negative factor. The net result is

$$\int\frac{p_1\,d\lambda_1}{2(\alpha'-\lambda_1)}-\int\frac{p_2\,d\lambda_2}{2(\lambda_2-\alpha')}-\int\frac{p_3\,d\lambda_3}{2(\lambda_3-\alpha')}$$
$$\equiv f_1(\lambda_1)-f_2(\lambda_2)-f_3(\lambda_3)=\beta',\quad \text{const}, \tag{4.21a}$$

$$\int\frac{p_2 d\lambda_1}{2(\alpha''-\lambda_1)}+\int\frac{p_2 d\lambda_2}{2(\alpha''-\lambda_2)}-\int\frac{p_3\,d\lambda_3}{2(\lambda_3-\alpha'')}$$
$$\equiv g_1(\lambda_1)+g_2(\lambda_2)-g_3(\lambda_3)=\beta'',\quad \text{const}, \tag{4.21b}$$

where f's and g's are monotonically increasing functions of

their arguments. As the λ 's goes through a complete period, the f's and g's increase by fixed amounts independent of their starting points,

$$\delta f_i(\lambda_i)|_{\text{one period}} \equiv F_i, \tag{4.22a}$$

$$\delta g_i(\lambda_i)|_{\text{one period}} \equiv G_i. \tag{4.22b}$$

If the trajectory has a period, the λ 's have to return to their original values after completing some integer numbers of cycles. We denote these integers by (l_1,l_2,l_3), and shall refer to them as winding numbers. From (4.21), we have

$$l_1F_1 - l_2F_2 - l_3F_3 = 0, \tag{4.23a}$$

$$l_1G_1 + l_2G_2 - l_3G_3 = 0. \tag{4.23b}$$

Next, we consider an arbitrary starting point but with the same separation constants α',α''. This implies automatically that it shares the same caustics. We follow the trajectory until the λ_1 variable has completed l_1 cycles. Then, Eq. (4.21) becomes

$$l_1F_1 - \delta f_2 - \delta f_3 = 0, \tag{4.24a}$$

$$l_1G_1 + \delta g_2 - \delta g_3 = 0. \tag{4.24b}$$

Subtracting (4.24) from (4.23), we have

$$(\delta f_2 - l_2F_2) + (\delta f_3 - l_3F_3) = 0, \tag{4.25a}$$

$$(\delta g_2 - l_2G_2) - (\delta_3 - l_3G_3) = 0, \tag{4.25b}$$

which implies

$$(\delta f_2 - l_2F_2)(\delta g_2 - l_2G_2) = -(\delta f_3 - l_3F_3)(\delta g_3 - l_3G_3). \tag{4.26}$$

It is easy to see that the left-hand side of (4.26) is either zero or positive. It is zero only if λ_2 completes exactly l_2 cycles. Similarly, the right-hand side of (4.26) is either zero or negative. It is zero only if λ_3 completes exactly l_3 cycles. Equality (4.26) thus implies that λ_2 and λ_3 complete exactly l_2 and l_3 cycles, respectively. Hence the trajectory is periodic with the same winding numbers (l_1,l_2,l_3). From (4.19a), we also know that these orbits also have the same period in time, or equivalently, in length.

V. PONCELET'S THEOREM

Let us return to two dimensions. Because of the theorem of Sinai (Sec. II), the specification of the particle trajectory is redundant. We have so far specified it by requiring pairs of successive chords to satisfy the law of reflection at the outer conic; then the theorem tells us that each chord is tangent to an inner conic confocal with the outer one. In the present section we shall suppose that the two conics are given, and that the trajectory consists of chords all tangent to the inner conic, and each intersecting its successor at the outer one. Then the law of reflection appears as a consequence, provided that the conics are confocal.

We wish here to concentrate on periodic trajectories. We note that (3.33) is both a necessary and a sufficient condition for the trajectory to close after p bounces, no matter where it starts. Therefore the analysis of Sec. III leads to the following conclusion.

If the two conics are chosen so that some trajectory closes after p bounces, then every trajectory closes after p bounces.

A periodic trajectory may be regarded as a polygon circumscribed about the inner conic and inscribed in the other. Therefore we may restate our conclusion: If a polygon is inscribed in one conic and circumscribed about another, then every point on the former conic is a vertex of some polygon (with the same number of sides) having the same property. (In projective geometry over a field not algebraically closed, one must stipulate that the starting point lies on a tangent to the inner conic.)

The reader may object that we should have specified that the conics be confocal. But the result as just stated involves only entities and properties (straight lines, conics, incidence, tangency) that are preserved by projective transformation; and by counting degrees of freedom one may verify that *any* pair of conics in the same plane may be projected into a pair of confocal conics. Therefore the above statement holds for any two conics; they need not be confocal or even concentric. (Of course, if they are not confocal, the law of reflection is not obeyed.)

This projective theorem about a polygon and two conics has long been known under the name of Poncelet's theorem. Poncelet's proof [5] was not strictly geometrical inasmuch as it made use of the point where a line is stationary under a given infinitesimal algebraic displacement. However, this concept can probably be made finite by means of the modern notion of differential forms.

The theorem was later discussed by Jacobi[6] who analyzed the case of two arbitrarily placed circles (projectively equivalent to the general case) and recognized certain relationships between successive chords as identical to the addition formulas for elliptic functions. He was thus led to an understanding of the role played in this theorem by elliptic integrals and elliptic functions, and obtained equations equivalent to our (3.31)–(3.33).

Although (3.33) involves a transcendental function, it can be converted for any integer p to an algebraic relation between b'/b and k, by applying Jacobi's addition formula $p-1$ times and using the fact that $\mathrm{cn}(2qK) = (-1)^q$. In 1858 Cayley gave an elegant form of this relation for arbitrary p.[5] Cayley's work stimulated a concise paper by Hart, in which Poncelet's theorem was proved for the first time entirely by finite operations—specifically, the constructions of projective geometry.[8]

To understand the principle behind Hart's construction, we may reexamine Eq. (3.31). As was quite clear to Jacobi, the elliptical integral

$$M(A,B) \equiv \int_{\theta_A}^{\theta_B} \frac{d\theta}{\sqrt{1-k^2\cos^2\theta}}$$

may be viewed as a metric along the outer conic, and Poncelet's theorem follows from the fact that all chords tangent to the inner conic subtend the same arc "length" on the outer conic according to this metric. For if we compare two trajectories ABC...Z and A'B'C'...Z' then $M(\mathrm{A,B}) = M(\mathrm{A',B'})$ since AB and A'B' are both tangent to the inner conic. Since the metric is obviously additive, it follows that $M(\mathrm{A,A'}) = M(\mathrm{B,B'})$. Continuing the argument we find $M(\mathrm{A,A'}) = M(\mathrm{B,B'}) = \cdots = M(\mathrm{Z,Z'})$; therefore if Z = A then Z' = A'. So if one trajectory closes the other does too.

Let us denote the outer conic by q_0 and the inner by q_1. Hart's implicit idea was that if AA', BB', ... all subtend the same length we should be able to find a third conic q_1' to which they are all tangent. The statement that $M(\mathrm{A},\mathrm{B}) = M(\mathrm{A}',\mathrm{B}')$ implies $M(\mathrm{A},\mathrm{A}') = M(\mathrm{B},\mathrm{B}')$ would then be replaced by the statement that if AB and A'B' are tangent to the same q_1, then AA' and BB' will be tangent to the same q_1'. In this way the proof could be carried through by purely geometric means, without referring to a metric or introducing elliptic integrals at all.

In choosing q_1', we must note that the modulus k appears in the definition of the metric. In Sec. II we observed that k was the eccentricity of q_1, but that holds only when q_0 and q_1 are confocal. In general, k is the eccentricity of the conic to which q_1 is carried by the projective transformation that renders it confocal with q_0. Thus $k = k(q_0,q_1)$ is a projective invariant of the two conics. Now, clearly, Hart's idea will not work unless $k(q_0,q_1') = k(q_0,q_1)$. For if q_1' defines the wrong metric on q_0, its tangents will not subtend "equal" arcs according to the metric defined by q_1, and it will not be tangent both to AA' and BB', etc. Therefore it is necessary to define a family F of conics such that $k(q_0,q) = k(q_0,q_1)$ for all $q \in F$. Let the outer and inner conics be the loci of the equations $Q_0 = 0$, $Q_1 = 0$, respectively. Let F be the family of conics $Q = 0$, where Q is a linear combination of Q_0 and Q_1. (In projective geometry F is called a *pencil* of conics.) Then, as Jacobi had pointed out, $k(q_0,q)$ is the same for all $q \in F$. (See the Appendix.) The pencil F is thus the basic tool in Hart's construction. (The importance of this pencil had already been stressed by Poncelet as well as Jacobi and Cayley.[5,6,12])

The actual proof, of course, makes no mention of the modulus k. It is based on the following (slightly restated) lemma.

Let the line l_0 intersect the conic q at P_1, P_3, and the conic q' at P_2, P_4. Let $l_1,\ldots,l_4$ be the tangents to the respective conics at these points. Let l_i, l_j intersect at P_{ij}. Then the pencil of conics generated by q and q' has a member q_0 passing through the four points P_{12}, P_{23}, P_{34}, P_{41} [Fig. 5(a)].

The proof of Hart's lemma is as follows: for $i = 0,\ldots,4$, let l_i be the locus of $L_i = 0$; then q is the locus of $L_0^2 = aL_1L_3$ for some a. (This can be shown by projecting l_3 to ∞ and P_1 to the origin, and rendering l_1 perpendicular to l_0; then q is a parabola with vertex at the origin.) Likewise q' is the locus of $L_0^2 = a'L_2L_4$; therefore the conic $aL_1L_3 = a'L_2L_4$ lies in the same pencil as q and q'. This conic clearly passes through the four points specified since L_1L_3 and L_2L_4 both vanish at each point.

The proof can be run backward to establish a converse: if q_0 and q are given together with all points and lines in the figure, the existence of q' follows. For q_0 must satisfy an equation of the form $L_1L_3 = bL_2L_4$, since b can be chosen so that this equation holds at a fifth point of q_0, and five points determine a conic. Eliminating the term L_1L_3 between this and the equation for q, we obtain the linear combination $L_0^2 = abL_2L_4$ which describes a conic of the pencil (q_0q). This conic is tangent to l_2 and l_4 as desired.

Hart, like Poncelet, actually proved a generalized theorem on the triangle, in which each side is tangent to its own fixed conic of the pencil. From this one may derive the theorem for polygons by triangulating the polygon. We shall argue directly to the polygon.

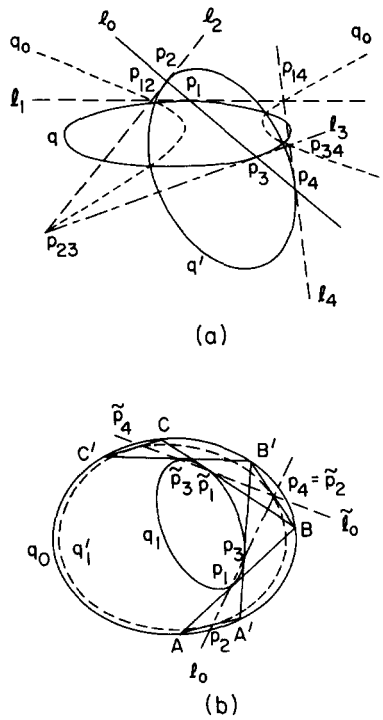


FIG. 5. (a) Hart's lemma. The conics q, q' have been shown as fully intersecting (four unlabeled points) although our usual picture for Poncelet's theorem shows them as nonintersecting in real space. The fact that q, q', q_0 belong to the same pencil is thus exhibited by the fact that they all pass through these four points, as well as the ones mentioned in the lemma. (b) Hart's construction for Poncelet's theorem. All the chords AA', BB',...,ZZ' will be tangent to q_1'.

Assume we have two trajectories $\mathrm{AB}\cdots\mathrm{Z}$ and $\mathrm{A'B'}\cdots\mathrm{Z'}$ [Fig. 5(b)]. Let q_1 be tangent to AB at P_1 and to A'B' at P_3. We draw the line l_0 through P_1 and P_3, intersecting AA' at P_2 and BB' at P_4. Then if we denote AB, AA', A'B', BB' by $l_1,\ldots,l_4$, respectively, the converse to the lemma shows that the pencil (q_0,q_1) contains a conic q_1' tangent both to AA' at P_2 and to BB' at P_4.

Now we work forward from BB', renaming the points and lines: thus $\tilde{l}_1 = \mathrm{BC}$ tangent to q_1 at $\tilde{P}_1$, and $\tilde{l}_2 = \mathrm{BB'}$ tangent to q_1' at $\tilde{P}_2$ (formerly called P_4). We draw $\tilde{l}_0$ through $\tilde{P}_1$ and $\tilde{P}_2$, intersecting q_1 again at $\tilde{P}_3$ and q_1' again at $\tilde{P}_4$, and denote $\tilde{l}_3$, $\tilde{l}_4$ the tangents to q_1, q_1' at $\tilde{P}_3$ and $\tilde{P}_4$, respectively.

The lemma now tells us that $\tilde{l}_3$ and $\tilde{l}_2$ intersect at a point $\tilde{P}_{23}$ which lies on q_0. For it must lie on a conic of the pencil (q_1,q_1') which also passes through $\tilde{P}_{12} = \mathrm{B}$, and only one conic of the pencil can pass through B. Therefore $\tilde{P}_{23} = \mathrm{B'}$, and $\tilde{l}_3 = \mathrm{B'C'}$. Likewise the lemma shows that $\tilde{P}_{14} = \mathrm{C}$, $\tilde{P}_3 = \mathrm{C'}$, and $\tilde{l}_4 = \mathrm{CC'}$; hence q_1' is also tangent to CC'.

Proceeding by repetition one establishes that q_1' is tangent to ZZ', so that if $\mathrm{Z} = \mathrm{A}$ then $\mathrm{Z'} = \mathrm{A'}$. The full proof contains an additional argument to show that ZZ' is the right tangent from Z to q_1 (the same as AA') and not the other one. Lebesgue, who gave essentially the same proof in a long memoir in 1921, explores this point in greater detail.[8] We

have also omitted the reasoning that excludes various degenerate situations in the proof as given above. However, this sketch should make it clear that nowhere is there any use of transcendental functions or of the notions of continuity and differentiation.

The existence of a proof of this kind is actually implicit in Jacobi's method, which has been put in a modern form by Griffiths and Harris.[10] They show that the set of all entities consisting of a point on q_0 and one of its tangents to q_1 constitutes an Abelian variety and is therefore subject to a family of algebraic transformations which guarantee the existence of a finite proof of Poncelet's theorem such as the one given explicitly by Hart. Although these transformations correspond to the addition laws for elliptic functions, they are algebraic consequences of the geometric premises and do not depend on differential arguments.

Because of Hart's proof or the "Abelian variety" argument, Poncelet's theorem must hold for projective geometries over arbitrary fields, even discrete ones; we have verified this in a number of cases. However, when we proceed to three dimensions, we encounter a somewhat mysterious situation.

VI. PONCELET'S THEOREM IN THREE DIMENSIONS

We are not concerned here with the three-dimensional version given by Griffiths and Harris, which involves a polyhedron and two quadric surfaces, but with a quite different one to be extracted from Sec. IV, involving three quadrics and a skew polygon. The analysis of that section shows that there is an analog to the theorem of Sinai, which states that if a trajectory is defined as a sequence of chords of an "outer" quadric q_0 linked successively by the law of reflection, then there exist two "inner" quadrics q_1 and q_2, both tangent to every chord of the trajectory; moreover q_0, q_1, q_2 are all "confocal" in the sense that they are surfaces of the same system of ellipsoidal coordinates. It also yields an analog to Poncelet's theorem, which says that if such a trajectory closes after p steps, then every point on q_0 belongs to a trajectory which also closes after p steps and whose rays are tangent to the same pair q_1, q_2. But we immediately encounter difficulties in putting this analog in projective form.

First of all, it is no longer true that any three quadrics in space can be projected into a confocal triplet in the sense above. Suppose the equation of a quadric q is written $Q = 0$, where

$$Q = \sum_{ij}^{3} q_{ij} x_i x_j + \sum_{i}^{3} q_{0i} x_i + q_{00}, \tag{6.1}$$

and let $\|Q\|$ stand for the 4×4 matrix $\|q_{ij}\|$; then if q is the locus of (4.2) we have

$$\|Q\|^{-1} = \left|\left| \begin{matrix} -1 & & & \\ & A & & \\ & & B & \\ & & & C \end{matrix} \right|\right| \left|\left| -\lambda \right|\right| \left|\left| \begin{matrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{matrix} \right|\right|. \tag{6.2}$$

But the three quadrics q_0, q_1, q_2 of Sec. IV are all loci of equations obtained from (4.2) by varying λ; therefore the matrices $\|Q_0\|^{-1}$, $\|Q_1\|^{-1}$, $\|Q_2\|^{-1}$ are linearly dependent. This is a projectively invariant statement which can be expressed by saying that q_0, q_1, q_2 belong to the same *range* of quadrics, a range being the dual of a pencil. Therefore our projective theorem will apply only to three quadrics belonging to the same range.

Second, the theorem of Sinai in two dimensions had a converse: if two tangents were drawn to q_1 from a point on q_0, they would satisfy the reflection law at q_0 provided the two conics were confocal. But in three dimensions there exist four lines l_a, l_b, l_c, l_d that can be drawn from a given point on q_0 so as to be tangent to both q_1 and q_2. These four can be paired in three ways $(ab)(cd)$, $(ac)(bd)$, $(ad)(bc)$ and only one of these pairings will satisfy the law of reflection. Moreover, the four lines $l_a,\ldots,l_d$ are determined by the roots of a quartic which in general has the maximal Galois group S_4; thus there seems to be no algebraic way to choose the right pairing.

To overcome this difficulty one must use one's knowledge of the quadric q_0, which in ellipsoidal coordinates is the locus of $\lambda_1 = \text{const}$. Through a given point on q_0 there must pass two other quadrics $\bar{q}, \bar{\bar{q}}$ in the same range as q_0, q_1, q_2, defined by $\lambda_2 = \text{const}$ and $\lambda_3 = \text{const}$. Each of the three pairings of $l_a,\ldots,l_d$ will satisfy the law of reflection on one of the three $q_0, \bar{q}, \bar{\bar{q}}$. The "right" pairing has its planes of reflection normal to q_0, so that they intersect in the normal to q_0. Now, the concept "normal" is not projectively invariant, but tangent is, and the normal to q_0 is also tangent to $\bar{q}$ and $\bar{\bar{q}}$. Hence, we may now state our three-dimensional version of Poncelet's theorem.

Suppose that three quadrics, q_0, q_1, q_2 belong to the same range R. Suppose that a polygon exists in three-space, inscribed in q_0 and with each side tangent to both q_1 and q_2, such that any two successive sides, intersecting at P on q_0, determine a plane that contains the common tangent at P to the other two quadrics of R that pass through P. Then every point of q_0 lying on a common tangent to q_1, q_2 is a vertex of a polygon of the same order possessing the same properties.

The arguments of Sec. IV suffice to establish this theorem in real space. Note that in real space the three quadrics $q_0, \bar{q}, \bar{\bar{q}}$ are always distinct; in more general projective spaces one may encounter degeneracies requiring a special prescription for pairing the four tangents. Putting this problem aside for the moment, we now ask whether a finite (algebraic or geometric) proof can be found for our three-dimensional Poncelet theorem. Again we find new obstacles in our way.

Poncelet's proof, even if rendered algebraic, identifies the point of tangency to q_1 with the stationary point of a moving line. But in the three-dimensional configuration neither point of tangency (to q_1 or q_2) is the stationary point.

If we think of generalizing Hart's proof, we would need an auxiliary pair of quadrics, q_1', q_2'. To make the proof work, this pair must define the same two-dimensional metric system on q_0 as the pair q_1, q_2; in other words the integrands of (4.17) and (4.18) must have the same form when q_1, q_2 are replaced by q_1', q_2' and the Cartesian axes are rescaled to make q_0, q_1', q_2' confocal. But of course it is also necessary that q_0, q_1', q_2' belong to a single range; otherwise they cannot

be rescaled to be confocal at all. As shown in the Appendix, there is only *one* free parameter (not two) in choosing q_1', q_2'. Therefore if Hart's construction is to be applied to a pair of trajectories AB... and A'B'..., then with A given the choices of A' that can be handled fall into a one-parameter family. But such a family cannot in general cover the two-dimensional surface of q_0. So it seems that Hart's method will not suffice to prove the full three-dimensional theorem.

Alternatively one may try to generalize the approach of Griffiths and Harris, who reformulate Jacobi's treatment for two dimensions in terms of the algebraic variety in which each point is specified by choosing a point on the outer conic together with a tangent to the inner conic, with the restriction that the latter be incident on the former. Poncelet's theorem follows from the existence of an Abelian group of holomorphic translations on this variety.

We are not sufficiently grounded in pure mathematics to understand whether this argument is genuinely algebraic, as it appears superficially to depend on the properties of the complex field. However, we can elucidate the parametrization of the corresponding variety in the three-dimensional theorem.

In two dimensions one may parametrize the variety E (Ref. 10) by x, y, and p_2 with the restrictions

$$x^2/(A-\lambda_1)+y^2/(B-\lambda_1)=1$$

and

$$p_2^2=(\alpha'-\lambda_2)/[(A-\lambda_2)(B-\lambda_2)]$$

(we have omitted the inessential constant $\alpha/2$) where λ_1 is a fixed constant, and λ_2 is defined as the other root of the quadratic $x^2/(A-\lambda)+y^2/(B-\lambda)=1$ after $\lambda-\lambda_1$ is factored out. Thus x and y determine a point on the outer conic, and the sign of p_2 tells which tangent to take.

In parametrizing the (two-dimensional) variety in the three-dimensional case, one must be careful not to include too much information. If $\lambda_2, \lambda_3, p_2, p_3$ are all known, then we can not only identify the tangent we have arrived by as well as the one we shall leave by, but we can also *distinguish between the other two tangents.* This last bit of information is superfluous to the problem.

Therefore the appropriate parametrization includes x, y, z, $p_2 p_3$, and p_2+p_3 with appropriate restrictions. Note that λ_2, λ_3 are now roots of a cubic of which one root λ_1 is fixed; therefore $\lambda_2\lambda_3$ and $\lambda_2+\lambda_3$ are rational functions of x,y,z. It follows that $p_2^2 p_3^2, p_2^2+p_3^2$ are also rational in x, y, z. The sign of $p_2 p_3$ tells us which of the two legitimate trajectories (plane of incidence must contain a line tangent to both $\lambda_2=\text{const}$ and $\lambda_3=\text{const}$) we are on; then the sign of p_2+p_3 tells which direction to choose along that trajectory. But no direction is specified along the other legitimate trajectory at the point; this would be superfluous.

The differential forms

$$[dQ']=P_2\,d\lambda_2/(\alpha'-\lambda_2)+p_3\,d\lambda_3/(\alpha'-\lambda_3)$$

and

$$[dQ'']=p_2\,d\lambda_2/(\alpha''-\lambda_2)+p_3\,d\lambda_3/(\alpha''-\lambda_3)$$

can be expressed in terms of the above parameters since they are symmetric in the exchange $2\leftrightarrow 3$. Since these forms are linearly independent in terms of those parameters, it appears (if we only understood the rules of argument) that a finite proof might be constructed along these lines. But we leave this to others more qualified. (We are grateful to Grayson for some expert advice on these mysterious matters.)

VII. ITERATIVE CONSTRUCTION

We can investigate our model from a different approach. Instead of solving the trajectory of the Hamilton–Jacobi method, we can study the trajectory by following the slope of the ray and the reflection point on the outer quadric. This is a simple generalization of Birkhoff's approach.[1] We denote a ray by a set of numbers (m_1,m_2,m_3), proportional to the direction cosines. We denote the point of reflection by (x,y,z) and its normal by (n_1,n_2,n_3). To within a normalization factor, we can write

$$n_1=x/A,\quad n_2=y/B,\quad n_3=z/C. \tag{7.1}$$

We denote the slope of the subsequent ray by $\mathbf{m}'$, and the normal of the subsequent reflection point by $\mathbf{n}'$ [see Fig. 6(a)]. It is straightforward to show that with suitable normalization

$$\mathbf{m}'=\mathbf{m}-2\mathbf{n}(m\cdot n)/n^2, \tag{7.2}$$

$$\mathbf{n}'=\mathbf{n}-2\mathbf{Q}\mathbf{m}'(m'\cdot n)/(m'Qm'), \tag{7.3}$$

where Q, $\mathbf{Qm}$, and (mQm) are

$$Q\equiv\begin{bmatrix}1/A & 0 & 0\\ 0 & 1/B & 0\\ 0 & 0 & 1/C\end{bmatrix}, \tag{7.4}$$

$$\mathbf{Qm}=(m_1/A,m_2/B,m_3/C), \tag{7.5}$$

$$(mQm)=m_1^2/A+m_2^2/B+m_3^2/C. \tag{7.6}$$

Repeated applications of (7.2) and (7.3) generate all the subsequent m's and n's.

Note that Eqs. (7.2) and (7.3) are homogeneous in both m and n. Thus the absolute scales of m and n can be arbitrary in the iteration equations. From (7.2) and (7.3), we can construct easily three invariants

$$m^2=m_1^2+m_2^2+m_3^2, \tag{7.7a}$$

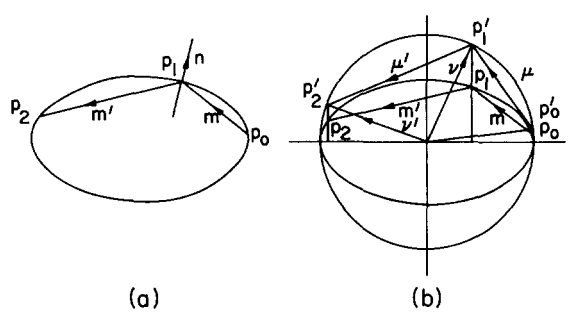


FIG. 6. For simplicity, we have made these drawings in two dimensions. (a) The propagation and reflection of a particle inside an ellipse (or an ellipsoid). Note that $\mathbf{m}$ and $\mathbf{m}'$ make an equal angle to the normal $\mathbf{n}$. (b) A geometrical construction of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ from $\mathbf{m}$ and $\mathbf{n}$. Using the outer auxiliary circle (or sphere in three dimensions), we can relate $\boldsymbol{\nu}$, $\boldsymbol{\nu}'$ to the points of reflections on the ellipse (ellipsoid) straightforwardly. Vector $\boldsymbol{\mu}'$ makes an equal angle to $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$.

$$m\cdot n = m_1n_1 + m_2n_2 + m_3n_3, \tag{7.7b}$$

and

$$(nAn) = An_1^2 + Bn_2^2 + Cn_3^2. \tag{7.7c}$$

Using the definition of $\mathbf{n}$ in (7.1), we find that the condition $(nAn) = 1$ leads to the equation of the outer ellipsoid.

There is a duality transformation which interchanges the role of m and n. We define

$$\boldsymbol{\mu} = (\sqrt{Q_1}m_1, \sqrt{Q_2}m_2, \sqrt{Q_3}m_3), \tag{7.8}$$

$$\boldsymbol{\nu} = (n_1/\sqrt{Q_1}, n_2/\sqrt{Q_2}, n_3/\sqrt{Q_3}), \tag{7.9}$$

where Q_1, Q_2, and Q_3 are diagonal elements of Q in (7.4). Then

$$\mu\cdot\nu = m\cdot n, \tag{7.10}$$

$$\mu^2 = (mQm), \tag{7.11}$$

and

$$(\nu Q\nu) = n^2. \tag{7.12}$$

In terms of μ and ν, we have

$$\boldsymbol{\mu}' = \boldsymbol{\mu} - 2\mathbf{Q}\boldsymbol{\nu}(\mu\cdot\nu)/(\nu Q\nu), \tag{7.13}$$

$$\boldsymbol{\nu}' = \boldsymbol{\nu} - 2\boldsymbol{\mu}'(\mu'\cdot\nu)/\mu'^2, \tag{7.14}$$

which are indeed the same as (7.2) and (7.3) with m, n interchanged. This scale transformation makes the outer quadric surface into a sphere, as shown in Fig. 6(b). The new direction cosines of the ray and the normal, μ and ν, transform according to (7.13) and (7.14). In this new frame, $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$ are pointing along the radial directions. It is easy to see that they are indeed related by a reflection around $\boldsymbol{\mu}'$.

We shall now give an independent proof of Sinai's theorem. We begin with a ray with slope m and passing through $\mathbf{x}_0 = (An_1, Bn_2, Cn_3)$. The equation for the ray is

$$\mathbf{x} = \mathbf{x}_0 + k\mathbf{m}. \tag{7.15}$$

The condition that this ray be tangent to a confocal quadric

$$\frac{x^2}{A-\lambda} + \frac{y^2}{B-\lambda} + \frac{z^2}{C-\lambda} = 1 \tag{7.16}$$

at $\mathbf{x}$ is

$$\frac{m_1x}{A-\lambda} + \frac{m_2y}{B-\lambda} + \frac{m_3z}{C-\lambda} = 0. \tag{7.17}$$

Substituting (7.15) into (7.17) and expressing $\mathbf{x}_0$ in terms of $\mathbf{n}$, we obtain

$$k = -\left(\frac{Am_1n_1}{A-\lambda} + \frac{Bm_2n_2}{B-\lambda} + \frac{Cm_3n_3}{C-\lambda}\right) \times\left[\left(\frac{m_1^2}{A-\lambda} + \frac{m_2^2}{B-\lambda} + \frac{m_3^2}{C-\lambda}\right)\right]^{-1}. \tag{7.18}$$

Substituting (7.15) into (7.16) and making use of (7.18), we obtain an equation in λ, m, n,

$$\frac{(An_1 + km_1)^2}{A-\lambda} + \frac{(Bn_2 + km_2)^2}{B-\lambda} + \frac{(Cn_3 + km_3)^2}{C-\lambda} = An_1^2 + Bn_2^2 + Cn_3^2. \tag{7.19}$$

Note that we have replaced 1 in the right-hand side of (7.16) by (nAn). This makes Eq. (7.19) homogeneous in both m and n.

Even though Eq. (7.19) looks complicated when (7.18) is inserted explicitly, it becomes quite simple after we multiply it by appropriate combinations of $(A-\lambda)$, $(B-\lambda)$, and $(C-\lambda)$. The final expression is

$$\begin{aligned} m^2(nAn)\lambda^2 &- \{AB[(n_1^2 + n_2^2)m^2 - (m_1n_2 - m_2n_1)^2] \\ &+ BC[(n_2^2 + n_3^2)m^2 - (m_2n_3 - m_3n_2)^2] \\ &+ CA[(n_3^2 + n_1^2)m^2 - (m_3n_1 - m_1n_3)^2]\}\lambda \\ &+ ABC(m\cdot n)^2 = 0. \end{aligned} \tag{7.20}$$

It is transparent that the coefficients of λ^2 and 1 are invariant under iterations (7.2) and (7.3). With a little work, we can also show that the linear term is also invariant under such iterations. This middle term is an independent quartic invariant which cannot be expressed in terms of the quadratic invariants. Thus we show that the quadratic equation (7.20) is invariant under interations (7.2) and (7.3). Since the roots of (7.13) determine the inner quadrics which are tangent to the rays, these inner quadrics are also invariant under iterations. This implies the existence of two caustic surfaces as promised. The reduction of our method and conclusion from three dimensions to two dimensions is straightforward. We can obtain equations analogous to (7.1)–(7.14) by treating m,n as two-dimensional vectors. The two-dimensional equivalent of (7.20) is a linear equation

$$m^2(nAn)\lambda - AB(m\cdot n)^2 = 0, \tag{7.21}$$

which implies a single caustic curve. In two dimensions, m^2, $(m\cdot n)$, and (nAn) are the only invariants under iterations (7.2) and (7.3). There are no independent quartic invariants.

In Secs. III and IV, we have shown by the Hamilton–Jacobi method that all the trajectories tangent to the same caustic curves or surfaces have the same period, independent of starting points. In Secs. V and VI, we have given a projective geometrical proof of this result for the two-dimensional model. The existence of a purely geometrical proof showed that the result is also true in different prime number fields. Numerically, we can apply the iteration equations in an arbitrary number field, and verify that the periods are indeed independent of the starting point when they are tangent to a fixed conic. What comes as a surprise is that the same result appears also to be valid numerically in three dimensions when the trajectories are tangent to two fixed quadric surfaces. At the moment, we are not able to provide a geometrical or an algebraic proof which can be extended into finite number fields.[13]

To understand the iteration equations in a finite number field, we need to examine some singular cases which can arise in a general field. We consider a finite number field defined by mod p, where p is a prime. The equation for a quadric confocal to the outer one is still

$$\frac{x^2}{A-\lambda} + \frac{y^2}{B-\lambda} + \frac{z^2}{C-\lambda} = 1, \quad \text{mod } p. \tag{7.22}$$

The outer conic is given by $\lambda = 0$. We can describe a ray as before by a linear equation. A ray is tangent to a quadric if they have a double root at their intersection. A quadric is a caustic to a ray if it is tangent to every segment of a trajectory. We can establish the existence of the caustic quadric

surfaces in a finite number field, and the associated λ obeys the same equation (7.20). We denote the λ associated with these two caustic quadrics by λ_1 and λ_2. For simplicity, we choose A, B, C, λ_1, and λ_2 such that none of $A-\lambda$, $B-\lambda$, and $C-\lambda$ is a multiple of p.

In the case of real numbers, n^2 and (mQm) are always positive and nonvanishing. The iterative equations (7.2), (7.3) are well defined. For a finite field, however, n^2 or (mQm) may become zero after some number of iterations even if we begin with finite n_0^2 and (m_0Qm_0). Here, we denote the m and n associated with the zeroth iteration by m_0 and n_0. Equations (7.2) and (7.3) are no longer valid when n^2 or $(mQm)=0$. We need to pay special attention to these singular cases. Note also that the duality transformation introduced earlier is still valid in a finite field. From an algebraic point of view, we can treat the case $n^2=0$ and $(mQm)=0$ in the same way. In the following, we shall concentrate on the $n^2=0$ case. The $(mQm)=0$ case follows with a minimal modification.

The first singular case is

$$n^2=0 \text{ but } (nQn)\neq 0. \tag{7.23}$$

Equation (7.2) is not defined when $n^2=0$. Hence we first keep n^2 undetermined and work out the formula for m'' in terms of m and n by applying (7.2), (7.3), (7.2) in succession. We shall find that m'' is given by an expression that remains well defined when $n^2=0$, provided that $nQn\neq 0$. Therefore we can simply continue the iteration from m'', provided we count the "missing" ray m' when evaluating the period.

To obtain the formula for m'', we shall use symbols $O[n^2]$, $O[n^4]$ to indicate terms of the corresponding order in a formal power series expansion. This is only a shortcut to avoid writing the full expressions, which are all rational functions. The procedure is valid in finite as well as continuous fields.

It is convenient to introduce an $\hat{m}'$ that is finite at $n^2=0$,

$$\hat{\mathbf{m}}'\equiv n^2\mathbf{m}' = -2(m\cdot n)\mathbf{n}+O(n^2). \tag{7.24}$$

We can compute easily that

$$(\hat{m}'Q\hat{m}') = 4(m\cdot n)^2(nQn)+O(n^2). \tag{7.25}$$

We can now evaluate $\mathbf{n}'$ as

$$\begin{aligned}\mathbf{n}' &= \mathbf{n}-2\mathbf{Q}\mathbf{m}'(m'\cdot n)/(m'Qm')\\ &= \mathbf{n}+2\mathbf{Q}\hat{\mathbf{m}}'(m\cdot n)n^2/(\hat{m}'Q\hat{m}')\\ &= \mathbf{n}-\mathbf{Q}\mathbf{n}n^2/(nQn)+O(n^4).\end{aligned} \tag{7.26}$$

From (7.26), we obtain both

$$\mathbf{n}\cdot\mathbf{n}' = O(n^4) \tag{7.27}$$

and

$$(\mathbf{n}-\mathbf{n}')^2 = n^2-2\mathbf{n}\cdot\mathbf{n}'+n'^2 = O(n^4), \tag{7.28}$$

which imply

$$n^2+n'^2=O(n^4), \quad n^2/n'^2=-1+O(n^2). \tag{7.29}$$

Now, it is straightforward to work out $\mathbf{m}''$,

$$\begin{aligned}\mathbf{m}'' &= \mathbf{m}'-2(m'\cdot n')\mathbf{n}'/n'^2\\ &= \mathbf{m}-2(m\cdot n)\mathbf{n}/n^2-2(m\cdot n)\mathbf{n}'/n'^2\\ &= \mathbf{m}-\frac{2(m\cdot n)\mathbf{n}}{n^2}-\frac{2(m\cdot n)}{n'^2}\left[\mathbf{n}-\frac{\mathbf{Q}\mathbf{n}n^2}{(nQn)}+O(n^4)\right]\\ &= \mathbf{m}-\frac{2(m\cdot n)\mathbf{Q}\mathbf{n}}{(nQn)}-2(m\cdot n)\mathbf{n}\left(\frac{1}{n^2}+\frac{1}{n'^2}\right)+O(n^2)\\ &= \mathbf{m}-\frac{2(m\cdot n)}{(nQn)}\mathbf{Q}\mathbf{n}-C\mathbf{n}+O(n^2).\end{aligned} \tag{7.30}$$

It is easy to see from (7.29) that the coefficient C is finite as $n^2\to 0$. We can obtain C either directly from (7.26) by keeping the $O(n^4)$ terms, or by using the invariant property

$$\begin{aligned}m''^2 &= m^2-\frac{4(m\cdot n)(mQn)}{(nQn)}\\ &\quad+\frac{4(m\cdot n)^2(nQ^2n)}{(nQn)^2}+2(m\cdot n)C=m^2\end{aligned} \tag{7.31}$$

giving

$$\begin{aligned}C &\equiv 2(m\cdot n)(1/n^2+1/n'^2)\\ &= 2\left(\frac{(mQn)}{(nQn)}-\frac{(m\cdot n)(nQ^2n)}{(nQn)^2}\right).\end{aligned} \tag{7.32}$$

It is simple to verify that all three conservation laws

$$m''^2=m^2, \quad (m''\cdot n)=-(m\cdot n),$$

and

$$(n'An')=(nAn),$$

are obeyed.

A second, more singular case is specified by

$$n^2=(nQn)=0 \quad \text{but} \quad (nQ^2n)\neq 0. \tag{7.33}$$

Here $\mathbf{m}'$ and $\mathbf{m}''$ are undefined but $\mathbf{m}'''$ is finite. We shall omit the algebra, and only summarize the results. The proper limits at $n^2=0$ are

$$\hat{\mathbf{m}}'\equiv n^2\mathbf{m}' = -2(m\cdot n)\mathbf{n}, \tag{7.34a}$$

$$\mathbf{n}' = \mathbf{n}+\mathbf{Q}\mathbf{n}(m\cdot n)/(mQn), \tag{7.34b}$$

$$\hat{\mathbf{m}}''\equiv n^2\mathbf{m}'' = -2(m\cdot n)\mathbf{n}, \tag{7.35a}$$

$$\mathbf{n}''=\mathbf{n}, \tag{7.35b}$$

$$\mathbf{m}'''=-\mathbf{m}. \tag{7.36}$$

It is straightforward to show that the conservation laws

$$m'''^2=m^2, \quad (n''An'')=(nAn),$$

$$(m'''\cdot n'')=-(m\cdot n)$$

are all satisfied. We believe that (7.34)–(7.36) give the correct continuation of (m,n) over the singular region defined by $n^2=(nQn)=0$ even in finite number fields.

In a three-dimensional space, we do not need to consider the more singular case of $n^2=(nQn)=(nQ^2n)=0$. In the three-dimensional space, Q obeys a cubic equation. If all three scalar products vanish, we then have $(nAn)=0$ which implies that the outer ellipsoid is degenerate. We shall not consider such a case here.

Knowing how to continue over the singular m and n's, we are able to test numerically over finite number fields what the caustic surfaces are and what the relations of the caustic

quadric with the periods are. Not surprisingly, we verify that a trajectory is tangent to the same caustic quadric surfaces after each iteration. As we have mentioned earlier, we are surprised to find that all the trajectories that share the same caustic surfaces have the same period independent of the starting points. We have made an exhaustive search covering all prime number fields with prime numbers smaller than or equal to 23, and covering all different and nonzero parameters A, B, C, α', α''. We consider only the nondegenerate case $m^2 \neq 0$, $nAn \neq 0$. The continuation procedures that we have introduced to go through singular points are essential for giving these constant periods. We almost always encounter these singularities at some point on the trajectory. The period is always the same whether the trajectory encounters a singularity or not, as long as it is tangent to the same caustic quadrics.

VIII. EXTENSION TO n DIMENSIONS AND DISCUSSION

We can generalize many of our results, including Poncelet's theorem, to n dimensions. The family of confocal quadrics obeys

$$\sum_{i=1}^{n} \frac{x_i^2}{A_i - \lambda} - 1 = 0. \tag{8.1}$$

The independent solutions to (8.1), $\lambda_1, \lambda_2, \ldots, \lambda_n$, form an orthogonal basis and are the n-dimensional elliptical coordinates. We have verified that all equations in Sec. IV can be extended to n dimensions. The Hamilton–Jacobi equation is separable. The $n-1$ separation constants α^i $(i = 1,2,\ldots,n-1)$ give rise to $(n-1)$ caustic confocal quadrics. We establish the following Poncelet's theorem in n dimensions: If one trajectory forms a closed polygon after p bounces, then all trajectories sharing the same $n-1$ caustic quadrics are also closed after p bounces independent of the starting point. As a billiard system, the total time period and length of these trajectories are the same.

We can generalize the iteration approach in Sec. VII to n dimensions simply by treating $\mathbf{m}$ and $\mathbf{n}$ as n-dimensional vectors. Equations (7.1)–(7.19) follow with trivial modifications. To obtain the generalization of (7.20), we begin with the analogs of (7.18) and (7.19),

$$k = -\sum_i \frac{A_i m_i n_i}{A_i - \lambda} \left(\sum_i \frac{m_i^2}{A_i - \lambda} \right)^{-1} \tag{8.2}$$

and

$$\sum_i \frac{(An_i + km_i)^2}{A_i - \lambda} - \sum A_i n_i^2 = \sum_i \frac{A_i^2 n_i^2}{A_i - \lambda} + 2k \sum_i \frac{A_i m_i n_i}{A_i - \lambda} + k^2 \sum_i \frac{m_i^2}{A_i - \lambda} - \sum A_i n_i^2 = 0. \tag{8.3}$$

By the use of (8.2), we can combine the k and k^2 terms in (8.3) into a single k-term, giving

$$\lambda \sum_i \frac{A_i n_i^2}{A_i - \lambda} + k \sum_i \frac{A_i m_i n_i}{A_i - \lambda} = 0. \tag{8.4}$$

Eliminating k between (8.2) and (8.4), we have

$$\lambda \sum_i \frac{A_i n_i^2}{A_i - \lambda} \sum_j \frac{m_j^2}{A_j - \lambda} - \left(\sum_i \frac{A_i m_i n_i}{A_i - \lambda} \right)^2 = 0. \tag{8.5}$$

The coefficients of the double poles in (8.5) all vanish identically. We can express (8.5) as the sum of single poles,

$$\sum_i \frac{C_i}{A_i - \lambda} = 0, \tag{8.6}$$

where the residuals C_i obey

$$C_i = -A_i n_i^2 \sum_j m_j^2 + \sum_{j \neq i} \left[\frac{A_i A_j}{A_j - A_i} (m_i n_j - m_j n_i)^2 \right]. \tag{8.7}$$

We can express (8.6) as an $(n-1)$th order polynomial equation which implies the existence of $(n-1)$ independent caustic confocal quadric surfaces. These C_i's provide n algebraic invariants. The iterative approach is applicable to all finite number fields.

We have not yet checked systematically whether this version of the Poncelet's theorem is valid for n-dimensional finite number fields.

We can solve the elliptical billiard system both classically and quantum mechanically. This system may provide us with a good testing ground for studying the quantum to classical transition. It may also serve as a starting point for analyzing nearly integrable systems.

The constancy of the total length for periodic orbits sharing the same caustic quadrics has important implications in quantum wave functions. In the WKB approximation, the phase factor of a wave function is proportional to the total length of the trajectory divided by $\hbar$. A resonance occurs when the phase factor associated with a closed orbit is a multiple of 2π. The constancy of orbital lengths implies that the WKB wave functions associated with trajectories sharing the same caustic quadrics become resonant simultaneously.

ACKNOWLEDGMENTS

We wish to thank Professor Richard Bishop for directing us to Hart's paper. One of us (R.F.) would like to thank the Physics Department of the University of Illinois at Urbana for its hospitality during an enjoyable stay. He would also like to thank Professor Daniel Grayson for explaining patiently how it is possible to differentiate a function with respect to a discrete variable, and Sophia Yancopoulos for pointing out that the application of our Hamilton–Jacobi method in three dimensions would yield an extension of Poncelet's theorem to a skew polygon.

This work was supported in part by the National Science Foundation under Grant No. PHY-87-01775.

APPENDIX: CHOICE OF QUADRICS IN HART'S CONSTRUCTION

Let us first consider two dimensions. We are given a pair of confocal conics $Q_0 = 0$ and $Q_1 = 0$, where

$$Q_0 = x^2/A + y^2/B - 1, \qquad Q_1 = x^2/(A - \alpha_1) + y^2/(B - \alpha_1) - 1. \tag{A1}$$

We wish to find a conic $Q_1' = 0$ which can be carried into Q_1 by a rescaling (say $x \to x/\xi, y \to y/\eta$) that at the same time renders it confocal with Q_0. This means that

$$Q_1' = \xi^2 x^2/(A-\alpha_1) + \eta^2 y^2/(B-\alpha_1) - 1 \qquad \text{(A2)}$$

and the rescaling must carry Q_0 into $\tilde{Q}_0$, which is confocal with Q_1 and hence with Q_0. Therefore

$$\tilde{Q}_0 = x^2/A\xi^2 + y^2/B\eta^2 - 1, \qquad \text{(A3)}$$

where

$$A\xi^2 - B\eta^2 = A - B. \qquad \text{(A4)}$$

Then we have

$$A(\xi^2 - 1) = B(\eta^2 - 1) = \mu \qquad \text{(A5)}$$

from which one easily deduces that

$$\alpha_1(Q_1' - Q_1) = \mu(Q_1 - Q_0). \qquad \text{(A6)}$$

Therefore Q_1' belongs to the pencil generated by Q_1 and Q_0.

Conversely, if Q_1' is any member of this pencil, (A6) is satisfied for some μ; and then working backwards one can derive all the previous equations, showing that Q_1' satisfies the initial requirements.

In three dimensions we have

$$\begin{aligned} Q_0 &= \frac{x^2}{A} + \frac{y^2}{B} + \frac{z^2}{C} - 1, \\ Q_1 &= \frac{x^2}{A-\alpha_1} + \frac{y^2}{B-\alpha_1} + \frac{z^2}{C-\alpha_1} - 1, \\ Q_2 &= \frac{x^2}{A-\alpha_2} + \frac{y^2}{B-\alpha_2} + \frac{z^2}{C-\alpha_2} - 1, \end{aligned} \qquad \text{(A7)}$$

but now the same rescaling must be applied to all three quadrics simultaneously. Hence

$$\begin{aligned} Q_1' &= \frac{\xi^2 x^2}{A-\alpha_1} + \frac{\eta^2 y^2}{B-\alpha_1} + \frac{\zeta^2 z^2}{C-\alpha_1} - 1, \\ Q_2' &= \frac{\xi^2 x^2}{A-\alpha_2} + \frac{\eta^2 y^2}{B-\alpha_2} + \frac{\zeta^2 z^2}{C-\alpha_2} - 1, \end{aligned} \qquad \text{(A8)}$$

must share a set of parameters, ξ, η, ζ. Again $\tilde{Q}_0$ must be confocal with Q_0, which means it belongs to the same system of Jacobi variables. This requires (A4) as well as

$$B\eta^2 - C\zeta^2 = B - C. \qquad \text{(A9)}$$

Hence

$$A(\xi^2 - 1) = B(\eta^2 - 1) = C(\zeta^2 - 1) = \mu, \qquad \text{(A10)}$$

which shows as before that Q_1', Q_2' belong, respectively, to the two pencils generated by Q_1 and Q_0 and by Q_2 and Q_0.

However, (A10) also shows that Q_1', Q_2' are simultaneously determined by a single parameter μ. The initial requirement, then, will not be satisfied if they are chosen independently from their respective pencils.

[1] G. D. Birkhoff, "Dynamical system," Vol. IX, AMS Coll. (1927). See also Ya. G. Sinai, Russ. Math. Surveys **25**, 137 (1970).

[2] H. Poincare, *Les Methodes Nouvelles de la Mecanique Celeste* (Gauthier–Villars, Paris, 1892) (also reprinted by Dover, New York, 1957; NASA Translation TTF-450, Washington, 1976).

[3] J. B. Keller and S. I. Rubinow, Ann. Phys. (NY) **9**, 24 (1960).

[4] Ya. G. Sinai, *Introduction to Ergodic Theory* (Princeton U. P., Princeton, NJ, 1976).

[5] J. V. Poncelet, *Traité des Propriété Projectives des Figures* (Gauthier–Villars, Paris, 1866), 2nd ed., pp. 311–318. A summary of Poncelet's proof was given by A. Cayley, in Quart. J Pure Appl. Math. **2**, 31 (1858).

[6] C. G. J. Jacobi, Gesammelte Werke **1**, 278 (1881).

[7] For a summary of formulas on elliptical functions, see, e.g., M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1964).

[8] A. S. Hart, Quart. J. Pure Appl. Math. **2**, 143 (1858). See also H. Lebesgue, Ann. Fac. Sci. Toulouse **13**, 61 (1921).

[9] See, e.g., E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1927), pp. 547–550; P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York 1953), Vol. 1, pp. 511–513.

[10] P. Griffiths and J. Harris, Comment. Math. Helv. **52**, 145 (1977); P. Griffiths and J. Harris, L'Enseignement Math. **XXIV**, 31 (1978).

[11] For recent work in this area, see, e.g., Y. Ayant and R. Arvieu, J. Phys. A **20**, 397 (1987); R. Arvieu and Y. Ayant, *ibid.* **20**, 1115 (1987); M. V. Berry, Eur. J. Phys. **2**, 91 (1981).

[12] A. Cayley, Philos. Mag. **6**, 99 (1853); Philos. Trans. R. Soc. London **151**, 225 (1861).

[13] For a survey of projective geometries over finite fields, see J. W. P. Hirschfield, *Projective Geometries Over Finite Fields* (Clarendon, Oxford, 1979); *Finite Projective Spaces of Three Dimensions* (Clarendon, Oxford, 1985).

Alternative Lagrangians for spherically symmetric potentials

M. Crampin
Faculty of Mathematics, The Open University, Walton Hall, Milton Keynes MK7 6AA, England

G. E. Prince
Department of Mathematics, LaTrobe University, Bundoora, Victoria 3083, Australia

(Received 2 December 1987; accepted for publication 9 March 1988)

A new, geometrical account of the existence of alternative (s-equivalent) Lagrangians is given for a particle moving in three-dimensional Euclidean space under the influence of a central force. The fact that the Lagrangian for this familiar problem is not unique was pointed out by Henneaux and Shepley [J. Math. Phys. **23**, 2101 (1982)]: the present analysis makes the reasons for the phenomenon more transparent than theirs.

I. INTRODUCTION

In a 1982 paper Henneaux and Shepley[1] showed that the Lagrangian describing the classical motion of a particle moving in three-dimensional Euclidean space under a central force, that is to say, a force derived from a spherically symmetric potential function, is not unique. The usual freedoms of multiplication of the Lagrangian by a constant and the addition to it of a total time derivative are not at issue here. Rather than the trivial equivalence thus defined we shall be concerned with what Henneaux and Shepley[1] (and others) have called solution equivalence or s-equivalence: Two Lagrangians, not trivially equivalent, are said to be s-equivalent if their Euler–Lagrange equations have the same solutions. The existence of s-equivalent Lagrangians has been a topic of some considerable interest over the last few years and has led to the publication of quite a number of papers.[2] However, it remains true to say that there are no known necessary and sufficient conditions for a dynamical system to admit s-equivalent, but not trivially equivalent, Lagrangians. One case, which might be described as the decomposable case, has been completely analyzed by Marmo and his co-workers.[3] The case of the spherically symmetric potential may perhaps be considered as being at the opposite end of the spectrum. Our excuse for revisiting this problem, and presenting a new solution, is that we believe that our geometrical constructions, while clarifying what happens in this particular case, also contain indications of how the general problem might be tackled.

Henneaux and Shepley[1] showed that the essential freedom in the Lagrangian is a function of two variables, which upon examination of their paper, is revealed to be a function on the unit sphere. The term to be added to the standard Lagrangian to obtain an s-equivalent one is determined from this function by quadratures. One example in which the new Lagrangian can be given explicitly is

$$L' = L + \gamma J/r^2,$$

where L is the usual Lagrangian, L' is the new one, γ is a constant, and J is the magnitude of the angular momentum. These results follow from the application of Henneaux's[1] method of tackling the inverse problem of Lagrangian dynamics, which uses the exterior calculus but is basically algebraic and analytical. Our approach in this paper is, by contrast, geometrical, in the sense that it is concerned with geometrical features of the velocity phase space of the system.

We have shown in a previous paper[4] that the Lagrangians of the class displayed above are the only ones that are themselves spherically symmetric; our argument, not surprisingly, was concerned with the action of the rotation group on the system. Group theory will play a much less prominent role here, but it will become apparent how that result fits in with the more general picture to be painted below.

II. BACKGROUND

We shall need the following general results[5] concerning the differential geometry of Lagrangian systems.

The velocity phase space of a Lagrangian system is the tangent bundle TM of a differentiable manifold M, the configuration space of the system. A tangent bundle is equipped with two natural geometric objects: a type (1,1) tensor field called the vertical endomorphism, denoted S and given in adapted coordinates by $S = \partial/\partial u^a \otimes dx^a$; and a vector field, the dilation field, denoted Δ and given in coordinates by $\Delta = u^a \partial/\partial u^a$. Conversely, a manifold, necessarily of even dimension, which carries a type (1,1) tensor field with the same properties as S and satisfies some appropriate topological conditions is affinely equivalent to a tangent bundle[6]; when a choice of vector field with the requisite properties to play the role of Δ is made it becomes equivalent linearly to a tangent bundle.

A vector field Γ on TM is called a second-order differential equation field if it satisfies $S(\Gamma) = \Delta$; this is a necessary and sufficient condition for the projections onto M of its integral curves to be solution curves of a system of second-order ordinary differential equations $\ddot{x}^a = F^a(x,\dot{x})$.

A Lagrangian is simply a smooth function L on a tangent bundle TM. We shall have to deal with functions defined not on the whole manifold TM, but only on an open submanifold of it; however, we shall use the same terminology. The Cartan form ω_L of L is the two-form given by $\omega_L = d(S(dL))$ and the energy function of L is given by

 0022-2488/88/071551-04$02.50

$E_L = \Delta(L) - L$. The Lagrangian is said to be regular if ω_L is symplectic, in which case there is a unique vector field Γ such that $i_\Gamma \omega_L = -dE_L$ (where i denotes the interior product) and Γ is a second-order differential equation field, the Euler–Lagrange field of L. Any Cartan form satisfies $i_{S(X)}\omega_L = -S(i_X\omega_L)$ for any vector field X on TM. The map $L \mapsto \omega_L$ is linear over the reals and its kernel consists of functions of the form $\tau + k$, where τ is a total time derivative and k is constant.

The inverse problem of Lagrangian dynamics (or of the calculus variations) is the problem of stating necessary and sufficient conditions for a given second-order differential equation field Γ to be the Euler–Lagrange field of some regular Lagrangian. A partial solution consists of giving necessary and sufficient conditions for a two-form Ω to be the Cartan form of some Lagrangian for which Γ is the Euler–Lagrange field. Such conditions have been given in various differing ways by several authors, including Henneaux.[1] We give them here in the form most suitable for later use.[7] The two-form Ω is (locally) a Cartan form for Γ if and only if it satisfies the following three conditions: (i) Ω is closed, (ii) Ω is zero when both its arguments are vertical (with respect to the tangent bundle projection), and (iii) $\mathscr{L}_\Gamma \Omega = 0$.

We shall also need one particular result of Henneaux's[1] analysis: The Lagrangian for nonradial motion for the central force problem in two dimensions is unique up to trivial equivalence.

III. SPHERICALLY SYMMETRIC POTENTIALS

The velocity phase space for the particular problem at hand is $T\mathscr{E}^3_0$, where $\mathscr{E}^3_0$ is the three-dimensional Euclidean space with its origin removed. We consider $T\mathscr{E}^3_0$ as the set of pairs of vectors $(\mathbf{x},\mathbf{u})$. Let $T\mathscr{E}^{3+}_0$ be the open submanifold of $T\mathscr{E}^3_0$ consisting of those points for which $\mathbf{x}\times\mathbf{u}\neq\mathbf{0}$; then there is a smooth map $\hat{\mathbf{J}}$ of $T\mathscr{E}^{3+}_0$ onto the unit two-sphere $\mathscr{S}^2$ defined by

$$\hat{\mathbf{J}}(\mathbf{x},\mathbf{u}) = (\mathbf{x}\times\mathbf{u})/|\mathbf{x}\times\mathbf{u}|.$$

Then $\hat{\mathbf{J}}$ defines a fibration of $T\mathscr{E}^{3+}_0$ over $\mathscr{S}^2$ in which the individual fibers are the tangent bundles of two-planes in $\mathscr{E}^3_0$, the two-planes of constant direction of angular momentum. For any spherically symmetric potential function V on $\mathscr{E}^3_0$ the Euler–Lagrange field Γ of the Lagrangian $L(\mathbf{x},\mathbf{u}) = \frac{1}{2}|\mathbf{u}|^2 - V(\mathbf{x})$ is everywhere tangent to the fibers of $\hat{\mathbf{J}}$, in fact, for any function f on $\mathscr{S}^2$ the function $\hat{\mathbf{J}}^* f$ is a constant of the motion. In geometrical form this expresses the usual reduction of a three-dimensional central force problem to two-dimensional considerations familiar from elementary orbit theory.

The ambiguity in the choice of Lagrangian for motion under a spherically symmetric potential is a direct consequence of the existence of this geometrical structure associated with the dynamical system.

Proposition 1: Let ω be an arbitrary two-form on $\mathscr{S}^2$; then $\hat{\mathbf{J}}^*\omega$ is a (necessarily degenerate) Cartan form for Γ.

Proof: We have to show that $\Omega = \hat{\mathbf{J}}^*\omega$ satisfies the conditions stated in Sec. II.

Closure: Since ω is a two-form on a two-dimensional manifold $d\omega = 0$, therefore $d\Omega = d(\hat{\mathbf{J}}^*\omega) = \hat{\mathbf{J}}^* d\omega = 0$.

Evaluation on vertical vectors: Consider the vertical subspace of $T_{(\mathbf{x},\mathbf{u})} T\mathscr{E}^{3+}_0$ for $(\mathbf{x},\mathbf{u}) \in T\mathscr{E}^{3+}_0$. The kernel of $\hat{\mathbf{J}}_*$ restricted to this subspace is two-dimensional: It consists of the vertical lifts to $T\mathscr{E}^{3+}_0$ of vectors tangent to the two-plane of constant direction of angular momentum determined by $(\mathbf{x},\mathbf{u})$. Let w be a vertical vector at $(\mathbf{x},\mathbf{u})$ complementary to the kernel of $\hat{\mathbf{J}}_*$. Then any vertical factor u may be uniquely written $u = u_0 + kw$, with $\hat{\mathbf{J}}_* u_0 = 0$ and k some real number. Thus if $v = v_0 + lw$ is also vertical and similarly decomposed, then

$$\Omega(u,v) = \Omega(u_0,v) + k\Omega(w,v) = \Omega(u_0,v) + k\Omega(w,v_0);$$

however, $\Omega = \hat{\mathbf{J}}^*\omega$ vanishes if one of its arguments lies in the kernel of $\hat{\mathbf{J}}_*$.

Vanishing Lie derivative: Since Γ is tangent to the fibers of $\hat{\mathbf{J}}$, $\hat{\mathbf{J}}_* \Gamma = 0$ and therefore $i_\Gamma \Omega = i_\Gamma \hat{\mathbf{J}}^*\omega = 0$. It follows that

$$\mathscr{L}_\Gamma \Omega = i_\Gamma d\Omega + d(i_\Gamma \Omega) = 0.$$

In fact, Γ is a characteristic vector field for Ω and it is well-known that the Lie derivative of a closed form with respect to any characteristic vector field is zero. ■

Thus given a two-form ω on $\mathscr{S}^2$ we can obtain a degenerate, possibly local, Lagrangian for Γ, namely the function from which the Cartan form $\hat{\mathbf{J}}^*\omega$ is derived. Then the sum of this and the standard Lagrangian is an alternative Lagrangian for Γ. The freedom in the choice of a two-form on $\mathscr{S}^2$ is essentially the same as the freedom in the choice of a function on $\mathscr{S}^2$, namely the coefficient when the two-form is expressed as a multiple of the volume two-form.

Note that the new Lagrangian is defined only for nonradial motion. In fact, if one calculates $\omega_L + \hat{\mathbf{J}}^*\omega$ explicitly one sees that it can never be extended smoothly to the whole of $T\mathscr{E}^3_0$ (unless $\omega = 0$). It is clear from the same calculation that the new Lagrangian may very well fail to be regular at points of $T\mathscr{E}^{3+}_0$.

The relation between the construction explained above and the spherical symmetry of the system, in the sense of its invariance under rotations, may be seen as follows. The rotation group SO(3) acts on $\mathscr{E}^3_0$ in the usual way; this action extends to an action on $T\mathscr{E}^3_0$ via the complete lift and $T\mathscr{E}^{3+}_0$ is invariant. The map $\hat{\mathbf{J}}$ is equivariant under the action of SO(3) on $T\mathscr{E}^{3+}_0$ and the usual action on $\mathscr{S}^2$: For any $g \in \mathrm{SO}(3)$,

$$\hat{\mathbf{J}} \circ g = g \circ \hat{\mathbf{J}}.$$

Thus the two-form $\hat{\mathbf{J}}^*\omega$ (on $T\mathscr{E}^{3+}_0$) is invariant under SO(3) if and only if ω (on $\mathscr{S}^2$) is also invariant. However, this is so only if ω is a constant multiple of the volume two-form on $\mathscr{S}^2$. It is this case that leads to the Lagrangian $L + \gamma J/r^2$.

The converse, that every alternative Lagrangian for a spherically symmetric potential can be obtained in this way, follows, of course, from the analysis of Henneaux and Shepley.[1] It is, however, instructive to rederive this result using the methods adopted here. Our argument takes advantage of the fact that one can specify the fibration on which our construction depends in another way, without appealing directly to angular momentum. Consider the vector fields

$$x^a\frac{\partial}{\partial x^a},\quad u^a\frac{\partial}{\partial x^a},\quad x^a\frac{\partial}{\partial u^a},\quad u^a\frac{\partial}{\partial u^a}$$

on $T\mathscr{C}_0^{3+}$: They are linearly independent (over the algebra of functions) and annihilated by $\hat{\mathbf{J}}_*$. In fact, these vector fields span a distribution $\mathscr{D}$ that is integrable in the sense of Frobenius's theorem and whose leaves are the fibers of the fibration $\hat{\mathbf{J}}$: $T\mathscr{C}_0^{3+}\to\mathscr{S}^2$. This distribution $\mathscr{D}$ has two significant properties (in addition to being integrable): It contains the Euler–Lagrange field Γ and it is invariant under the vertical endomorphism S. In fact, provided the force is nowhere vanishing, $\mathscr{D}$ may be specified as the smallest integrable distribution containing Γ invariant under S.

In order to derive our result we shall have to assume that the energy associated with the alternative Lagrangian is smooth everywhere on $T\mathscr{C}_0^3$, in particular on the zero section: However, this appears to be assumed implicitly in the work of Henneaux and Shepley[1] in any case. We assume also that the potential function never vanishes.

Proposition 2: Let L' be an alternative Lagrangian for Γ such that $E_{L'}$ is smooth on $T\mathscr{C}_0^3$. Then, up to trivial equivalence,

$$\omega_{L'}=\omega_L+\omega^*$$

where L is the standard Lagrangian and ω^* is the pullback to $T\mathscr{C}_0^{3+}$ of a two-form ω on the leaf space $T\mathscr{C}_0^{3+}/\mathscr{D}=\mathscr{S}^2$.

Proof: It follows from the fact that the Lagrangian for the two-dimensional central force problem is essentially unique that the restriction of L' to any leaf of $\mathscr{D}$ is of the form $kL+\tau$, where k is constant and τ is a total time derivative, both on the leaf. Thus there are functions, still denoted k and τ, on $T\mathscr{C}_0^{3+}$ such that $L'=kL+\tau$, with k constant and τ a total time derivative on restriction to any leaf of $\mathscr{D}$. Now $\Delta=S(\Gamma)$ belongs to $\mathscr{D}$ and therefore $\Delta(k)=0$ since k is constant on each leaf, and $\Delta(\tau)=\tau$ because τ, being a total time derivative, is linear in the fiber coordinates on each leaf. The energy associated with L' is given by

$$E_{L'}=k\Delta(L)+L\Delta(k)+\Delta(\tau)-kL-\tau=kE_L.$$

Now E_L and $E_{L'}$ are smooth everywhere on $T\mathscr{C}_0^3$ and E_L is nonzero on the zero section, so k is smooth in some open neighborhood of the zero section. However, $\Delta(k)=0$; it follows that k is independent of the fiber coordinates on $T\mathscr{C}_0^{3+}$. Furthermore, since both E_L and $E_{L'}$ are constants of the motion, so is k, so that $\Gamma(k)=0$; however, from this and the independence of k of the fiber coordinates it follows that k is constant. By invoking trivial equivalence we may therefore, without loss of generality, take k to be 1. We therefore have $\omega_{L'}=\omega_L+\omega_\tau$ and $E_{L'}=E_L$. Now L' is a Lagrangian for Γ, so

$$i_\Gamma\omega_{L'}+dE_{L'}=i_\Gamma\omega_L+dE_L=0;$$

this means that

$$i_\Gamma\omega_\tau=0.$$

Thus the characteristic distribution of ω_τ contains Γ. However, ω_τ, being a Cartan form, must satisfy $i_{S(X)}\omega_\tau=-S(i_X\omega_\tau)$ for any vector field X, as we pointed out earlier; in particular, this means that the characteristic distribution of ω_τ is invariant under S. Furthermore, ω_τ is closed, from which it follows that its characteristic distribution is integrable. Thus the characteristic distribution of ω_τ is integrable, contains Γ, and is invariant under S; however, $\mathscr{D}$ is the smallest such distribution and so the characteristic distribution of ω_τ contains $\mathscr{D}$. Finally, since the Lie derivative of a closed form by a characteristic vector field is zero, the Lie derivative of ω_τ by any vector field in $\mathscr{D}$ is zero. These are just the conditions which ensure that ω_τ passes to the quotient $T\mathscr{C}_0^{3+}/\mathscr{D}=\mathscr{S}^2$. ■

It is interesting to note that the dimension of $\mathscr{D}$ drops from 4 to 2 when one goes outside $T\mathscr{C}_0^{3+}$. Also, the action of SO(3) leaves $\mathscr{D}$ invariant and therefore permutes its leaves; in fact, each leaf of $\mathscr{D}$ is a section of the action of SO(3) in the sense defined in our earlier paper.[4]

IV. TOWARD GENERALIZATION

It seems clear from the foregoing analysis that the analog of the distribution $\mathscr{D}$ has an important role to play in the problem of ambiguity of Lagrangians. In general, for a given second-order differential equation field Γ on a tangent bundle TM, the distribution $\mathscr{D}$ (the smallest integrable distribution on TM which contains Γ and is invariant under S) will be the whole module of vector fields on TM. However, there are certainly cases of interest in which $\mathscr{D}$ is smaller: The extreme case occurs when Γ is a spray, in which case $\mathscr{D}$ is spanned by Γ and Δ.

To conclude this paper, we therefore prove some results about $\mathscr{D}$ that should be relevant to this problem.

Proposition 3: Each leaf of $\mathscr{D}$, if it projects onto a smooth submanifold of M, is the tangent bundle of its projection.

Proof: We show first that, for every vertical vector field $V\in\mathscr{D}$, there is a vector field $X\in\mathscr{D}$ such that $V=S(X)$. Since V is vertical it satisfies $S(V)=0$ and $\mathscr{L}_\Gamma S(V)=V$. Thus

$$V=[\Gamma,S(V)]-S([\Gamma,V])=S([V,\Gamma])$$

and we may take $X=[V,\Gamma]\in\mathscr{D}$.

The restriction of the tangent bundle projection π to a leaf ℓ of $\mathscr{D}$ is a submersion of it onto its image, which by assumption is a submanifold of M. The restriction of S to ℓ is an integrable almost tangent structure by the result first proved. We may therefore identify ℓ with the tangent bundle of the leaf space of the kernel distribution of the restriction of S to ℓ once a zero section of ℓ has been chosen. However, the intersection of ℓ with the zero section of TM gives a zero section for ℓ and the quotient space may be identified with $\pi(\ell)$. ■

Proposition 4: If the distribution $\mathscr{D}$ is invariant under a type (1,1) tensor field T then it is invariant under $\mathscr{L}_\Gamma T$. In particular, $\mathscr{D}$ is invariant under $\mathscr{L}_\Gamma^k S$, $k=1,2,\ldots$.

Proof: By assumption, $T(X)\in\mathscr{D}$ for every $X\in\mathscr{D}$. Thus

$$\mathscr{L}_\Gamma T(X)=[\Gamma,T(X)]-T([\Gamma,X])\in\mathscr{D}$$

and so $\mathscr{D}$ is invariant under $\mathscr{L}_\Gamma T$. ■

Proposition 5: If ϕ is a diffeomorphism of M for which the complete lift $\tilde{\phi}$ to TM is a symmetry of Γ, then $\tilde{\phi}$ permutes the leaves of $\mathscr{D}$. If G is a group of Lie symmetries of Γ, then there is an action of G on the leaf space $TM/\mathscr{D}$ which makes the projection $TM\to TM/\mathscr{D}$ equivariant.

Proof: Consider the distribution $\tilde{\phi}_* \mathscr{D}$: It contains Γ because $\tilde{\phi}$ is a symmetry, it is invariant under S because $\tilde{\phi}$ is a complete lift, and it is integrable because $\tilde{\phi}$ is a diffeomorphism. Thus $\mathscr{D} \subseteq \tilde{\phi}_* \mathscr{D}$; however, since $\tilde{\phi}$ is a diffeomorphism, $\mathscr{D} = \tilde{\phi}_* \mathscr{D}$. The results for Lie symmetries follow. ■

ACKNOWLEDGMENTS

This work was begun while MC was visiting LaTrobe University and completed while GEP was visiting the Open University. Each author would like to thank the other's home institution for its hospitality.

[1]M. Henneaux and L. C. Shepley, "Lagrangians for spherically symmetric potentials," J. Math. Phys. **23**, 2101 (1982).
[2]See, for example, S. Hojman and H. Harleston, "Equivalent Lagrangians: Multidimensional case," J. Math. Phys. **22**, 1414 (1981); M. Henneaux, "Equations of motion, commutation relations and ambiguities in the Lagrangian formalism," Ann. Phys. (NY) **140**, 45 (1982); M. Crampin and G. E. Prince, "Equivalent Lagrangians: Some comments," Phys. Lett. A **108**, 191 (1985); and other papers cited below.
[3]C. Ferrario, G. Lo Vecchio, G. Marmo, and G. Morandi, "Separability of completely integrable dynamical systems admitting alternative Lagrangian descriptions," Lett. Math. Phys. **9**, 140 (1985); C. Ferrario, G. Lo Vecchio, G. Marmo, G. Morandi, and C. Rubano, "A separability theorem for dynamical systems admitting alternative Lagrangian descriptions," preprint, University of Ferrara, 1986.
[4]M. Crampin and G. E. Prince, "Generalizing gauge variance for spherically symmetric potentials," J. Phys. A: Math. Gen. **18**, 2167 (1985).
[5]For an exposition of Lagrangian dynamics containing the results used in the present paper see M. Crampin, "Tangent bundle geometry for Lagrangian dynamics," J. Phys. A: Math. Gen. **16**, 3755 (1983).
[6]M. Crampin and G. Thompson, "Affine bundles and integrable almost tangent structures," Math. Proc. Cambridge Philos. Soc. **98**, 61 (1985).
[7]M. Crampin, "On the differential geometry of the Euler–Lagrange equations, and the inverse problem of Lagrangian dynamics," J. Phys. A: Math Gen. **14**, 2567 (1981).

$N=2$ supertori and their representation as algebraic curves

Ezer Melzer
Enrico Fermi Institute and Department of Physics, The University of Chicago, Chicago, Illinois 60637

(Received 29 December 1987; accepted for publication 16 March 1988)

$N=2$ supertori ($N=2$ super-Riemann surfaces of genus 1) are classified according to their superconformal equivalence classes. A detailed description of the corresponding supermoduli space is based on this classification. Superelliptic functions are constructed and used for representing the supertori as algebraic varieties in projective superspace. The boundary of $N=2$, genus 1 supermoduli space is discussed.

I. INTRODUCTION

The theory of Riemann surfaces has become a natural framework within which (bosonic) string theory is formulated.[1] Its counterpart in superstring (fermionic string) models is the theory of super-Riemann surfaces (SRS's), whose rich structure is only beginning to be explored, both by physicists and mathematicians.[2–5] SRS's enter the physical theory as the world sheets of a closed oriented superstring, and in order to sum over all superstring configurations we have to integrate over the corresponding supermoduli space,[6] describing superconformally inequivalent SRS's. Much work has still to be done before our understanding of the structure of supermoduli spaces is complete. In addition, the representation of SRS's as algebraic curves, besides the pure mathematical interest in it, may be appropriate for application of algebraic or number-theoretic methods in the context of superstring theory. Such methods are desirable as offering a hope for nonperturbative superstring calculations.[7]

The present work is concerned mainly with genus 1, $N=2$ SRS's ($N=2$ supertori), where N is the number of fermionic (anticommuting) coordinates that join the single bosonic (commuting), complex coordinate describing an underlying ordinary RS. Just as ordinary tori, supertori are obtained by taking the quotient of the superplane by the action of a supergroup generated by two commuting supertranslations. We work out in detail the supermoduli space and use superelliptic functions for representing the supertori as algebraic curves, following the approach of Refs. 3 and 4 where the $N=1$ case is treated. The $N=2$ case does not exhibit merely the features of a "product" of two $N=1$ cases. There is now an interplay not only between each one of the two fermionic coordinates and the bosonic one, but also an interplay among the two fermionic coordinates themselves. This new feature manifests itself in the $U(1)$ symmetry present in the $N=2$ superstring[8]; in superstring models with $N=2$ SUSY on the world sheet such symmetry is used for achieving space-time-supersymmetric compactifications.[9,10] As a result we get a bosonic supermodulus, in addition to two fermionic ones and the ordinary bosonic modulus (genus 1), representing the freedom to rotate the fermionic coordinates one to another as we go along a cycle of the underlying ordinary torus.[5] This freedom also enables a smooth transition between all the different spin structures that are embedded in the $N=2$ supertori, by going along continuous paths in the supermoduli space. It may be useful for studying the $N=1$ superstring as embedded in the $N=2$ supermoduli space.[5]

In Sec. II we discuss $N=2$ SRS's in general and find the generic form of a superconformal transformation that can be used to define their transition functions. Section III concentrates on $N=2$ supertori and their superconformal equivalence classes. The technical tools needed for the classification are described in Appendix A. In Sec. IV we describe the genus 1 super-Teichmüller space, and use the supermodular group in order to obtain from it the supermoduli space. A discussion of the embedded spin structures concludes this section. Superelliptic functions, meromorphic superfunctions that are defined on supertori, are constructed in Sec. V for the various supertorus classes. We find sets of them that are complete, in the sense that any meromorphic superfunction on a supertorus can be rationally expressed in terms of the functions in the corresponding set. Rational equations involving superelliptic functions enable us to embed the supertori into projective superspaces, where the image of the supertori is obtained as the locus of points satisfying a set of polynomial equations. The embeddings, discussed in Sec. VI, allow us to study the compactification of the supermoduli space by adding to it points at infinity, corresponding to singular, pinched supertori. Section VII is devoted to conclusions and open problems. Subtleties that arise from the use of rigorous supermanifold theory[11] were deferred to Appendix B.

II. $N=2$ SUPER-RIEMANN SURFACES

A $N=2$ SRS is described locally by one complex bosonic coordinate z and two complex fermionic coordinates θ^i ($i=1,2$), lumped together to form one superspace coordinate $Z=(z,\theta)=(z,\theta^1,\theta^2)$, with superconformal transition functions between overlapping coordinate patches. A transformation from one coordinate patch $U(Z)$ to another $\tilde{U}(\tilde{Z})$ is superconformal if the following condition holds in the overlap region[2,5]:

$$D\tilde{z}=\tilde{\theta}^1 D\tilde{\theta}^1+\tilde{\theta}^2 D\tilde{\theta}^2, \tag{2.1}$$

where D is the two-component covariant derivative whose components are

 0022-2488/88/071555-14$02.50

$$D_i = \frac{\partial}{\partial\theta^i} + \theta^i \frac{\partial}{\partial z},$$
$$\{D_i, D_j\}_+ = 2\delta_{ij} \frac{\partial}{\partial z} \quad (i,j = 1,2). \tag{2.2}$$

By solving the condition (2.1), one can show that any superconformal transformation is of the form

$$\tilde{z}(z,\theta) = f(z) + h(z)[\theta^T O(z)\psi(z)] - \theta^1\theta^2(\det O)\left[\frac{d\psi^T}{dz}\epsilon\psi(z)\right], \tag{2.3}$$
$$\tilde{\theta}(z,\theta) = \psi(z) + h(z)O^T(z)\theta + \theta^1\theta^2(\det O)\epsilon\frac{d\psi}{dz},$$

where the bosonic analytic functions $f(z)$, $h(z)$, and the column vector $\psi(z)$ of two fermionic analytic functions satisfy

$$\frac{df}{dz} = h^2(z) + \frac{d\psi^T}{dz}\psi(z). \tag{2.4}$$

Here $O(z)$ is an arbitrary bosonic, analytic O(2) matrix,

$$O^T(z)O(z) = \mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \det O(z) = \pm 1, \tag{2.5}$$

and ϵ is the antisymmetric 2×2 matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Here and in what follows θ is treated as a two-component column vector. The Lie superalgebra of this supergroup of transformations is the $N = 2$ superconformal algebra (SCA) in two dimensions (one complex dimension).[8]

As an important special case of these transformations we have the supergroup $\mathrm{SPL}_{N=2}(2,\mathbb{C})$ of superconformal automorphisms of the $N = 2$ super-Riemann sphere SC_2^* (we follow the notation of Ref. 3),

$$\begin{aligned}\tilde{z}(z,\theta) &= (az+b)/(cz+d) \\ &\quad + (cz+d)^{-2}\{\theta^T O(\gamma z+\delta)(1+\tfrac{1}{2}\delta^T\gamma) \\ &\quad - \theta^1\theta^2(\det O)[\delta^T\epsilon\gamma(cz+d) \\ &\quad + 2\,d\gamma_1\gamma_2 z - 2c\delta_1\delta_2]\}, \\ \tilde{\theta}(z,\theta) &= (\gamma z+\delta)/(cz+d) \\ &\quad + (cz+d)^{-1}(1+\tfrac{1}{2}\delta^T\gamma + \tfrac{1}{4}\gamma_1\gamma_2\delta_1\delta_2)O^T\theta \\ &\quad + \theta^1\theta^2(\det O)(cz+d)^{-2}\epsilon(d\gamma - c\delta),\end{aligned} \tag{2.6}$$

where a, b, c, d are bosonic constants (normalized so that $ad - bc = 1$), γ and δ fermionic constant column vectors, and O a bosonic constant O(2) matrix. The corresponding Lie superalgebra osp(2,2) is a subalgebra of the $N = 2$ SCA, generated by the four bosonic and four fermionic generators $L_{\pm 1}, L_0, T_0, G^i_{-1/2}, G^i_{1/2}$ $(i = 1,2)$.[5] Notice that if we allow O to be z dependent, (2.6) is still a superconformal automorphism of the supersphere, but we lose the closure of the whole family of these transformations. This can be seen more easily from the algebra point of view [where such z dependence implies the joining of all T_m $(m\in\mathbb{Z})$ to the above set of generators], from the commutation relation

$$[T_m, G^i_n] = i\epsilon_{ij}G^j_{m+n} \quad (i,j = 1,2,\ m\in\mathbb{Z},\ n\in\tfrac{1}{2}+\mathbb{Z}). \tag{2.7}$$

Assuming a uniformization theorem holds, as was proved to be the case for $N = 1$,[3] every $N = 2$ SRS M is the quotient of SC_2^*, $\mathrm{SC}_2 = \mathbb{C}^{1,2}$, or SH_2 (the $N = 2$ super-Riemann sphere, superplane, or upper half-superplane, respectively) by a discrete group G which is a subgroup of $\mathrm{SPL}_2(2,\mathbb{C})$. The group G must act properly discontinuously on the Riemann surface M_0 which is the body of M (i.e., the surface described locally by the ordinary complex coordinate z_0, the non-nilpotent part of z). This ensures that M can be endowed with the DeWitt topology (trivial in the soul directions) which is necessary for physics applications. Furthermore, as in the $N = 1$ case,[3] we demand that G leave the "metric" defined on the covering space invariant (the "metrizability" condition). We will be interested only in SC_2 as the covering space, and the metric (more precisely, the norm of the even component of the supergravity frame field) in this case is

$$ds = |dz + \theta^T d\theta|. \tag{2.8}$$

The group G is isomorphic to $\pi_1(M) = \pi_1(M_0)$ and is unique up to a conjugation $G \to \hat{G} = h^{-1}Gh$, where h is an arbitrary superconformal transformation such that $\hat{G}$ still satisfies the above conditions imposed on G.

III. $N=2$ SUPERTORI

We now concentrate on supertori, $N = 2$ SRS's of genus 1. Every supertorus can be obtained as the quotient of SC_2 by an Abelian group G, generated by two transformations of the form (2.6). The metrizability condition, i.e., the invariance of (2.8) under G, requires both generators to have $c = 0$, $a/d = 1 = ad$, and $\gamma = 0$. Without loss of generality we can set $a = d = 1$, since the other choice $a = d = -1$ accompanied by certain b, δ, and O, is equivalent to having $a = d = 1$ and $-b$, $-\delta$, $-O$ instead [this is true due to the fact that $\det O = \det(-O)$ in our $N = 2$ case, contrary to what one has when $N = 1$,[3] for example]. We further want G to act properly discontinuously on the body of the supertorus and so b must have a nonvanishing body, $b_0 \neq 0$. We conclude that each one of the two generators of G induces a transformation of the form

$$\tilde{z}(z,\theta) = z + b + \theta^T O\delta \quad (b_0 \neq 0),$$
$$\tilde{\theta}(z,\theta) = O^T\theta + \delta, \tag{3.1}$$

which is a $N = 2$ supersymmetry transformation ("supertranslation") accompanied by a global rotation in the θ plane, with or without reflection. The Lie superalgebra of the supergroup of transformations of the form (3.1) is generated by L_{-1}, T_0, and $G^i_{-1/2}$ $(i = 1,2)$. Notice that z-dependent rotations were excluded here also by the metrizability condition, since by (2.5) $O(z)$ is either of the form

$$R(r(z)) = \begin{pmatrix} \cos r(z) & -\sin r(z) \\ \sin r(z) & \cos r(z) \end{pmatrix}, \quad \det R = +1, \tag{3.2}$$

or of the form

$$M(r(z)) \equiv R(r(z))\mathbb{I},$$
$$\text{where } \mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \det M = -1, \tag{3.3}$$

and so

$$dz+\theta^T d\theta \to \left(1+2\theta^1\theta^2\frac{dr}{dz}\right)dz+\theta^T d\theta \tag{3.4}$$

under (3.1).

Representing the transformation (3.1) by an ordered triplet (O,b,δ), the commutator of two generators $A=(O,b,\delta)$ and $A'=(O',b',\delta')$ is given by

$$\begin{aligned} &A'^{-1}A^{-1}A'A \\ &\quad=(OO'O^TO'^T,\delta^T(\mathbf{1}+O^T+O'-O'O^T)\delta' \\ &\qquad+\delta'^TO\delta'-\delta^TO'\delta,O'[OO'^T\delta \\ &\qquad+O(\delta'-\delta)-\delta']). \end{aligned} \tag{3.5}$$

The b term of the resulting transformation is a pure soul (nilpotent), showing that if A and A' do not commute the DeWitt topology would be violated. [Since the commutator (3.5) belongs to the group G by which we take the quotient of the superplane, we are led to identification of points whose z coordinates differ only in soul. This implies a nontrivial topology in the soul directions as opposed to the trivial, noncompact DeWitt topology.]

In order to find and classify all the commuting generator pairs we have to distinguish between four different general cases. First, without loss of generality, we order the pair (A,A') in such a way that

$$\text{Im}\,\tau>0,\ \tau\equiv b/b'. \tag{3.6}$$

We now employ the following notation[5]: we say that the supertorus that is the quotient of SC_2 by G generated by (A,A') is of the type

$$\begin{aligned} &\text{UU if det } O=1 \text{ and det } O'=1; \\ &\text{UT if det } O=1 \text{ and det } O'=-1; \quad \text{U, Untwisted,} \\ &\text{TU if det } O=-1 \text{ and det } O'=1; \quad \text{T, Twisted,} \\ &\text{TT if det } O=-1 \text{ and det } O'=-1. \end{aligned} \tag{3.7}$$

For ensuring commutativity of A and A' we need first of all

$$OO'O^TO'^T=\mathbf{1}. \tag{3.8}$$

Equation (3.8) is automatically satisfied in the UU case because SO(2) is Abelian, and so in this case O and O' are of the form $R(r)$ and $R(r')$ [see (3.2)] with r, r' arbitrary bosonic constants. In the UT case we have $O=R(r)$ and $O'=M(r')$ [see (3.3)], and (3.8) holds only when $r=0$ or π, i.e., $O=\pm\mathbf{1}$, while r' is arbitrary. Similarly, in the TU case $O=M(r)$ with arbitrary r while $O'=\pm\mathbf{1}$, and finally in the TT case $O=M(r)$ with arbitrary r and we must have $O'=\pm O$. The other two commutativity conditions, namely,

$$\delta^T(\mathbf{1}+O^T+O'-O'O^T)\delta'+\delta'^TO\delta'-\delta^TO'\delta=0, \tag{3.9}$$

$$OO'^T\delta+O(\delta'-\delta)-\delta'=0, \tag{3.10}$$

will be studied below separately in each case.

We now use conjugations by superconformal transformations in order to bring the pair of generators in each case to a canonical form. This will allow us to classify all the superconformally inequivalent supertori and so describe the super-Teichmuller space $ST_{g=1,N=2}$ of $N=2$ SRS's of genus 1. Details about the specific conjugations used can be found in Appendix A; here we only refer to their types as listed there. We start with the simpler cases where at least one of the generators in the pair is "twisted."

UT case: We have to distinguish once more between two different subcases.

(I) If $O=-\mathbf{1}$ we can use conjugations of the types (A2) and (A3) to get a pair of the form $A(-\mathbf{1},\tau,0)$ and $A'(\mathbb{I},1,\delta')$, whereupon (3.10) requires $\delta'=0$. The canonical form for the generator pair in this case is therefore

$$\text{UTI: } A(-\mathbf{1},\tau,0) \text{ and } A'(\mathbb{I},1,0), \tag{3.11}$$

where $\text{Im}\,\tau>0$, otherwise τ arbitrary.

(II) If $O=\mathbf{1}$, conjugations (A2) and (A3) bring the pair of generators to the form $A(\mathbf{1},\tau,\binom{\delta_1}{0})$ and $A'(\mathbb{I},1,\binom{\delta_1'}{0})$ with $\delta_1\delta_1'=0$, as required by (3.9) and (3.10). Applying a conjugation (A4) to this case we get the canonical form

$$\text{UTII: } A(\mathbf{1},\tau,\tbinom{\delta}{0}) \quad \text{and} \quad A'(\mathbb{I},1,0), \tag{3.12}$$

with arbitrary τ ($\text{Im}\,\tau>0$) and δ (one fermionic constant here, and not a column vector). Conjugations (A2) or (A3) show that (3.12) with the parameters $(\tau,-\delta)$ or the parameters (τ,δ) gives rise to equivalent supertori. We express this fact by writing $\delta\sim-\delta$, for this UTII case.

TU case: Similarly to the UT case we get two possible canonical forms:

$$\text{TUI: } A(\mathbb{I},\tau,0) \text{ and } A'(-\mathbf{1},1,0), \tag{3.13}$$

$$\text{TUII: } A(\mathbb{I},\tau,\tbinom{\delta}{0}) \quad \text{and} \quad A'(\mathbf{1},1,0), \tag{3.14}$$

with arbitrary τ ($\text{Im}\,\tau>0$) and $\delta\sim-\delta$ in (3.14).

TT case: Here we get

$$\text{TTI: } A(\mathbb{I},\tau,0) \text{ and } A'(-\mathbb{I},1,0), \tag{3.15}$$

$$\text{TTII: } A(\mathbb{I},\tau,\tbinom{\delta}{0}) \quad \text{and} \quad A'(\mathbb{I},1,0), \tag{3.16}$$

with arbitrary τ ($\text{Im}\,\tau>0$) and $\delta\sim-\delta$ in (3.16).

The above three cases are trivial generalizations of what one has in the $N=1$ case.[3] They can be thought of as "tensor products" of the different cases found there (see discussion on the embedded spin structures in Sec. IV). "New" features are found only in the following case.

UU case: We start with a generator pair of the "raw" form $A(R(r),b,\delta)$ and $A'(R(r'),b',\delta')$. Here we treat r and r' as ordinary complex numbers and refer the reader to Appendix B for a discussion of the complications that arise if we allow them to have nilpotent parts.

(I) If r and r' are not both zero (mod 2π) then we can always bring either δ or δ' to zero by conjugation (A2), and then (3.10) requires also the other to vanish. Conjugation (A5) enables us now to get the canonical form

$$\text{UUI: } A(R(r),\tau,0) \text{ and } A'(\mathbf{1},1,0), \tag{3.17}$$

with arbitrary τ ($\text{Im}\,>0$) and r (including $r=2\pi m$, $m\in\mathbb{Z}$). Conjugation (A2) shows the equivalence $(\tau,r)\sim(\tau,-r)$.

We remark that by using a "complexified" basis for the (already complex) θ plane, $\theta^{\pm}=\theta^1\pm i\theta^2$, the action of the transformation A on a superspace coordinate $Z=(z,\theta^{\pm})$ takes the diagonalized form

$$A(R(r),\tau,0):\quad z\to z+\tau,\quad \theta^+\to e^{-ir}\theta^+,\quad \theta^-\to e^{+ir}\theta^-. \tag{3.18}$$

This form suggests the terminology U(1) supertori that is also used to describe the UUI type of supertori.[5] However, one should not be misled by this terminology to think that r is necessarily real.

(II) If $r=r'=0$ then by conjugation (A6) we get the canonical form,

$$\text{UUII: } A(1,\tau,\delta) \text{ and } A'(1,1,0), \tag{3.19}$$

with arbitrary τ ($\mathrm{Im}\,\tau>0$) and $\delta=\binom{\delta_1}{\delta_2}$. Conjugation (A2) shows the equivalence $(\tau,\delta)\sim(\tau,O\delta)$ for *arbitrary* $O\in O(2)$.

The UU case is the only one where the two I and II subcases cross. Namely, the supertori, UUI with $r=0$ mod 2π are equivalent to the supertori UUII with $\delta=0$, for the same τ.

IV. SUPER-TEICHMÜLLER AND SUPERMODULI SPACES

We are now at a position to describe the super-Teichmüller space $ST_{g=1,N=2}$. It consists of seven disconnected pieces, which we denote by ST_{12}^{UTI}, ST_{12}^{TUI}, ST_{12}^{TTI}, ST_{12}^{UTII}, ST_{12}^{TUII}, ST_{12}^{TTII}, and ST_{12}^{UU}, corresponding to the various types of supertori discussed in the previous section. The three pieces representing the UTI, TUI, and TTI types are each a copy of the upper half-plane H with the coordinate τ, i.e., three copies of the ordinary Teichmüller space T_1 of ordinary tori. The three pieces representing the UTII, TUII, and TTII types are each a copy of the disconnected piece $ST_{11}^{+\,+}$ of $ST_{g=1,N=1}$ representing tori with the trivial spin structure[3]; it is obtained from $\mathbb{C}^{1,1}$ with coordinates (τ,δ) by restricting τ to lie in H and identifying the points $(\tau,-\delta)$ with (τ,δ). The seventh piece can be thought of being composed out of two different pieces that cross each other. The first, representing the UUI type of supertori, is obtained from $\mathbb{C}^{2,0}$ with coordinates (τ,r) by restricting τ to lie in H and identifying $(\tau,r)\sim(\tau,\pm r+2\pi n)$ for any $n\in\mathbb{Z}$. This identification gives this piece the topology of $T_1\times(\mathbb{Z}_2$-orbifold of a cylinder$)\cong T_1\times(\mathbb{C}^*/\mathbb{Z}_2)$, where $\mathbb{C}^*=\mathbb{C}-\{0\}$; the coordinates on $T_1\times(\mathbb{C}^*/\mathbb{Z}_2)$ are $(\tau,\rho\equiv e^{ir})$ and the $\mathbb{Z}_2$-orbifold structure is obtained by identifying $(\tau,\rho^{-1})\sim(\tau,\rho)$. The second piece, that corresponds to UUII supertori, is obtained from $\mathbb{C}^{1,2}$ with coordinates (τ,δ_+,δ_-), where $\delta_\pm=\delta_1\pm i\delta_2$, by restricting τ to lie in H and identifying $(\tau,\delta_+,\delta_-)\sim(\tau,\delta_-,\delta_+)\sim(\tau,k\delta_+,k^{-1}\delta_-)$ for any invertible bosonic constant k. The two pieces are connected through the identification of points $(\tau,r=0)$ of the first with points $(\tau,\delta_+=0,\delta_-=0)$ of the second, for the same τ.

Note that the Grassmannian nature of δ_+, δ_- makes it difficult to analyze the topology of the fermionic piece of ST_{12}^{UU} and find admissible coordinate charts on it. Such analysis is essential for endowing this space with a supermanifold structure, which one expects it to have. Were δ_+, δ_- ordinary complex numbers, this sector would have been $\mathbb{C}^2/\sim$, where $\sim$ is the above equivalence relation. Now, $\mathbb{C}^2/\sim$ is isomorphic to the complex plane with an additional point P "above" the origin; the open sets (in the topology induced from the standard one on $\mathbb{C}^2$) are the usual open sets in $\mathbb{C}$, and also $P\cup[N(0)-\{0\}]$, where $N(0)$ is any open neighborhood of the origin in $\mathbb{C}$. This result is obtained by noting that two pairs of complex numbers are equivalent under $\sim$ iff the products within each pair are equal, unless they both vanish. For pairs of fermionic numbers this statement holds in one direction only, because of the noninvertibility of such numbers. Here $\mathbb{C}^2/\sim$ is a non-Hausdorff manifold of one complex dimension, a dimension smaller by one than that of $\mathbb{C}^2$, the space we started with. This last feature does not seem to be carried on to the Grassmannian case, where we deal with $\mathbb{C}^{0,2}/\sim$.

In order to find the supermoduli space $SM_{g=1,N=2}$ of $N=2$ supertori we have to describe the action of the supermodular group on $ST_{1,2}$. The supermodular group is generated by two transformations that can be realized as acting on generator pairs (A,A') that characterize supertori,

$$S:(A,A')\to(AA',A'),\quad T:(A,A')\to(A'^{-1},A). \tag{4.1}$$

Starting from a canonical pair (A,A'), that corresponds to a point of $ST_{1,2}$, we transform it according to (4.1) and conjugate the resulting generator pair to bring it back to a canonical form, from which the transformed point of $ST_{1,2}$ can be read off. It turns out that S sends the τ coordinate to $\tau+1$, leaving unchanged all other coordinates (when present). However, in some cases it transfers points from one disconnected piece of $ST_{1,2}$ to another, according to the following assignment:

$$\begin{array}{lll}
\text{UTI (II)}\to\text{TTI (II)}, & \tau\to\tau+1, & (\delta\to\delta);\\
\text{TUI (II)}\to\text{TUI (II)}, & \tau\to\tau+1, & (\delta\to\delta);\\
S:\ \text{TTI (II)}\to\text{UTI (II)}, & \tau\to\tau+1, & (\delta\to\delta);\\
\text{UUI}\to\text{UUI}, & \tau\to\tau+1, & r\to r;\\
\text{UUII}\to\text{UUII}, & \tau\to\tau+1, & (\delta_+,\delta_-)\to(\delta_+,\delta_-).
\end{array} \tag{4.2}$$

In a similar notation

$$\begin{array}{lll}
\text{UTI (II)}\to\text{TUI (II)}, & \tau\to-1/\tau, & (\delta\to\tau^{-3/2}\delta);\\
\text{TUI (II)}\to\text{UTI (II)}, & \tau\to-1/\tau, & (\delta\to\tau^{-3/2}\delta);\\
T:\ \text{TTI (II)}\to\text{TTI (II)}, & \tau\to-1/\tau, & (\delta\to\tau^{-3/2}\delta);\\
\text{UUI}\to\text{UUI}, & \tau\to-1/\tau, & r\to r/\tau;\\
\text{UUII}\to\text{UUII}, & \tau\to-1/\tau, & (\delta_+,\delta_-)\to\tau^{-3/2}(\delta_+,\delta_-).
\end{array} \tag{4.3}$$

The sign ambiguity in $\tau^{-3/2}$ is immaterial since we anyway identify $-\delta$ with δ.

$SM_{1,2}$ is obtained by taking the quotient of $ST_{1,2}$ by the supermodular group, generated by S and T. The resulting space consists of three disconnected pieces which we denote by SM_{12}^{I}, SM_{12}^{II}, and SM_{12}^{UU}.

SM_{12}^{I} is a three-sheeted cover of the ordinary $g=1$ moduli space, i.e., of $M_1 = H/SL_2(\mathbb{Z}) = \{\tau \in H \mid -\frac{1}{2} \leqslant \mathrm{Re}\,\tau \leqslant \frac{1}{2},\ |\tau| \geqslant 1\}$ with the standard identification of the edges (a point in SM_{12}^{I} is therefore specified by one bosonic supermodulus). The edges of the covering sheets are identified according to the scheme shown in Fig. 1.

SM_{12}^{II} is obtained from the product space $SM_{12}^{I} \times \mathbb{C}^{0,1}$, which is a trivial fiber bundle of fiber coordinate δ over SM_{12}^{I}, by identifying $\delta \sim i\delta$ in each fiber. This last identification, which extends $\delta \sim -\delta$ that we had already in the super-Teichmüller space, can be inferred from the action of T^2 on ST_{12}^{II}; it gives the fiber a structure of a $\mathbb{Z}_4$-orbifold which is singular at $\delta = 0$. On SM_{12}^{II} we have two supermoduli, one bosonic and one fermionic.

SM_{12}^{UU} consists of two connected pieces originating from the two pieces of ST_{12}^{UU}. A point in the first piece is specified by a pair of bosonic supermoduli (τ, r), where $\tau \in M_1$, and in addition to the identification $(\tau, r) \sim (\tau, \pm r + 2\pi n)$ we have also $(\tau, r) \sim (\tau, \pm r + 2\pi n\tau)$ for any $n \in \mathbb{Z}$. The new equivalence arises from the old one when combined with two T operations: $(\tau, r) \sim (-1/\tau, r/\tau) \sim (-1/\tau, \pm r/\tau + 2\pi n) \sim (\tau, \pm r + 2\pi n\tau)$. We conclude that for a given τ, r lies on the $\mathbb{Z}_2$-orbifold obtained by taking the quotient of a torus of periods 2π and $2\pi\tau$ by the equivalence relation $r \sim -r$. This orbifold has the topology of a sphere with four singular points [at $r = 0$, π, $\pi\tau$, and $\pi(1+\tau)$] where the curvature diverges.[12] We can arrive at the same result taking an algebraic point of view.[5] The UUI supertori can be thought of as a pair of conjugate line bundles [of fiber coordinates θ^+ and θ^-—see (3.18)] over an ordinary torus, which are both of zero degree since their transition functions $e^{\pm ir}$ are constants.[13] Therefore, r is essentially a coordinate on the Picad torus $\mathrm{Pic}_0(T(\tau))$ of the torus characterized by τ, which is indeed a conformally equivalent torus of the same period ratio, in our $g=1$ case (our discussion actually constitutes an alternative derivation of this classical result in algebraic geometry). However, since the order within the pair of conjugate line bundles characterizing a UUI supertorus is immaterial, basically because θ^+ and θ^- are interchangeable [e.g., via conjugation (A2)], we have to take the above $\mathbb{Z}_2$ quotient of $\mathrm{Pic}_0(T(\tau))$ in order to get a description of inequivalent supertori. The second piece of SM_{12}^{UU} can be written as $M_1 \times (\mathbb{C}^{0,2}/\sim)$, where $\sim$ identifies points (δ_+, δ_-) in $\mathbb{C}^{0,2}$ according to $(\delta_+, \delta_-) \sim (\delta_-, \delta_+) \sim i(\delta_+, \delta_-) \sim (k\delta_+, k^{-1}\delta_-)$ for any invertible bosonic k. The second equivalence was obtained by considering the action of T^2 on ST_{12}^{UU}. The connection between the two pieces of SM_{12}^{UU} is induced from the connectedness of ST_{12}^{UU}. One can think of the continuous transition from the first piece to the second as being accompanied by a trade in of the bosonic supermodulus r by two (not completely independent) fermionic supermoduli (δ_+, δ_-).

We close this section with a discussion of the relation between $N=1$ and $N=2$ supertori. $N=1$ supertori are naturally described by spin bundles of fiber coordinate θ over ordinary tori,[3] having one of four possible spin structures. The spin structures specify the way by which a fiber above a point of the torus is identified with itself after being transported back to the same point along the two cycles of the torus. In terms of the fiber coordinate θ, they specify whether it changes sign or not when we go along each cycle. In the $N=2$ supertori a two-dimensional fiber of coordinates θ^i ($i = 1,2$ or $i = +, -$) replaces the one-dimensional spin bundle. One can contemplate eliminating the two θ coordinates one at a time in order to find the "embedded spin structures" in the $N=2$ supertori. For this procedure to make sense, the elimination must leave a single θ that does not mix with the one we eliminated under translations along the cycles of the torus, i.e., we must perform the elimination using supetori that are described by a "diagonalized" basis of the θ plane. The generator pairs (A, A') in their canonical forms (3.11)–(3.16), (3.18), (3.19) give us precisely the desired representation, if we use the $\theta^{1,2}$ basis in (3.11)–(3.16) and $\theta^{\pm}$ in (3.18) [either basis in (3.19)]. We get the following embedded spin structures for the various types of $N=2$ supertori (the pairs of signs specify the sign flips of the remaining θ coordinate under translations along the 1 and τ cycles, respectively):

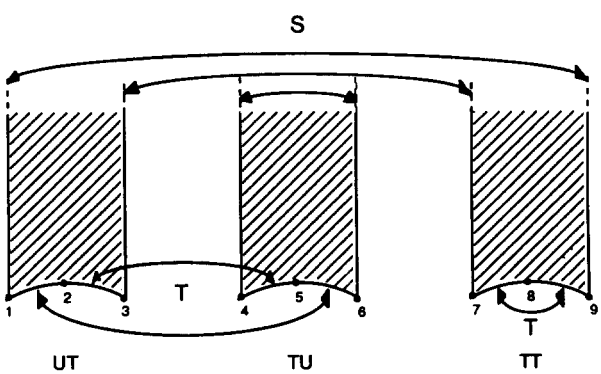


FIG. 1. The construction of SM_{12}^{I} from three covering sheets of M_1 with identifications of their edges (Sec. IV). Notice, in particular, the resulting identifications of the points $1 \sim 3 \sim 4 \sim 6 \sim 7 \sim 9$ and $2 \sim 5$.

Type	Embedded spin structure eliminating θ^2 (θ^-)	eliminating θ^1 (θ^+)
UTI	+ −	− −
TUI	− +	− −
TTI	− +	+ −
UTII	+ +	− +
TUII	+ +	+ −
TTII	+ +	− −
UUI ($r=0$)	+ +	+ +
UUI ($r=\pi$)	+ −	+ −
UUII	+ +	+ +.

(4.4)

Note that the notion of embedded spin structure makes sense in the UUI case only when $r = 0$ or π [see (3.18)]. Different spin structures can be "connected" by moving along continuous paths in $SM_{1,2}$, the parametrization space of $N = 2$ supertori. In this way we find the spin structure content of the three disconnected components of $SM_{1,2}$ to be

$$\begin{array}{lccc} & SM_{12}^{I} & SM_{12}^{II} & SM_{12}^{UU} \\ \text{spin structure content:} & + -, - +, - - & + +, + -, - +, - - & + +, + - \end{array}. \tag{4.5}$$

In the $N = 1$ case,[4] it is possible to connect only the nontrivial spin structures $+ -$, $- +$, $- -$, whereas the trivial one $+ +$ is isolated. We see that the presence of two fermionic coordinates and the freedom to rotate one to the other, that we have in the $N = 2$ case, enables one to connect all the possible spin structures (in the SM_{12}^{II} piece of the supermoduli space).

V. SUPERELLIPTIC FUNCTIONS

In this section we construct meromorphic functions on the supertori. They correspond to meromorphic functions on the covering space $\mathbb{C}^{1,2}$ that are invariant under the generator pair (A,A') characterizing the supertorus in question, and we shall refer to them as "superelliptic." We have to consider each type of supertori separately.

UTI: Any meromorphic function on $\mathbb{C}^{1,2}$, depending implicitly on τ, can be expanded in the form

$$\begin{aligned} \mathfrak{R}(Z;\tau) &= F_0(z;\tau) + \theta^1 F_1(z;\tau) \\ &\quad + \theta^2 F_2(z;\tau) + \theta^1\theta^2 F_3(z;\tau), \end{aligned} \tag{5.1}$$

with F_i $(i = 0,1,2,3)$ meromorphic functions on $\mathbb{C}^{1,0}$. For being superelliptic in the UTI case, $\mathfrak{R}(Z;\tau)$ must be invariant under both transformations (3.11),

$$\begin{aligned} &A: \quad z \to z + \tau, \quad \theta^1 \to -\theta^1, \quad \theta^2 \to -\theta^2; \\ &A': \quad z \to z + 1, \quad \theta^1 \to \theta^1, \quad \theta^2 \to -\theta^2. \end{aligned} \tag{5.2}$$

Therefore, the functions F_i must change sign when translated along the two cycles of the underlying ordinary torus in the following way:

$$\begin{array}{lccc} & & z \to z+1 & z \to z+\tau \\ & F_0 & + & + \\ \text{UTI:} & F_1 & + & - \\ & F_2 & - & - \\ & F_3 & - & + \end{array}.$$

We see that F_0 must be an ordinary elliptic function, whereas F_1, F_2, and F_3 are meromorphic sections of the spin bundles over the torus, with $+ -$, $- -$, and $- +$ spin structures, respectively. Using the results obtained in Ref. 4 we conclude that these functions can always be written in the form

$$\begin{aligned} \text{UTI: } F_0(z;\tau) &= \wp_0(z;\tau)\cdot(\text{an arbitrary elliptic function}), \\ F_1(z;\tau) &= \wp_1(z;\tau)\cdot(\text{an arbitrary elliptic function}), \\ F_2(z;\tau) &= \wp_3(z;\tau)\cdot(\text{an arbitrary elliptic function}), \\ F_3(z;\tau) &= \wp_2(z;\tau)\cdot(\text{an arbitrary elliptic function}), \end{aligned} \tag{5.3}$$

where $\wp_0(z;\tau) \equiv 1$ and

$$\begin{aligned} \wp_r(z;\tau) &\equiv \sqrt{\wp(z;\tau) - \wp(\omega_r;\tau)} \\ &= \frac{\vartheta'(0;\tau)\vartheta_r(z;\tau)}{\vartheta_r(0;\tau)\vartheta(z;\tau)} \quad (r = 1,2,3). \end{aligned} \tag{5.4}$$

Here $\wp(z;\tau)$ is the Weierstrass function of periods 1 and τ, and $\omega_1 = \frac{1}{2}$, $\omega_2 = \tau/2$, $\omega_3 = (1+\tau)/2$; the ϑ functions (not to be confused with the fermionic coordinates θ) are as in Ref. 14.

UTII: Here it is more convenient to write the generic superelliptic function in the form

$$\mathfrak{R}(Z;\tau,\delta) = E_1(z,\theta^1;\tau,\delta) + \theta^2 E_2(z,\theta^1;\tau,\delta). \tag{5.5}$$

Then the invariance under the pair of transformations (3.12),

$$\begin{aligned} &A: \quad z \to z + \tau + \theta^1\delta, \quad \theta^1 \to \theta^1 + \delta, \quad \theta^2 \to \theta^2; \\ &A': \quad z \to z + 1, \quad \theta^1 \to \theta^1, \quad \theta^2 \to -\theta^2, \end{aligned} \tag{5.6}$$

forces E_1 to be a $N = 1$ superelliptic function (of the trivial spin structure $+ +$),[4] that can be expressed as a rational function of

$$\mathfrak{R}_{N=1}(z,\theta^1;\tau,\delta) = \wp(z;\tau + \theta^1\delta) \tag{5.7}$$

and the two derivatives $D_1\mathfrak{R}_1$, $D_1^2\mathfrak{R}_1 = \partial\mathfrak{R}_1/\partial z$. E_2 can be expressed as $\wp_2(z;\tau + \theta^1\delta)$ times some $N = 1$ superelliptic function.

A similar treatment of the TU and TT cases brings us to the following summary of superelliptic functions for the twisted supertori:

$$\begin{aligned} \text{I subcases: } \mathfrak{R}(Z;\tau) &= F_0(z;\tau) + \theta^1 F_1(z;\tau) + \theta^2 F_2(z;\tau) \\ &\quad + \theta^1\theta^2 F_3(z;\tau). \end{aligned} \tag{5.8a}$$

$$\text{II subcases: } \mathfrak{R}(Z;\tau,\delta) = E_1(z,\theta^1;\tau,\delta) + \theta^2 E_2(z,\theta^1;\tau,\delta), \tag{5.8b}$$

with

$$F_i(z;\tau) = \wp_r(z;\tau)\cdot f(z;\tau), \tag{5.8c}$$

$$\begin{aligned} E_j(z,\theta^1;\tau,\delta) &= \wp_s(z;\tau + \theta^1\delta)\cdot Sf(z,\theta^1;\tau,\delta) \\ &\equiv \mathfrak{R}_{1,s}(z,\theta^1;\tau,\delta)\cdot Sf(z,\theta^1;\tau,\delta), \end{aligned} \tag{5.8d}$$

where $f(z;\tau)$ is an arbitrary elliptic function, $Sf(z,\theta^1;\tau,\delta)$ an arbitrary $N = 1$ ($+ +$) superelliptic function, and the values for r and s are assigned according to the following tables:

i	r assignment UT	TU	TT	j	s assignment UT	TU	TT
0	0	0	0	1	0	0	0
1	1	2	2	2	2	1	3
2	3	3	1				
3	2	1	3				

(5.8e)

In the I subcases the functions $\wp$, $\wp'$, and $\wp_r$ $(r=1,2,3)$ form a complete set of superelliptic functions, in the sense that all the functions appearing as coefficients of 1, θ^1, θ^2, and $\theta^1\theta^2$ in the generic form (5.8a) can be expressed as rational functions of them. Similarly, the functions $\mathfrak{R}_1$, $D_1\mathfrak{R}_1$, $\partial\mathfrak{R}_1/\partial z$, and one appropriate $\mathfrak{R}_{1,s}$ [according to (5.8e)], form a complete set of superelliptic functions for the II subcases.

UUI: In this case we write the superelliptic functions in the form

$$\mathfrak{R}(Z;\tau,r)=F_0(z;\tau,r)+\theta^+F_-(z;\tau,r)+\theta^-F_+(z;\tau,r)+\theta^+\theta^-F_3(z;\tau,r), \tag{5.9}$$

and then the invariance under the two transformations in (3.17), namely,

$$\begin{aligned} &A:\quad z\to z+\tau,\quad \theta^+\to e^{-ir}\theta^+,\quad \theta^-\to e^{+ir}\theta^-,\\ &A':\quad z\to z+1,\quad \theta^+\to\theta^+,\quad \theta^-\to\theta^-, \end{aligned} \tag{5.10}$$

implies that

$$\begin{aligned} &F_0(z+1;\tau,r)=F_0(z;\tau,r)=F_0(z+\tau;\tau,r),\\ &F_+(z+1;\tau,r)=F_+(z;\tau,r)=e^{+ir}F_+(z+\tau;\tau,r),\\ &F_-(z+1;\tau,r)=F_-(z;\tau,r)=e^{-ir}F_-(z+\tau;\tau,r),\\ &F_3(z+1;\tau,r)=F_3(z;\tau,r)=F_3(z+\tau;\tau,r). \end{aligned} \tag{5.11}$$

Thus F_0 and F_3 must be ordinary elliptic functions (possibly r dependent), whereas $F_\pm$, although unchanged under $z\to z+1$, gain constant phases under $z\to z+\tau$. The latter are therefore meromorphic sections of line bundles of zero degree over the torus, and we can construct them by taking the quotient of two theta functions[14] with displaced arguments. A particular form, which will be used later, for the desired F_+ is

$$\Phi(z;\tau,r)\equiv\vartheta(z+r/2\pi;\tau)/\vartheta(z;\tau), \tag{5.12}$$

and for F_- we take $\Phi(z;\tau,-r)$. Any other F_+ (or F_-) satisfying (5.11) can be expressed as the particular choice Φ multiplied by some elliptic function. This $\Phi(z;\tau,r)$ has simple poles and simple zeros at $z=0$ and $-r/2\pi \bmod(1,\tau)$, respectively. When r approaches $2\pi(m+n\tau)$ $(m,n\in\mathbb{Z})$ (the lattice points defining the Picard torus) the poles and zeros cancel each other and Φ becomes entire:

$$\Phi(z;\tau,r=2\pi(m+n\tau))=\vartheta(z+m+n\tau;\tau)/\vartheta(z;\tau)=(-1)^{m+n}e^{-\pi in^2\tau}e^{-2\pi inz}. \tag{5.13}$$

Only for these special values of r do there exist entire functions $F_\pm$ satisfying (5.11), namely (5.13) up to a constant. Notice also that as r varies from 0 to π, Φ interpolates between 1 and $\wp_1(z;\tau)$ (up to constants), which are sections of the trivial line bundle and the $(+\ -)$-spin bundle over the torus, respectively.[15] Together $\wp$, $\wp'$, and $\Phi(z;\tau,\pm r)$ form a complete set of superelliptic functions on the UUI supertori; Φ can be thought of as the "supersymmetric partner" of the Weierstrass function $\wp$ in this case.

UUII: The superelliptic functions in this case naturally generalize those of the $N=1$ $(+\ +)$ supertori. The typical function is

$$\mathfrak{R}(Z;\tau,\delta_1,\delta_2)=\wp(z;\tau+\theta^T\delta), \tag{5.14}$$

which is clearly invariant under the two transformations in (3.19). All the covariant derivatives of $\mathfrak{R}$, obtained by applying $D_i=\partial/\partial\theta^i+\theta^i(\partial/\partial z)$ $(i=1,2)$ to it, are superelliptic as well. The first few examples, with their expansions in the fermionic coordinates θ^i, are listed below:

$$\mathfrak{R}=\wp+\theta^T\delta\dot\wp-\theta^1\theta^2\delta_1\delta_2\ddot\wp, \tag{5.15a}$$

$$D\mathfrak{R}=\delta\dot\wp+(\wp'-\delta_1\delta_2\dddot\wp\epsilon)\theta+\theta^1\theta^2(\epsilon\delta)\ddot\wp' \quad\text{(matrix notation!)}, \tag{5.15b}$$

$$D_1D_2\mathfrak{R}=\delta_1\delta_2\ddot\wp+(\theta^T\epsilon\delta)\dot\wp+\theta^1\theta^2\wp'', \tag{5.15c}$$

$$\mathfrak{R}'=D_1^2\mathfrak{R}=D_2^2\mathfrak{R}=\wp'+\theta^T\delta\dot\wp'-\theta^1\theta^2\delta_1\delta_2\ddot\wp', \tag{5.15d}$$

where the prime and the dot denote partial differentiation with respect to z and τ, respectively, and on the rhs's $\wp$ and its derivatives are always evaluated at $(z;\tau)$. It is shown in Appendix B that $\mathfrak{R}$, $D_1\mathfrak{R}$, $D_2\mathfrak{R}$, and $\mathfrak{R}'$ form a complete set of superelliptic functions for the UUII supertori.

VI. SUPERTORI AS ALGEBRAIC CURVES

In this section we derive differential equations, relating superelliptic functions to their derivatives, and use them to embed the $N=2$ supertori into projective superspaces. In addition to the two UU cases we treat only the UTI and UTII cases as representing the twisted sector (the other cases can be handled in similar ways).

UTI: This case is very similar to that of $N=1$ supertori with the nontrival spin structures (namely $+\ -$ in θ^1 and $-\ -$ in θ^2). The final result is obtained immediately from Ref. 4. The supertorus, characterized by the single bosonic supermodulus τ, is embedded into the projective superspace[4] $\mathbb{P}^{2,6}$ by the map

$$\begin{aligned}(z,\theta^1,\theta^2)&\to[x,y,1;\phi_1,\psi_1,\zeta_1,\phi_3,\psi_3,\zeta_3]\\&=[\wp(z),\wp'(z),1;\theta^1\wp_1(z),\theta^1\wp_1'(z),\theta^1\wp\wp_1(z),\\&\quad\theta^2\wp_3(z),\theta^2\wp_3'(z),\theta^2\wp\wp_3(z)].\end{aligned} \tag{6.1}$$

The image of the supertorus is described by seven polynomial equations

$$y^2=4x^3-g_2x-g_3, \tag{6.2a}$$

$$2(x-e_1)\psi_1=y\phi_1,\quad 2(x-e_3)\psi_3=y\phi_3, \tag{6.2b}$$

$$y\psi_1=2(x-e_2)(x-e_3)\phi_1,$$

$$y\psi_3=2(x-e_1)(x-e_2)\phi_3, \tag{6.2c}$$

$$\zeta_1=x\phi_1,\quad \zeta_3=x\phi_3, \tag{6.2d}$$

where $e_r=\wp(\omega_r;\tau)$, $g_2(\tau)=60\Sigma'(m+n\tau)^{-4}$, $g_3(\tau)=140\Sigma'(m+n\tau)^{-6}$ are τ-dependent constants [Σ' denotes summation over all integer pairs $(m,n)\neq(0,0)$]. They satisfy the relations

$$e_r\neq e_s,\quad\text{for}\quad r\neq s\quad(r,s=1,2,3), \tag{6.3a}$$

$$e_1+e_2+e_3=0, \tag{6.3b}$$

$$g_2=-4(e_1e_2+e_2e_3+e_3e_1)=2(e_1^2+e_2^2+e_3^2), \tag{6.3c}$$

$$g_3=4e_1e_2e_3. \tag{6.3d}$$

Equations (6.2c) are necessary only for the embedding of the points $z=\omega_r$ $(r=1,3)$, forcing ϕ_r to vanish there; Eqs. (6.2b) fail to set the constraints which are required by (6.1).

Notice that the points $z=0$ are embedded according to

$$(0,\theta^1,\theta^2)\to(0,1,0;0,0,-\theta^1/2,0,0,-\theta^2/2). \tag{6.4}$$

The superconformal structure of the supertorus can be specified by a rational, meromorphic one-form on $\mathbb{P}^{2,6}$ that agrees with the "metric" $dz+\theta^T d\theta$ on the image of the supertorus. One such one-form is

$$\frac{dx}{y}+\frac{\phi_1\,d\phi_1}{x-e_1}+\frac{\phi_3\,d\phi_3}{x-e_3}. \tag{6.5}$$

UTII: As in the previous case, we make small variations on results that were obtained in Ref. 4 in order to get the differential equations relating different superelliptic functions that are used for the embedding. The supertorus is embedded this time into $\mathbb{P}^{3,5}$ via the map

$$\begin{aligned}(z,\theta^1,\theta^2)&\to[x,y,u,1;\phi,\psi,\zeta,\eta,\xi]\\&=[\Re_1,\Re_1',\Re_1^2,1;D_1\Re_1,D_1\Re_1',\\&\quad\theta^2\Re_{1,2},\theta^2\Re_{1,2}',\theta^2(\Re_{1,2})^2\Re_{1,2}'],\end{aligned} \tag{6.6}$$

where $\Re_1$, $\Re_{1,2}$ [given by (5,7), (5.8d), and (5.4)] and their derivatives are all evaluated at (z,θ^1), and they depend implicitly on the two supermoduli τ,δ. When embedding the points $z=\omega_r$ one must take care and expand all the superelliptic functions in the fermionic coordinates before taking the limits $z\to\omega_r$. In particular

$$\zeta=\theta^2\sqrt{\wp-e_2}(1+\tfrac{1}{2}\theta^1\delta(\dot\wp-\dot e_2)/(\wp-e_2)), \tag{6.7a}$$

$$\begin{aligned}\eta=\theta^2\sqrt{(\wp-e_1)(\wp-e_3)}&[1+\tfrac{1}{2}\theta^1\delta((\dot\wp-\dot e_1)/(\wp-e_1)\\&+(\dot\wp-\dot e_3)/(\wp-e_3))],\end{aligned} \tag{6.7b}$$

$$\begin{aligned}\xi=\frac{1}{2}\theta^2\wp'\sqrt{\wp-e_2}&\left[1+\frac{1}{2}\theta^1\delta\right.\\&\left.\times\left(\frac{\dot\wp-\dot e_1}{\wp-e_1}+2\,\frac{\dot\wp-\dot e_2}{\wp-e_2}+\frac{\dot\wp-\dot e_3}{\wp-e_3}\right)\right].\end{aligned} \tag{6.7c}$$

The image of the supertorus is defined by the equations

$$y^2=4x^3-g_2x-g_3+2\phi\psi, \tag{6.8a}$$

$$2y\psi=(12x^2-g_2)\phi-\delta(\dot g_2x+\dot g_3), \tag{6.8b}$$

$$2[(x-e_2)y+\dot e_2\delta\phi]\eta=y^2\zeta, \tag{6.8c}$$

$$2[(x-e_2)(\dot g_2x+\dot g_3)+2\dot e_2\phi\psi]\eta=y(\dot g_2x+\dot g_3)\zeta, \tag{6.8d}$$

$$y^2\eta=2\{(x-e_1)(x-e_3)y-\delta\phi[\dot e_2(x+2e_2)-\tfrac{1}{4}\dot g_2]\}\zeta, \tag{6.8e}$$

$$\begin{aligned}y(\dot g_2x+\dot g_3)\eta=2\{&(x-e_1)(x-e_3)(\dot g_2x+\dot g_3)\\&-[2\dot e_2(x+2e_2)-\tfrac{1}{2}\dot g_2]\phi\psi\}\zeta,\end{aligned} \tag{6.8f}$$

$$u=x^2, \tag{6.8g}$$

$$\xi=\tfrac{1}{2}y\zeta. \tag{6.8h}$$

Equations (6.8a)–(6.8c), (6.8g), and (6.8h) are sufficient for the embedding except at the points $z=\omega_r$ $(r=1,2,3)$. They fail at these points since they admit solutions (e.g., $x=e_r$, $y=0$, $u=e_r^2$, $\phi=\dot e_r\delta$, $\psi=k\delta$ with arbitrary k, ζ, and η arbitrary, $\xi=0$) that are not the image under (6.6) of any point on the supertorus. Equations (6.8c), (6.8e) are analogous to (6.2b), (6.2c) of the UTI case, and can be obtained from the latter by the "supersymmetric continuation" $\tau\to\tau+\theta^1\delta$. In the derivation we used the relations (6.3) and the identity

$$\theta^1\delta=(D_1\Re_1/\Re_1')\delta. \tag{6.9}$$

Equation (6.9) is in fact the source of the problem we have at the points $z=\omega_r$, where $y=0$, because of the $y=\Re_1'$ in the denominator. Equations (6.8d), (6.8f) are obtained from (6.8c), (6.8e) using (6.8b). They "cure" the problem at $z=\omega_r$ but are not sufficient by themselves because they introduce new (noncoincident!) problems at points where $(\dot g_2x+\dot g_3)$ vanishes. The points $z=0$ are mapped by the above embedding into

$$(0,\theta^1,\theta^2)\to(0,0,1,0;0,6\theta^1,0,0,-\theta^2). \tag{6.10}$$

A rational, meromorphic one-form on $\mathbb{P}^{3,5}$ that agrees with the metric $dz+\theta^T d\theta$ on the image of the supertorus is

$$\frac{y\,dx+\phi\,d\phi}{y^2+\phi\psi}+\frac{y\zeta\,d\xi}{(x-e_2)y+\dot e_2\delta\phi}. \tag{6.11}$$

UUI: In addition to $\wp$ and $\wp'$ we want to use Φ [defined in (5.12)] and possibly its derivatives for embedding UUI supertori into projective superspace, and so we need equations that relate these functions. The theory of elliptic functions enables us to derive such equations (we use the notation and conventions of Ref. 14). Starting from the more symmetric function (whose pole structure, however, is not appropriate for the embedding purpose)

$$\begin{aligned}\tilde\Phi(z;\tau,4\pi u)&\equiv\Phi(z-u;\tau,4\pi u)\\&=\vartheta(z+u;\tau)/\vartheta(z-u;\tau),\end{aligned} \tag{6.12}$$

where we set $u=r/4\pi$, we get an elliptic function

$$\begin{aligned}\frac{\tilde\Phi'(z;\tau,4\pi u)}{\tilde\Phi(z;\tau,4\pi u)}&=\frac{\partial}{\partial z}\ln\vartheta\Big|_{z+u}-\frac{\partial}{\partial z}\ln\vartheta\Big|_{z-u}\\&=\frac{\partial}{\partial z}\ln\sigma\Big|_{z+u}-2\eta_1(z+u)\\&\quad-\left[\frac{\partial}{\partial z}\ln\sigma\Big|_{z-u}-2\eta_1(z-u)\right]\\&=\zeta(z+u;\tau)-\zeta(z-u;\tau)-4\eta_1u\\&=2\zeta(u;\tau)-[\wp'(u;\tau)/(\wp(z;\tau)\\&\quad-\wp(u;\tau))]-4\eta_1u,\end{aligned} \tag{6.13}$$

where $\eta_1=\zeta(\frac{1}{2};\tau)$.[16] For getting an equation that involves Φ we translate z in (6.13) by $+u$, and then use the addition formula for the Weierstrass function $\wp$ and its differential equation (6.2a), to obtain

$$\begin{aligned}\frac{\Phi'(z;\tau,4\pi u)}{\Phi(z;\tau,4\pi u)}&=2\zeta(u)-4\eta_1u-\frac{\wp'(u)}{\frac{1}{4}[(\wp'(z)-\wp'(u))/(\wp(z)-\wp(u))]^2-\wp(z)-2\wp(u)}\\&=2\zeta(u)-4\eta_1u-\frac{2\wp'(u)[\wp(z)-\wp(u)]^2}{\wp''(u)[\wp(z)-\wp(u)]-\wp'(u)[\wp'(z)-\wp'(u)]},\end{aligned} \tag{6.14}$$

suppressing the τ dependence of the rhs. A few observations are in order: (1) Φ'/Φ has simple poles at $z=0$ and $-2u$ mod$(1,\tau)$, and accordingly both sides of (6.14) are infinite there. (2) As it stands, the rhs of (6.14) is not well-defined for $r=4\pi u=0$ mod$(4\pi,4\pi\tau)$ because the functions $\zeta(u)$, $\wp(u)$, $\wp'(u)$ diverge there, but in the limit $u\to 0$ we get a meaningful result and (6.14) still holds. (3) A similar problem is encountered at $z=u$ mod$(1,\tau)$, for a fixed $u\neq 0$ mod$(1,\tau)$, because both the numerator and the denominator on the rhs vanish there. At these points the polynomial version of (6.14), when used as an embedding equation, would not put any constraint on the projective superspace coordinates, and we will have to complement it by an additional equation. As the source of the problem lies in the singularity of the addition formula for $\wp(z+u)$ at the points where it degenerates into the "duplication formula" $\wp(2u)=\frac{1}{4}[\wp''(u)/\wp(u)]^2-2\wp(u)$, we use a different version of it which does not have this flaw,

$$\wp(z+u)=4\left[\frac{\wp^2(z)+\wp(z)\wp(u)+\wp^2(u)-\frac{1}{4}g_2}{\wp'(z)+\wp'(u)}\right]^2-\wp(z)-\wp(u) \tag{6.15}$$

(it is obtained from the original formula by noting that

$$[\wp'(z)]^2-[\wp'(u)]^2=4[\wp(z)-\wp(u)][\wp^2(z)+\wp(z)\wp(u)+\wp^2(u)-\tfrac{1}{4}g_2]).$$

Equation (6.14) can now be rewritten in a form nonsingular at $z=u$ mode $(1,\tau)$ (singular, however, at some other points),

$$\frac{\Phi'(z;\tau,4\pi u)}{\Phi(z;\tau,4\pi u)}=2\zeta(u)-4\eta_1 u - \frac{\frac{1}{2}\wp'(u)[\wp'(z)+\wp'(u)]^2}{\wp''(u)[\wp^2(z)+\wp(z)\wp(u)+\wp^2(u)-\frac{1}{4}g_2]-\wp'(u)[\wp(z)+2\wp(u)][\wp'(z)+\wp'(u)]}. \tag{6.14'}$$

Ignoring for a moment the points $z=0$, a UUI supertorus (characterized by the bosonic supermoduli τ, r) is embedded into $\mathbb{C}^{2,4}$ by

$$\begin{aligned}(z,\theta^+,\theta^-)\to[x,y;\phi_+,\psi_+,\phi_-,\psi_-]&\\ =[\wp(z),\wp'(z);\theta^+\Phi_-(z),\theta^+\Phi'_-(z),&\\ \theta^-\Phi_+(z),\theta^-\Phi'_+(z)],&\end{aligned} \tag{6.16}$$

where $\Phi_\pm=\Phi(z;\tau,\pm r)$. The derivatives $\Phi'_\pm$ are necessary for embedding the points $z=\mp r/2\pi=\mp 2u$ mod$(1,\tau)$ where $\Phi_\pm$ vanish, for $r\neq 0$ mod$(2\pi,2\pi\tau)$. When $r=0$ mod$(2\pi,2\pi\tau)$ we can avoid using them, and in fact embed the supertorus minus the points $z=0$ into the smaller space $\mathbb{C}^{2,2}$, by

$$\begin{aligned}(z,\theta^+,\theta^-)\to[x,y;\phi_+,\phi_-]&\\ =[\wp[(z),\wp'(z);\theta^+e^{2\pi inz},\theta^-e^{-2\pi inz}]&\\ \text{for } r=2\pi(m+n\tau);\ m,n\in\mathbb{Z}].&\end{aligned} \tag{6.17}$$

The embedding equations are in general

$$y^2=4x^3-g_2x-g_3, \tag{6.18a}$$

$$\begin{aligned}\pm 2\wp'(u)[x-\wp(u)]^2\phi_\pm&\\ =\{\psi_\pm\pm[2\zeta(u)-4\eta_1u]\phi_\pm\}&\\ \times\{\wp''(u)[x-\wp(u)]\pm\wp'(u)[y\pm\wp'(u)]\},&\end{aligned} \tag{6.18b}$$

$$\begin{aligned}\pm\tfrac{1}{2}\wp'(u)[y\mp\wp'(u)]^2\phi_\pm&\\ =\{\psi_\pm\pm[2\zeta(u)-4\eta_1u]\phi_\pm\}&\\ \times\{\wp''(u)[x^2+\wp(u)x+\wp^2(u)-\tfrac{1}{4}g_2]&\\ \pm\wp'(u)[x+2\wp(u)][y\mp\wp'(u)]\},&\end{aligned} \tag{6.18c}$$

where $u=r/4\pi$, as before. If (6.17) is used for the special cases $r=2\pi(m+n\tau)$, then only (6.18a) is needed. In order to embed the points $z=0$ as well, we have to add two fermionic coordinates and go to $\mathbb{P}^{2,6}$

$$\begin{aligned}(z,\theta^+,\theta^-)\to[x,y,1;\phi_+,\psi_+,\xi_+,\phi_-,\psi_-,\xi_-]&\\ =[\wp(z),\wp'(z),1;\theta^+\Phi_-(z),&\\ \theta^+\Phi'_-(z),\theta^+\wp\Phi_-(z),\theta^-\Phi_+(z),&\\ \theta^-\Phi'_+(z),\theta^-\wp\Phi_+(z)],&\end{aligned} \tag{6.19}$$

and in addition to (6.18), the image of the supertorus obeys

$$\xi_\pm=x\phi_\pm. \tag{6.20}$$

The points $z=0$ are embedded by (6.19) according to

$$\begin{aligned}(0,\theta^+,\theta^-)\to(0,1,0;0,0,[\vartheta(2u)/2\vartheta'(0)]\theta^+,0,0,&\\ -[\vartheta(2u)/2\vartheta'(0)]\theta^-).&\end{aligned} \tag{6.21}$$

Equation (6.21) is singular for the special cases $r=4\pi u=2\pi(m+n\tau)$, because the theta function vanishes at $m+n\tau$. What happens is that Φ, which generally has a simple pole at the lattice points, becomes analytic there when $r=2\pi(m+n\tau)$, and the additional coordinates needed for the embedding into projective superspace must be obtained from $\theta^\pm\Phi_\mp$ by multiplying them by an elliptic function of higher singularity at $z=0$. Since we are forced to use a different form for the embedding of supertori with the special r values, we might as well extend the simpler form (6.17), instead of (6.16), to projective superspace [in these cases $\psi_\pm$ are proportional to $\phi_\pm$ in (6.16)]. We get a nonsingular embedding into $\mathbb{P}^{2,4}$ by

$$\begin{aligned}(z,\theta^+,\theta^-)\to[x,y,1;\phi_+,\xi_+,\phi_-,\xi_-]&\\ =[\wp(z),\wp'(z),1;\theta^+e^{2\pi inz},&\\ \theta^+e^{2\pi inz}\wp'(z),\theta^-e^{-2\pi inz},\theta^-e^{-2\pi inz}\wp'(z)]&\\ [\text{for } r=2\pi(m+n\pi),m,n\in\mathbb{Z}],&\end{aligned} \tag{6.22}$$

with the equations

$$y^2=4x^3-g_2x-g_3, \tag{6.23a}$$

$$\xi_\pm=y\phi_\pm. \tag{6.23b}$$

As a side (nonsupersymmetric) result of the above discussion we get an explicit algebraic embedding of zero-degree line bundles over the torus into projective space. It is obtained by ignoring one of the $\theta^{\pm}$ coordinates and treating the remaining one as an ordinary, complex local coordinate on the fiber. Nontrivial line bundles $[r \neq 2\pi(m+n\tau)]$ are embedded into $\mathbb{P}^5$ and trivial ones into $\mathbb{P}^4$, where the embedding equations are easily obtained from (6.18)–(6.20) and (6.23), respectively.

A one-form on $\mathbb{P}^{2,6}$ that reduces to the metric on a supertorus that is embedded by (6.19) is

$$\left\{1+\left[\frac{\vartheta'(0)}{\vartheta(2u)}\right]^2\frac{\phi_+\psi_-+\phi_-\psi_+}{2[x-\wp(2u)]}\right\}\frac{dx}{y} - \left[\frac{\vartheta'(0)}{\vartheta(2u)}\right]^2\frac{\phi_+\,d\phi_-+\phi_-\,d\phi_+}{2[x-\wp(2u)]}, \tag{6.24}$$

for its derivation we used the relation

$$\Phi_+(z)\Phi_-(z) = -\left[\frac{\vartheta(2u)}{\vartheta'(0)}\right]^2[\wp(z)-\wp(2u)]. \tag{6.25}$$

A one-form on $\mathbb{P}^{2,4}$ that plays a similar role for the embedding (6.22) is

$$(1+4\pi i n\phi_+\phi_-)(dx/y)+\tfrac{1}{2}(\phi_+\,d\phi_-+\phi_-\,d\phi_+), \tag{6.26}$$

where n is the integer such that $r/2\pi = n\tau \bmod 1$.

UUII: Following the $N=1$ $(++)$ case we embed a UUII supertorus (characterized by the supermoduli τ, δ_1, δ_2) into $\mathbb{P}^{3,4}$ by

$$(z,\theta^1,\theta^2)\to[x,y,u,1;\phi_1,\phi_2,\psi_1,\psi_2] = [\Re,\Re',\Re^2,1;D_1\Re,D_2\Re,D_1\Re',D_2\Re'], \tag{6.27}$$

where $\Re(z,\theta^1,\theta^2;\tau,\delta_1,\delta_2)$ is defined in (5.14). The embedding equations

$$y^2(y^2-4x^3+g_2x+g_3-2\phi^T\psi) + \delta_1\delta_2\phi_1\phi_2(\ddot{g}_2x+\ddot{g}_3)=0, \tag{6.28a}$$

$$y(2y\psi-12x^2\phi+g_2\phi+\delta\dot{g}_2x+\delta\dot{g}_3) - \epsilon[\delta_1\delta_2(\ddot{g}_2x+\ddot{g}_3)\phi-\delta\dot{g}_2\phi_1\phi_2]=0, \tag{6.28b}$$

$$(y^2-4x^3+g_2x+g_3-2\phi^T\psi)\delta=0, \tag{6.28c}$$

$$u=x^2, \tag{6.28d}$$

can be obtained from (5.15), using the Weierstrass equation (6.2a) and its derivatives [see also (B9), (B10)]. Notice that Eqs. (6.28b), (6.28c), are written in matrix notation. Equation (6.28c), obtained from (6.28a), is needed for removing the arbitrariness that is encountered when $y=0$. It is redundant, however, in the trivial case $\delta=0$ for which we have to use different embedding equations [obtained from (6.28) by setting $\delta=0$ and cancelling all overall y factors], or use an altogether different embedding into a smaller space as in (6.22). It seems, without any obvious reason, that we cannot avoid using exceptional embeddings for both the UUI and UUII cases at the trivial point where they cross.

The points $z=0$ are embedded by (6.27) according to

$$(0,\theta^1,\theta^2)\to(0,0,1,0;0,0,6\theta^1,6\theta^2), \tag{6.29}$$

as can be seen from

$$(D\Re'/\Re^2)\,|_{z=0} = 6\theta. \tag{6.30}$$

A rational, meromorphic one-form on $\mathbb{P}^{3,4}$ that reduces to the metric $dz+\theta^T d\theta$ on the image of the supertorus is

$$\frac{2y^3(y\,dx+\phi^T d\phi)+(\dot{g}_2x+\dot{g}_3)\phi_1\phi_2(\delta^T\epsilon\,d\phi)+h(x)\delta_1\delta_2(\phi^T\epsilon\,d\phi)}{2y^3(y^2+\phi^T\psi)+(\dot{g}_2x+\dot{g}_3)\phi_1\phi_2(\delta^T\epsilon\psi)+h(x)\delta_1\delta_2(\phi^T\epsilon\psi)}, \tag{6.31}$$

where h is the rational function such that

$$h(\wp)=2[\ddot{\wp}(\wp')^2-2\dot{\wp}\dot{\wp}'\wp'+\dot{\wp}^2\wp'']. \tag{6.32}$$

The rhs of (6.32) is clearly even in z, and the fact that it is an elliptic function can be verified directly by using

$$\left(\frac{\partial}{\partial\tau}\right)^k\wp\,|_{z+1}=\left(\frac{\partial}{\partial\tau}\right)^k\wp\,|_z,\quad \left(\frac{\partial}{\partial\tau}\right)^k\wp\,|_{z+\tau} = \sum_{j=0}^{k}\binom{k}{j}\left(\frac{\partial}{\partial\tau}\right)^{k-j}\left(-\frac{\partial}{\partial z}\right)^j\wp\,|_z. \tag{6.33}$$

Finally we discuss how the compactification of $\mathrm{SM}_{1,2}$ is achieved by adding to it points with $\tau=i\infty$. The $\mathrm{SM}_{12}^{\mathrm{I}}$ piece is compactified by adding only two distinct points, one for each of the UT and TU covering sheets of M_1 (see Sec. IV). The point at infinity of the TT sheet is identified with that of UT under the action of the generator S of the supermodular group. For the compactification of the $\mathrm{SM}_{12}^{\mathrm{II}}$ piece we use the embedding equations of the corresponding supertori to find, as in Ref. 4, that only two points have to be added also in this case. Recall that $\mathrm{SM}_{12}^{\mathrm{II}}$ is a $\mathbb{Z}_2$-orbifold of $\mathrm{SM}_{12}^{\mathrm{I}}\times\mathbb{C}^{0,1}$, and the compactification is achieved by compactifying the $\mathrm{SM}_{12}^{\mathrm{I}}$ component as described above and identifying all the points along each $\mathbb{C}^{0,1}$ fiber over the two additional points at $\tau=i\infty$. We arrive at this result by noting that the fermionic supermodulus δ disappears from the embedding equations [see, e.g., (6.8)] in the limit $\tau\to i\infty$, and so there are only two singular supertori in that limit (once more the UT and TT ones are identified) and not two families labeled by δ. Essentially the same thing happens when we compactify the $\mathrm{SM}_{12}^{\mathrm{UUII}}$ part of $\mathrm{SM}_{12}^{\mathrm{UU}}$; the (δ_+,δ_-) dependence disappears from the embedding equations as $\tau\to i\infty$ and only one point, or a fiber with all its points (δ_+,δ_-) identified, has to be added at $\tau=i\infty$. The compactification of the $\mathrm{SM}_{12}^{\mathrm{UUI}}$ part has not been worked out yet.

VII. DISCUSSION

In this paper we have studied $N=2$ supertori and some of their algebro-geometric properties. Superconformal equivalence classes of supertori were characterized by a total

number of four parameters (supermoduli), two bosonic and two fermionic. Some parameters vanish identically in each piece of the parameter space which therefore does not contain a four-dimensional component (excluding the possibility of having nilpotent bosonic supermoduli). Most notably, in the presence of two nonvanishing bosonic supermoduli all the fermionic ones can be "gauged away," and the resulting configuration corresponds to a special family of "untwisted" supertori. Spin structures of the ordinary torus were seen to be naturally embedded in $N=2$ supertori, and all the four different spin structures were shown to be continuously connected along paths in the supermoduli space.

Results in the theory of superelliptic functions were obtained. We found sets of a small number of functions in terms of which one can rationally express any meromorphic superfunction on a supertorus of any type. In particular, in the set related to the special untwisted supertori, the fermionic superelliptic functions that join the bosonic Weierstrass function and its derivative were constructed from the ratio of two displaced theta functions. The superelliptic functions were used for embedding the supertori into projective superspace where their image was written as the locus of zeros of sets of polynomials. The superconformal structure on the supertorus was encoded in a rational one-form on projective superspace that reduces on the image of the supertorus to the even component $dz+\theta^T d\theta$ of the supergravity frame field. The embedding equations facilitated the study of the compactification divisor of the genus 1 supermoduli space.

Some mathematical points have still to be clarified.

(1) How can we define admissible coordinate charts, if such exist, on the untwisted sector of super-Teichmüller space. This is important for understanding the supermanifold structure of the supermoduli space and for defining an integration measure on it.

(2) Straighten out the discrepancy between the two apparently different descriptions of the untwisted sector of super-Teichmüller space, as obtained using the physically intuitive approach (Secs. III and IV) and using rigorous superspace theory (Appendix B).

(3) Complete determining the boundary of the untwisted sector of supermoduli space.

More advanced questions in the theory of $N=2$ SRS's are not likely to be answered before further developments of the algebraic geometry of SRS's in the $N=1$ case are made. Some of the open problems are the derivation of an addition law for the superelliptic functions and finding a more natural way of specifying the superconformal structure on the embedded supertori,[4] as well as constructing the supermoduli spaces for higher genera and embedding the corresponding SRS's into projective superspace.

ACKNOWLEDGMENTS

I would like to thank P. Freund and J. Rabin for suggesting the topic and many useful discussions. I also enjoyed helpful conversations with J. Cohn, D. Friedan, and R. Geroch. This work was submitted in partial fulfillment of the requirements for a Ph.D. in Physics at the University of Chicago.

This work was supported by the National Science Foundation, Grant No. PHY85-21588.

APPENDIX A: CONJUGATING SUPERTRANSLATIONS

In this appendix we summarize the conjugations, by various superconformal tranformations h, that are used in Sec. III for bringing generator pairs (A,A') of the form (3.1) to canonical forms

$$(A,A')\to(h^{-1}Ah,h^{-1}A'h). \tag{A1}$$

The resulting pair is required to have the form (3.1), so that it also gives rise, like the original pair (A,A'), to a well-behaved supertorus in the sense described in Sec. III.

The first type of conjugation we consider is performed by transformations of the form (3.1), for which we use the notation introduced in the text.

$$\begin{aligned} &h=(\hat{O},0,\hat{\delta}),\quad h^{-1}=(\hat{O}^T,0,-\hat{O}\hat{\delta}),\\ &A=(O,b,\delta)\to h^{-1}Ah\\ &\quad=(\hat{O}O\hat{O}^T,b+\hat{\delta}^T(1+O)\delta-\hat{\delta}^TO\hat{\delta},\\ &\qquad\hat{O}[\delta-(1-O^T)\hat{\delta}]), \end{aligned} \tag{A2a}$$

and similary for $A'=(O',b',\delta')$. Notice that b changes only by pure soul terms so that $\mathrm{Im}(b/b')_0$ remains unchanged, and we do not have to worry about reordering the (A,A') pair. As a special case, when $\hat{\delta}=0$, we get

$$A=(O,b,\delta)\to h^{-1}Ah=(\hat{O}O\hat{O}^T,b,\hat{O}\delta). \tag{A2b}$$

The following observations are important.

(1) When $\det O=1$, i.e., $O=R(r)$ [using the notation introduced in (3.2), (3.3)], O is brought by the conjugation (A2) either to itself or to its transpose (inverse) $R(-r)$, depending on the sign of $\det\hat{O}$. When $\det O=-1$, i.e., $O=M(r)=R(r)\mathbb{I}$, we can choose $\hat{O}=R(-r/2)$ and "canonize" O into $\mathbb{I}=\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$ (at the expense of rotating δ).

(2) δ can be conjugated to zero only if the matrix $\mathbf{1}-O^T$ is invertible, which is the case only when $\det O=1$ and $O\neq\mathbf{1}$ [more precisely, if we allow O to have soul terms, $O_0=\mathrm{body}(O)\neq\mathbf{1}$]. In particular, choosing $\hat{O}=\mathbf{1}$ and $\hat{\delta}=(\mathbf{1}-O^T)^{-1}\delta$, we get

$$\begin{aligned} &(R(r),b,\delta)\to(R(r),b-\delta^T[\mathbf{1}-R(-r)]^{-1}\delta,0)\\ &\quad\text{if } r_0\neq 0 \bmod 2\pi. \end{aligned} \tag{A2c}$$

If $O=\mathbb{I}$ [as we saw above, any $M(r)$ can be brought to this form] then the best we can do is choose $\hat{O}=\mathbf{1}$, $\hat{\delta}=\left(\begin{smallmatrix}0\\(1/2)\delta_2\end{smallmatrix}\right)$, and get

$$\left(\mathbb{I},b,\begin{pmatrix}\delta_1\\\delta_2\end{pmatrix}\right)\to\left(\mathbb{I},b,\begin{pmatrix}\delta_1\\0\end{pmatrix}\right). \tag{A2d}$$

Now choose h to be a transformation of the form (2.6) with $c=0=\gamma=\delta$, $O=\mathbf{1}$, and $d=a^{-1}$ (a must be invertible), i.e.,

$$h:\begin{cases}\hat{z}=a^2z,\\ \hat{\theta}=a\theta,\end{cases}$$

and we get

$$A=(O,b,\delta)\to h^{-1}Ah=(O,a^2b,a\delta). \tag{A3}$$

The following conjugations (A4) – (A6) will in general spoil the required form (3.1) of (A,A'), but for certain commuting pairs of generators that we encounter in Sec. III they are well behaved. Hence we just state the tranformation

h that is used without the explicit action on a generic generator.

In (2.6) take $a=b=d=1, c=0, \delta=\gamma$, and $O=1$ to get

$$h:\ \begin{cases}\hat{z}=(z+1)(1+\theta^T\gamma-2\theta^1\theta^2\gamma_1\gamma_2),\\ \hat{\theta}=\theta+(z+1)\gamma+\theta^1\theta^2\epsilon\gamma,\end{cases} \tag{A4a}$$

$$h^{-1}:\ \begin{cases}z=\hat{z}(1-\hat{\theta}^T\gamma-2\hat{\theta}^1\hat{\theta}^2\gamma_1\gamma_2)-1,\\ \theta=\hat{\theta}-\hat{z}(\gamma+2\gamma_1\gamma_2\epsilon\hat{\theta})-\hat{\theta}^1\hat{\theta}^2\epsilon\gamma.\end{cases}$$

Taking $\mathbb{I}$ instead of 1 we get a slightly different form:

$$h:\ \begin{cases}\hat{z}=(z+1)(1+\theta^T\mathbb{I}\gamma+2\theta^1\theta^2\gamma_1\gamma_2),\\ \hat{\theta}=\mathbb{I}\theta+(z+1)\gamma-\theta^1\theta^2\epsilon\gamma,\end{cases} \tag{A4b}$$

$$h^{-1}:\ \begin{cases}z=\hat{z}(1-\hat{\theta}^T\gamma-2\hat{\theta}^1\hat{\theta}^2\gamma_1\gamma_2)-1,\\ \theta=\mathbb{I}[\hat{\theta}-\hat{z}(\gamma+2\gamma_1\gamma_2\epsilon\hat{\theta})-\hat{\theta}^1\hat{\theta}^2\epsilon\gamma].\end{cases}$$

For treating the UU case we need conjugations by transformations that are not elements of $\mathrm{SPL}_2(2,\mathbb{C})$, namely ones that involve rotations in the θ plane whose angle is linear in z (the corresponding generator is T_1). The simplest such conjugation makes use of

$$h:\ \begin{cases}\hat{z}=z,\\ \hat{\theta}=R(tz)\theta,\end{cases} \tag{A5}$$

with t a bosonic constant. A more complicated one is defined through exponentiation of elements of the $N=2$ SCA (we could not find an explicit expression for the corresponding supergroup element),

$$h=\exp(it_1T_1+\gamma^TG_{1/2}), \tag{A6}$$

where t_1 and γ_i $(i=1,2)$ are constants (bosonic and fermionic, respectively). For being able to work with this representation we note (using the Baker–Campbell–Hausdorff formula[17]) that

$$\begin{aligned}A(R(r),b,\delta)&=\exp\{[b+(r/3)\mathrm{sinc}^{-2}(r/2)\delta_1\delta_2]L_{-1}+irT_0\\&\quad+\mathrm{sinc}^{-1}(r/2)[\delta^TR(-r/2)G_{-1/2}]\},\end{aligned} \tag{A7}$$

where

$$\mathrm{sinc}^{-1}r=\frac{r}{\sin r}=\left[\sum_{n=0}^{\infty}\frac{(-1)^nr^{2n}}{(2n+1)!}\right]^{-1}. \tag{A8}$$

What makes (A7) useful is the fact that for nilpotent r both the sum in (A8) and its inverse are finite! We remark also that on the lhs of (A7) we may have $r=2\pi n+$ (nilpotent part) $(n\in\mathbb{Z})$ without affecting A, but on the rhs r always stands for the nilpotent part only. The conjugation is done in the "algebra level" using the expansion

$$e^{-X}e^Ye^X=\exp\{Y+[Y,X]+(1/2!)[[Y,X],X]+\cdots\}, \tag{A9}$$

and the (anti-) commutation relations of the $N=2$ SCA.[8]

APPENDIX B: RIGOROUS SUPERSPACE THEORY

In this appendix we study the consequences of employing more rigorously the Rogers supermanifold approach to our discussion, and also use it for carrying out a proof that was omitted from the main text. Basically, we introduce an underlying Grassmann algebra generated by some large number (but finite) of anticommuting elements, and consider all the coordinates and parameters involved in the discussion as being valued in it. For more details the reader is referred to Ref. 11. The main difference between this approach and the more intuitive, physical one is that bosonic variables are not treated as ordinary complex (or real) numbers but as possibly having additional nilpotent parts ("souls"). This allows us to have consistent, *direct* interpretation of expressions like those appearing in (3.1), for instance, where we add a nilpotent term $\theta^TO\delta$ to the bosonic coordinate z.

Most of our results in the main text remain unchanged; we need only change the interpretation of the variables involved. New features appear, however, in the treatment of the *UUII* type of supertori in Sec. III, where we bring generator pairs that characterize this type of supertori into a canonical form. More explicitly, we have to consider supertori whose generator pair is of the form

$$A(R(r),\tau,\delta)\ \text{and}\ A'(R(r'),1,\delta')\quad\text{with}\ r_0=r_0'=0\ \mathrm{mod}\ 2\pi, \tag{B1}$$

i.e., r and r' are both pure soul instead of being simply zero (mod 2π). The commutativity conditions (3.9), (3.10) require in this case

$$\delta^T[\delta'+R(r')(\delta'-\delta)]=0, \tag{B2a}$$

$$R(r)[1-R(r')]\delta=R(r')[1-R(r)]\delta', \tag{B2b}$$

which can be shown to be equivalent to

$$\delta^TR((r'-r)/2)\delta'=0, \tag{B3a}$$

$$r\,\mathrm{sinc}^{-1}\left(\frac{r'}{2}\right)R\left(\frac{r'}{2}\right)\delta'=r'\,\mathrm{sinc}^{-1}\left(\frac{r}{2}\right)R\left(\frac{r}{2}\right)\delta. \tag{B3b}$$

Here r and r' are pure souls, any $2\pi m$ terms having been removed, and the sinc^{-1} function is defined in (A8). Assuming (B3) to hold we can use conjugation (A6) with

$$\begin{aligned}t_1&=r'-\left(\frac{1}{2}+\frac{r'^2}{3}\right)\mathrm{sinc}^{-2}\left(\frac{r'}{2}\right)\delta_1'\delta_2',\\ \gamma&=\mathrm{sinc}^{-1}\left(\frac{r'}{2}\right)R\left(\frac{r'}{2}\right)\delta',\end{aligned} \tag{B4}$$

which brings δ' to zero without changing the form (3.1) of the generator pair. Now conjugations (A3) and (A5) are used for bringing r' to zero, and we arrive at the canonical form

$$A(R(r),\tau,\delta)\ \text{and}\ A'(1,1,0). \tag{B5}$$

Here $r_0=0$ mod 2π, $\mathrm{Im}\,\tau>0$, and δ is an arbitrary fermionic column vector; the conditions (B2) are automatically satisfied and do not impose further constraints. Conjugation (A2) can be seen to give rise to superconformal equivalences within the family of supertori described by (B5), namely,

$$(\tau,r,\delta)\sim(\tilde{\tau},(\det\hat{O})r,\hat{O}\delta+r\hat{\delta}), \tag{B6}$$

where $\hat{O}$ and $\hat{\delta}$ are arbitrary, and $\tilde{\tau}$ differs from τ only by a nilpotent term depending on r, δ, $\hat{O}$, and $\hat{\delta}$. We could not find a way to use (B6) nor find some other conjugations that further simplify (B5) and avoid the "gauge freedom" expressed in (B6). An even more disturbing problem is the fact that the bosonic parameter r is restricted to have a vanishing

body, and as such it cannot play the role of a coordinate on the super-Teichmüller space, which is expected to have the structure of a supermanifold. It would be nice to be able to conjugate r to zero, hence reproducing the result of the main text, or conjugate it to be proportional to $\delta_1\delta_2$ through an unconstrained bosonic constant. Finding the necessary conjugations or proving that they do not exist is difficult, because it is hard to sort the conjugations that do not spoil the form (B5) of the generator pair out of the vast number of superconformal transformations (2.3).

Replacing the UUII canonical form (3.19) by (B5) requires a revised form for the action of the supermodular group on the corresponding piece $\mathrm{ST}_{12}^{\mathrm{UUII}}$ of the super-Teichmüller space, which can be shown to be described by

$$S\text{: UUII}\to\text{UUII}\ \tau\to\tau+1,\ r\to r,\ \delta\to\delta; \tag{B7a}$$

$$T\text{: UUII}\to\text{UUII}\begin{cases}\tau\to -\dfrac{1}{\tau}\left[1-\dfrac{r}{3\tau}\left(1-\dfrac{1}{\tau^2}\right)\mathrm{sinc}^{-2}\left(\dfrac{r}{2}\right)\delta_1\delta_2\right],\\ r\to\dfrac{r}{\tau}\left[1-\dfrac{r}{3\tau}\,\mathrm{sinc}^{-2}\left(\dfrac{r}{2}\right)\delta_1\delta_2\right],\\ \delta\to\tau^{-1/2}\,\dfrac{\sin(r/2\tau)}{\sin(r/2)}\,\delta,\end{cases} \tag{B7b}$$

where T^2 sends (τ,r,δ) to $(\tau,-r,i\delta)$. Notice that when $\delta=0$ or $r=0$ (B7) reduces to (4.2),(4.3) for the UUI or the UUII case, respectively. We have not worked out the superelliptic functions nor the embedding into projective superspace of the UUII supertori described by (B5).

We turn now to a proof of the completeness of the set of superelliptic functions $\mathfrak{R}$, $D_1\mathfrak{R}$, $D_2\mathfrak{R}$, and $\mathfrak{R}'$ in the UUII case with $r=0 \bmod 2\pi$. We use the same method used in Ref. 4 for proving that the set $\mathfrak{R}_1$, $D_1\mathfrak{R}_1$, and $\mathfrak{R}_1'$ is complete for $N=1$ (+ +) supertori. An arbitrary superelliptic function is expanded in the form (5.1),

$$f(z,\theta;\tau,\delta)=F_0(z;\tau,\delta)+\theta^1F_1(z;\tau,\delta)+\theta^2F_2(z;\tau,\delta)+\theta^1\theta^2F_3(z;\tau,\delta), \tag{B8}$$

and we express θ^1, θ^2, $\theta^1\theta^2$, and the functions F_i in terms of $\mathfrak{R}$, $D_1\mathfrak{R}$, $D_2\mathfrak{R}$, and $\mathfrak{R}'$ up to terms of higher order in the generators of the underlying Grassmann algebra. This is done by first noting that the bodies $F_{i,0}$ of F_i $(i=0,1,2,3)$ are all ordinary elliptic functions that can be written as rational functions of $\wp$ and $\wp'$, and replacing the latter by $\mathfrak{R}$ and $\mathfrak{R}'$, respectively, we get superelliptic functions $\tilde{F}_i$ that differ from $F_{i,0}$ only in soul. Next we use the identities

$$\theta=\frac{D\mathfrak{R}}{\mathfrak{R}'}-\delta\frac{\dot{\wp}}{\wp'}+\frac{1}{2}\left[\delta^T\epsilon\left(\frac{D\mathfrak{R}}{\mathfrak{R}'}\right)'\right]\frac{\dot{\wp}'}{\wp''}\epsilon\frac{D\mathfrak{R}}{\mathfrak{R}'}, \tag{B9}$$

$$\theta^1\theta^2=\frac{D_1\mathfrak{R}D_2\mathfrak{R}}{(\mathfrak{R}')^2}-\frac{\dot{\wp}}{\wp'}\left(\frac{D^T\mathfrak{R}}{\mathfrak{R}'}\epsilon\delta\right)+\delta_1\delta_2\left(\frac{\dot{\wp}}{\wp'}\right)^2, \tag{B10}$$

for writing

$$f=\tilde{F}_0+\frac{D_1\mathfrak{R}}{\mathfrak{R}'}\tilde{F}_1+\frac{D_2\mathfrak{R}}{\mathfrak{R}'}\tilde{F}_2+\frac{D_1\mathfrak{R}D_2\mathfrak{R}}{(\mathfrak{R}')^2}\tilde{F}_3+\tilde{f}. \tag{B11}$$

The functions $\tilde{F}_i$ depend on (z,θ) only through $\mathfrak{R}$ and $\mathfrak{R}'$, and $\tilde{f}$ is defined to be the function that satisfies (B11) and so is necessarily superelliptic. It can be expanded once more in the form (B8) where this time the expansion functions are at least of first order in the Grassmann generators. We repeat the above process inductively to fully express the (z,θ) dependence of f through $\mathfrak{R}$, $D_1\mathfrak{R}$, $D_2\mathfrak{R}$, and $\mathfrak{R}'$, which proves our claim.

[1] D. Friedan, in 1982 Les Houches summer school, *Recent Advances in Field Theory and Statistical Mechanics*, edited by J-B. Zuber and R. Stora (North-Holland, Amsterdam, 1984); L. Alvarez-Gaumé and P. Nelson, in *Supersymmetry, Supergravity, and Superstrings '86*, edited by B. de Wit, P. Fayet, and M. Grisaru (World Scientific, Singapore, 1986); P. Nelson, "Lectures on strings and moduli spaces," Harvard preprint HUPT-86/A047.

[2] D. Friedan, in *The Proceedings of the Workshop on Unified String Theories*, edited by D. Gross and M. Green (World Scientific, Singapore, 1986); E. Martinec, Nucl. Phys. B **281**, 157 (1987); L. Crane and J. M. Rabin, Commun. Math. Phys. **113**, 601 (1988).

[3] J. M. Rabin, "Super Riemann surfaces," University of Chicago preprint EFI 86-63, to appear in the *Proceedings of the Workshop on Mathematical Aspects of String Theory*, U. C. San Diego, 1986.

[4] J. M. Rabin and P. G. O. Freund, Commun. Math. Phys. **114**, 131 (1988).

[5] J. D. Cohn, Nucl. Phys. B **284**, 349 (1987).

[6] E. Verlinde and H. Verlinde, Phys. Lett. B **192**, 95 (1987); J. J. Atick, J. M. Rabin, and A. Sen, "An ambiguity in fermionic string perturbation theory," SLAC-PUB-4420.

[7] D. Friedan and S. H. Shenker, Phys. Lett. B **175**, 287 (1986).

[8] M. Ademollo, L. Brink, A. d'Adda, R. D'Auria, E. Napolitano, S. Sciuto, E. del Giudice, P. di Vecchia, S. Ferrara, F. Gliozzi, R. Musto, R. Pettorino, and J. Schwarz, Nucl. Phys. B **111**, 77 (1976); M. Ademollo, L. Brink, A. d'Adda, R. D'Auria, E. Napolitano, S. Sciuto, E. del Giudice, P. di Vecchia, S. Ferrara, F. Gliozzi, R. Musto, and R. Pettorino, Phys. Lett. B **62**, 105 (1976); P. di Vecchia, J. Petersen, and H. Zheng, *ibid.* **162**, 327 (1985).

[9] A. Sen, Nucl. Phys. B **284**, 423 (1987).

[10] D. Gepner, "Space-time supersymmetry in compactified string theory and superconformal models," Princeton University preprint PUPT-1056.

[11] A. Rogers, J. Math. Phys. **21**, 1352 (1980).

[12]S. Hamidi and C. Vafa, Nucl. Phys. B **279**, 465 (1986).
[13]P. Griffiths and J. Harris, *Principles of Algebraic Geometry* (Wiley, New York, 1978).
[14]K. Chandrasekharan, *Elliptic Functions* (Springer, Berlin, 1985).
[15]This interpolation is similar, in a sense, to the "spectral flow" (see Ref. 10 and references therein) of the identity operator into a field in the Ramond sector in a $N = 2$ superconformal field theory.
[16]Another nice property of $\tilde{\Phi}$, which can be inferred from the functional equation of ϑ, is its behavior under the supermodular transformation T: $\tilde{\Phi}(z;T(\tau,r)) = \tilde{\Phi}(z;-1/\tau,r/\tau) = e^{i\pi z}\tilde{\Phi}(\tau z;\tau,r)$.
[17]V. S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations* (Springer, New York, 1984).

Curvature of generic space-times in general relativity

A. D. Rendall[a)]
Department of Mathematics, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen AB9 2TY, Scotland

(Received 15 July 1987; accepted for publication 24 February 1988)

Transversality theory is used to study the curvature of generic space-times in general relativity and some applications to unique determination of the metric by the curvature are given.

I. INTRODUCTION

In general relativity the physical space-time metric can only be measured with finite accuracy. Thus, as pointed out by Hawking,[1] when proving theorems it is sufficient for physical purposes to show that they hold generically, i.e., that they hold for almost all metrics in an appropriate sense. This can be formulated mathematically by defining a property to be generic if it holds for an open dense subset of the space of all Lorentz metrics on a given manifold in a suitable topology. In this paper the Whitney C^k topologies are used. The relative merits of this choice and others are discussed in Refs. 1–3.

The main results in this paper are concerned with certain generic properties of the curvature of space-time. Propositions 6.1–6.3 discuss the Riemann tensor, energy-momentum tensor, and Weyl tensor, respectively. Section II is concerned with matrices that represent the components of the curvature at a point of space-time. It is shown that the set of matrices of a given algebraic type is a regular[4] submanifold of the relevant matrix space and a method is given of calculating the dimension of such a submanifold. In Sec. III the maps sending a metric to the various types of curvature are studied. The methods used are local, working in Riemann normal coordinates. Section IV presents some results on transversality required in the main proofs. The idea is to identify natural submanifolds of the bundle J^2L of two-jets of Lorentz metrics defined by conditions on the curvature and compute their codimensions. It is then possible to define a generic set of metrics by requiring them to be transverse to these submanifolds. This transversality condition ensures that the subset of space-time on which one of these curvature conditions is satisfied is a regular submanifold and that its codimension is the same as that of the corresponding submanifold of J^2L. In Sec. V this procedure is carried out using the information obtained in Secs. II and III. In Sec. VI the main results are stated together with some corollaries on the determination of the metric by the curvature and on holonomy groups.

II. SOME RESULTS ON MATRICES

A. The Riemann tensor case

Consider the action of the Lie group $G = \mathrm{Gl}(6,\mathbb{R}) \times \mathrm{Gl}(6,\mathbb{R})$ on $M_6(\mathbb{R})$, the space of real 6×6 matrices, given by $\phi((P,Q),X) = P^{-1}XQ$. The orbit of a matrix X under this action consists of all matrices with the same rank as X. Let $\mathcal{S}$ be the set of symmetric matrices in $M_6(\mathbb{R})$ and let H be the subgroup of G consisting of elements that map $\mathcal{S}$ onto itself. The transformations arising from the action of H include all congruence transformations (i.e., the case $P^{-1} = Q^T$) and so the action of H on $\mathcal{S}$ has only finitely many orbits. In each one finds a diagonal matrix each of whose entries is ± 1 or 0, i.e., the Sylvester canonical form. It will be shown that each orbit is a regular submanifold of $\mathcal{S}$ and that these manifolds "fit together nicely." To make the latter precise some definitions are required.

Definition 2.1: Let C be a closed subset of smooth manifold X. A finite set $A_0,\ldots,A_q$ of regular submanifolds of X is said to be a *stratification* of C if

(i) $\overline{A}_{i+1} \setminus A_{i+1} \subset A_0 \cup \cdots \cup A_i$, for each i,
(ii) $\dim A_i < \dim A_{i+1}$, for each i,
(iii) $A_0 \cup \cdots \cup A_q = C$.

Definition 2.2: A stratification $A_0,\ldots,A_q$ of C is said to be *smoothly locally trivial* if given $x \in A_j$ there is an open neighborhood U of x, manifolds Y and Z, a stratification $A'_0,\ldots,A'_q$ of a closed subset C' of Y and a diffeomorphism $f: Y\times Z \to U$ so that

(i) $f(C'\times Z) = C \cap U$,
(ii) $f(A'_i \times Z) = A_i \cap U$, for each i,
(iii) A'_j is a single point.

Intuitively this says that near x the given stratification looks like the product of A_j with a stratification of lower dimension.

Now let J be the matrix

$$\mathrm{diag}\{\underbrace{1,\ldots,1}_{p_1}, \underbrace{-1,\ldots,-1}_{p_2}\}$$

and consider the orbit $\mathcal{O}$ in $\mathcal{S}$ containing the matrix $X_0 = \left[\begin{smallmatrix} J & 0 \\ 0 & 0 \end{smallmatrix}\right]$. If $P(t)$ is a path in $\mathrm{Gl}(6,\mathbb{R})$ with $P(0) = I$, define $X(t) = P(t)X_0 P^T(t)$. Then

$$\frac{dX}{dt}(0) = \frac{dP}{dt}(0)X_0 + X_0\frac{dP^T}{dt}(0).$$

Let Π be the plane through X_0 that is the union of the vectors $(dX/dt)(0)$ for all such paths. Let $\Pi^\perp$ be the plane through X_0 that is the orthogonal complement of Π with respect to the inner product $\langle X,Y\rangle = \mathrm{tr}(XY)$ on $\mathcal{O}$. Here $\Pi^\perp$ consists of all matrices of the form $\left[\begin{smallmatrix} J & 0 \\ 0 & S \end{smallmatrix}\right]$ where S is symmetric. Then X_0 is the only matrix of rank p_1+p_2 in $\Pi^\perp$. Let K be the subgroup of H that fixes X_0 and let V be a submanifold of H with the property that $T_eH = T_eK + T_eV$.[5] Define a map f from $V\times\Pi^\perp$ to $\mathcal{S}$ by $f(v,X) = \phi(v,X)$. The Jacobian of f at (e,X_0) is nonsingular. It follows by the inverse function theorem that there are open neighborhoods U_1 of e in V and

[a)] Present address: Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 8046 Garching bei München, Federal Republic of Germany.

 0022-2488/88/071569-06$02.50

U_2 of X_0 in $\Pi^\perp$ such that the restriction of f to $U_1 \times U_2$ is a diffeomorphism onto an open neighborhood U of X_0 in $\mathscr{S}$. Also $f^{-1}(\mathscr{O}) = U_1 \times \{X_0\}$ and so $\mathscr{O} \cap U$ is a regular submanifold of $\mathscr{O}$. Since H acts transitively on $\mathscr{O}$ it follows that $\mathscr{O}$ is a regular submanifold of $\mathscr{S}$. By the same token any other orbit $\mathscr{O}'$ is a regular submanifold of $\mathscr{S}$. It follows that if A_i is the set of symmetric matrices of rank i then the A_i form a smoothly locally trivial stratification of $\mathscr{S}$. To see this take U_2 as the manifold Y in Definition 2.2 and let $A_i' = A_i \cap U_2$. Then the A_i' form a stratification of Y and all the conditions of the definition are fulfilled. From the explicit form of $\Pi^\perp$ that is the normal plane to the orbit at X_0 it can be seen that for $p_1 + p_2 = 5, 4$, and 3 the codimension of the orbit (i.e., $\dim \mathscr{S} - \dim \mathscr{O}$) is 1, 3, and 6, respectively.

The normal space of $\mathscr{O}$ at X_0 (i.e., the plane through the origin parallel to $\Pi^\perp$) contains no matrices of rank 6. Using the action of H it follows that the normal plane of $\mathscr{O}$ at any point contains no matrices of rank 6. Thus if $Z_0 = \left[\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right]$, where I is the 3×3 identity matrix, Z_0 is never normal to any orbit except the trivial orbit $\{0\}$. This fact will be important later.

B. The energy-momentum tensor case

Let η be the matrix diag $\{-1,1,1,1\}$. A 4×4 matrix X will be called self-adjoint if $\eta X = X^T \eta$. The self-adjoint matrices form a ten-dimensional linear subspace T of $M_4(\mathbb{R})$. Let $\mathscr{L} = \{P \in \mathrm{Gl}(4,\mathbb{R}): P^T \eta P = \eta\}$. Here $\mathscr{L}$ is the Lorentz group and it acts on T by similarity: $(P,X) \to P^{-1} X P$. All matrices in a given orbit of this action have the same eigenvalues.

If p is a point of space-time and we choose a coordinate system so that the matrix of components of the metric at p is η then the matrix of components $T_a{}^b$ of the energy-momentum tensor at p is self-adjoint. Hall[6] has given canonical forms for $T_a{}^b$ under the action of the Lorentz group. His results are expressed in terms of a null tetrad. If we transform to an orthonormal tetrad we obtain canonical forms for self-adjoint matrices under the action of $\mathscr{L}$ by similarity. There are four cases corresponding to Eqs. (3)–(6) in Hall's paper:

$$\begin{bmatrix} p_1 \pm 1 & \pm 1 & 0 & 0 \\ \mp 1 & p_1 \mp 1 & 0 & 0 \\ 0 & 0 & p_2 & 0 \\ 0 & 0 & 0 & p_3 \end{bmatrix}, \tag{2.1}$$

$$\begin{bmatrix} p_1 & 0 & -1 & 0 \\ 0 & p_1 & 1 & 0 \\ 1 & 1 & p_1 & 0 \\ 0 & 0 & 0 & p_2 \end{bmatrix}, \tag{2.2}$$

$$\begin{bmatrix} p_1 & -p_2 & 0 & 0 \\ p_2 & p_1 & 0 & 0 \\ 0 & 0 & p_3 & 0 \\ 0 & 0 & 0 & p_4 \end{bmatrix}, \tag{2.3}$$

$$\mathrm{diag}\{p_1,p_2,p_3,p_4\}. \tag{2.4}$$

Each orbit under the action of $\mathscr{L}$ on T contains at least one matrix in one of the canonical forms (2.1)–(2.4).

Following Ref. 5 the set of matrices in T of a given Segré type[7] will be called a bundle. It is a union of orbits. For $X_0 \in T$ let Π be the tangent plane at X_0 to the orbit $\mathscr{O}$ containing X_0 and let $\Pi^\perp$ be the normal plane to $\mathscr{O}$ at X_0 with respect to the inner product $\langle X,Y \rangle = \mathrm{tr}(XY^T)$. Suppose that X_0 is in one of the canonical forms introduced above and let H be the subgroup of $\mathscr{L}$ that leaves X_0 fixed. If V is a submanifold of $\mathscr{L}$ with $T_e\mathscr{L} = T_e V \oplus T_e H$ then as in Sec. II A there are open neighborhoods U_1 of e in V and U_2 of X_0 in $\Pi^\perp$ and a diffeomorphism f from $U_1 \times U_2$ onto an open neighborhood U of X_0.

Here X_0 is block diagonal with blocks of order $n_1,\ldots,n_k$, where $n_1 + \cdots + n_k = 4$. Identify the vector space $M_{n_1}(\mathbb{R}) \oplus \cdots \oplus M_{n_k}(\mathbb{R})$ with the linear subspace of $M_4(\mathbb{R})$ consisting of block diagonal matrices with blocks of these orders. It is possible to express X_0 as $X_1 \oplus \cdots \oplus X_k$, where, for each i, X_i has only one eigenvalue and X_i and X_j do not have the same eigenvalue unless $i = j$. Let $T_i = T \cap M_{n_i}(\mathbb{R})$, let $\mathscr{O}_i$ be the set of matrices in T_i similar to X_i, and let B_i be the set of matrices in T_i with the same Segré type as X_i. It can be shown[8] that the normal space to $\mathscr{O}$ at X_0 is the direct sum of the normal spaces to the $\mathscr{O}_i$ at X_i. Also if $Y = Y_1 \oplus \cdots \oplus Y_k$ belongs to a sufficiently small neighborhood of X_0 then Y is similar to X_0 if and only if Y_i is similar to X_i for each $i = 1,\ldots,k$.

Let T_i' be the matrices of trace zero in T_i and let Z_i be the scalar multiples of the identity matrix of order n_i so that $T_i = T_i' \oplus Z_i$. Then $B_i = (B_i \cap T_i') \times Z_i$ because adding a scalar multiple of the identity to a matrix does not change its Segré type. Now $B_i \cap T_i'$ consists of a single orbit [except in the case of the canonical form (2.1) where it may consist of two orbits corresponding to the two choices of sign in the canonical form]. It can be shown by a case by case check,[8] using the canonical forms given, that there is an open neighborhood of a matrix in the normal plane to $B_i \cap T_i'$ at that matrix that only intersects $B_i \cap T_i'$ at a single point. Thus the normal plane to $\mathscr{O}_i$ at X_i intersects B_i precisely in the set $X_i + Z_i$, i.e., $\{X_i + Y: Y \in Z_i\}$ within some neighborhood of X_i. The conclusion is that there is an open neighborhood U_2 of X_0 in $\Pi^\perp$ such that $U_2 \cap B = U_2 \cap (X_0 + (Z_1 \oplus \cdots \oplus Z_k))$. It then follows from the existence of the diffeomorphism from $U_1 \times U_2$ to U constructed earlier that B is a regular submanifold of T. In fact the union of the bundles of a given dimension is a regular submanifold of T and as in Sec. II A it follows that these submanifolds form a smoothly locally trivial stratification of T.

The above argument also gives a method of calculating the dimensions of the bundles. Consider the Segré type $\{1,1(11)\}$. This means that the canonical form is (2.4) with $p_3 = p_4$. By direct computation the Lorentz matrices that leave this fixed are those of the form

$$\begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & A \end{bmatrix},$$

where A is a 2×2 matrix with $A^T A = I$. Thus H is one dimensional in this case. Thus the dimension of the orbit is $\dim \mathscr{O} - \dim H = 5$. Since there are three independent eigenvalues and hence three blocks in the decomposition of

X_0, dim $B = 5 + 3 = 8$. For Segré type $\{1,(111)\}$ the canonical form is (2.4) with $p_2 = p_3 = p_4$. In this case H consists of all matrices of the form $\left[\begin{smallmatrix} \pm 1 & 0 \\ 0 & A \end{smallmatrix}\right]$, where A is a 3×3 matrix with $A^{T}A = I$. Thus H is three dimensional and the dimension of the orbit is 3. Since there are two independent eigenvalues the dimension of the bundle is $3 + 2 = 5$. This method can be used to calculate the dimension of any of the other bundles.

C. The Weyl tensor case

Consider the set Z of symmetric complex 3×3 matrices with zero trace. An argument similar to that of Sec. II B, this time making use of the Petrov canonical forms, shows that the bundles of matrices of given Segré types in Z (which correspond to the Petrov types) form a smoothly locally trivial stratification of Z. We regard Z here as a real ten-dimensional manifold. More details can be found in Ref. 3 where it is shown that the bundles of matrices of types I, II, D, and III are of dimensions 10, 8, 6, and 6, respectively.

III. THE DEPENDENCE OF CURVATURE ON THE METRIC

A. The Riemann tensor

Let M be a fixed four-dimensional manifold and let J^2L be the bundle of two-jets of Lorentz metrics[9]. The Riemann tensor of any Lorentz metric has the symmetries

$$R_{abcd} = -R_{bacd}\,, \tag{3.1}$$

$$R_{abcd} = R_{cdab}\,, \tag{3.2}$$

$$R_{a[bcd]} = 0\,, \tag{3.3}$$

where square brackets denote antisymmetrization. Equation (3.3) is the algebraic Bianchi identity. Let K be the bundle of tensors with the symmetries (3.1)–(3.3) over M. The components of the Riemann tensor at a point depend only on the components of the metric and their first and second partial derivatives at that point. This means that the map sending a metric to its Riemann tensor arises from a map $R: J^2L \to K$. The aim is to show that R is a submersion.

Let $\mathscr{R}$ be the vector space of sets of components with the symmetries (3.1)–(3.3) and let $\mathscr{Q}$ be the vector space of sets of components with the symmetries

$$Q_{abcd} = Q_{bacd}\,, \tag{3.4}$$

$$Q_{abcd} = Q_{cdab}\,, \tag{3.5}$$

$$Q_{a(bcd)} = 0\,, \tag{3.6}$$

where round brackets denote symmetrization. Define a linear map $\Phi: \mathscr{R} \to \mathscr{Q}$ by $\Phi(R_{abcd}) = \frac{1}{2}(R_{acbd} + R_{adbc})$ and a map $\Psi: \mathscr{Q} \to \mathscr{R}$ by $\Psi(Q_{abcd}) = \frac{2}{3}(Q_{acbd} - Q_{adbc})$.[10] Then $\Phi \circ \Psi$ and $\Psi \circ \Phi$ are the identity maps on $\mathscr{Q}$ and $\mathscr{R}$, respectively, and hence Φ and Ψ are isomorphisms.

Now let (M,g) be a space-time. Given $p \in M$, we can set up Riemann normal coordinates in a neighborhood of p. Then[11] $g_{ab,cd}$ belongs to $\mathscr{Q}$ at p and $R_{abcd} = -\frac{2}{3}\Psi(g_{ab,cd})$ at that point. The given coordinate system on a neighborhood U of p identifies the parts of J^2L and K over U with open subsets of Euclidean space and it is sufficient to check that the coordinate representative of the map is a submersion. The formula for R_{abcd} shows that by leaving the components g_{ab} and $g_{ab,c}$ at p unchanged and altering $g_{ab,cd}$ at p the components R_{abcd} at p can be changed in any desired direction. More precisely, given any vector V at $R_{abcd}(p)$, there exists a curve of two-jets of metrics $(g_{ab}, g_{ab,c}, g_{ab,cd}(t))$ such that the resulting curve of Riemann tensors $R_{abcd}(t)$ has tangent vector v at $t = 0$. It follows that R is a submersion.

B. The energy-momentum tensor

Let S be the bundle of symmetric second-order covariant tensors over M. Then the energy-momentum tensor in the form T_{ab} is a section of this bundle. The map sending a metric to its energy-momentum tensor arises from a map E: $J^2L \to S$. This map is a submersion. For working in Riemann normal coordinates as in Sec. III A it is clear that by changing $g_{ab,cd}$ at p in an appropriate way the components T_{ab} at p can be changed in any desired direction.

C. The Weyl tensor

The corresponding discussion for the Weyl tensor is a little more complicated and will be left until Sec. V.

The point of these results is that in what follows we need to know, for instance, that if A is a submanifold of K then $R^{-1}(A)$ is a submanifold of J^2L. It is a consequence of the theorems quoted in the next section that this follows from the fact that R is a submersion. On the other hand, the inverse image of a submanifold under an arbitrary smooth map may be very complicated and is not, in general, a submanifold. For example, if f: $\mathbb{R}^2 \to \mathbb{R}^2$ is defined by $f(x,y) = (x, x(y - x^2))$ and $A = \{(x,y) \in \mathbb{R}^2 : y = 0\}$ then $f^{-1}(A)$ is not a submanifold of $\mathbb{R}^2$.

IV. REMARKS ON TRANSVERSALITY

Definition 4.1: Let X, Y be smooth manifolds, $f: X \to Y$ a C^k map, and Y_1 a regular submanifold of Y. Here f is said to be transverse to Y_1 if at each point $p \in X$ one of the following conditions holds:

(i) $f(p) \notin Y_1$

or

(ii) $f_{*,p}(T_pX) + T_{f(p)}Y_1 = T_{f(p)}Y$.

(Here $f_{*,p}$ denotes the differential of f at p).

A fundamental property of transverse maps is the following.[12]

Theorem 4.1: Let X, Y, f, Y_1 be as above. Then $f^{-1}(Y_1)$ is a regular submanifold of X and its codimension in X is equal to that of Y_1 in Y.

In what follows it will be necessary to consider maps that are transverse to several submanifolds. In order to discuss this situation another definition is required.

Definition 4.2: A stratification $A_0,\ldots,A_q$ is called a submanifold complex[12] if given a sequence $\{x_n\}$ in A_j converging to $y \in A_i$ there is a sequence $\{E_n\}$ of d planes ($d = \dim A_i$) with $E_n \subset T_{x_n}A_j$ converging to T_yA_i.

It is not difficult to show that a smoothly locally trivial stratification is a submanifold complex; the converse is not in general true. If $A_0,\ldots,A_q$ is a submanifold complex and f is transverse to each A_i then $f^{-1}(A_0),\ldots,f^{-1}(A_q)$ also form a

submanifold complex.[12] There is an analogous definition for subbundles.

Definition 4.3: Let E be a fiber bundle over a manifold X with projection π and typical fiber Y. Submanifolds $A_0,\ldots,A_q$ of E are said to form a *subbundle complex* of E if there is a submanifold complex $Z_0,\ldots,Z_q$ of Y with the following property: given $x\in X$ there is a neighborhood U of x and a diffeomorphism $f\colon \pi^{-1}(U)\to U\times Y$ such that $f(A_i)=U\times Z_i$ for each i and the projection of $f(e)$ on U coincides with $\pi(e)$ for each $e\in\pi^{-1}(U)$.

A subbundle complex of a jet bundle J^kE is called *natural* if for each x the map f in Definition 4.3 can be chosen to be the trivialization of J^kE associated with a trivialization of E. The following jet transversality theorem for bundles has been proved in Ref. 3 by modifying arguments from Hirsch.[12]

Theorem 4.2: Let E be a fiber bundle over X and $A_0,\ldots,A_q$ a natural subbundle complex of J^kE. Let Γ^rE denote the space of C^r sections of E with Whitney C^r topology. Then for $r\geqslant k+1$ the set of C^r sections of E whose k-jet prolongations are transverse to $A_0,\ldots,A_q$ is open and dense in Γ^rE.

V. GEOMETRY OF THE ALGEBRAIC TYPES OF CURVATURE

A. The Riemann tensor

It was shown in Sec. II A that rank gives a smoothly locally trivial stratification (and hence a submanifold complex) $A_0,\ldots,A_6$ of the space $\mathcal{S}$ of symmetric matrices. It was also shown that $Z_0=\left[\begin{smallmatrix}0 & I\\ -I & 0\end{smallmatrix}\right]$ is never normal to any nontrivial orbit. Thus if $\mathcal{S}'$ is the orthogonal complement of Z_0 in $\mathcal{S}$ the inclusion map $\mathcal{S}'\to\mathcal{S}$ is transverse to $A_1,\ldots,A_6$. It follows that $A_0\cap\mathcal{S}',\ldots,A_6\cap\mathcal{S}'$ form a submanifold complex in $\mathcal{S}'$. If K is the bundle (introduced in Sec. III A) of tensors with symmetries (3.1)–(3.3) then a coordinate system on an open subset U of M gives a correspondence $K\,|\,U\to U\times\mathcal{S}'$. [Taking the orthogonal complement of Z_0 corresponds to imposing the algebraic Bianchi identity (3.3).] In this way a submanifold complex of $K\,|\,U$ is determined. This is independent of the coordinate system chosen and so a submanifold complex of K is determined. The manifolds composing it have the same codimensions in K as $A_0,\ldots,A_6$ have in $\mathcal{S}$. If $R\colon J^2L\to K$ is the map introduced in Sec. III, then, since R is a submersion, it is transverse to each of these manifolds. It follows that their inverse images under R, call them $B_0,\ldots,B_6$, form a submanifold complex of J^2L. In fact, because of the way they were constructed, $B_0,\ldots,B_6$ form a natural subbundle complex of J^2L. From Theorem 4.1 the codimension of B_i in J^2L is equal to the codimension of A_i in $\mathcal{S}$.

B. The energy-momentum tensor

Now let L_0 be the set of 4×4 symmetric matrices of Lorentz signature and let $\mathcal{S}''$ be the full set of symmetric 4×4 matrices. Consider the set of elements (g,T) of $L_0\times\mathcal{S}''$ for which ηT has a given Segré type (where η is as in Sec. II B). This is a product of a bundle (in the sense of Sec. II) with L_0 and so is a regular submanifold of $L_0\times\mathcal{S}''$. Furthermore, the collection of such manifolds form a submanifold complex. Fix $g_0\in L_0$. Choose a neighborhood Y of g_0 in L_0 and for $g\in Y$ let $X(g)$ be a matrix that maps the standard basis of $\mathbb{R}^4$ to a g-orthonormal one and that depends smoothly on g. (The possibility of doing this follows from the fact that the Gram–Schmidt orthogonalization process is smooth.) Let ϕ be the map sending (g,T) to $(g,X^T(g)TX(g))$. It is a diffeomorphism from $Y\times\mathcal{S}''$ to itself. Also $g^{-1}T$ has the same Segré as $\eta X^T(g)TX(g)$. Thus the set $\{(g,T)\colon g^{-1}T$ has a given Segré type$\}$ is a regular submanifold of $Y\times\mathcal{S}''$ and these manifolds form a submanifold complex of $Y\times\mathcal{S}''$. It is then easily seen that the same is true for the whole of $L_0\times\mathcal{S}''$.

Let S be the bundle of symmetric second-order covariant tensors as in Sec. III B. A coordinate system on an open subset U of M gives a correspondence $S\,|\,U\to U\times\mathcal{S}''$ and C^r Lorentz metrics on U are in one-to-one correspondence with C^r maps $U\to L_0$. Let F be the typical fiber of J^2L. Then the restriction of the map $E\colon J^2L\to S$ of Sec. II to the part of J^2L over U corresponds to a map from $U\times F\to U\times\mathcal{S}''$ that is actually of the form (Id_U,E'), where Id_U is the identity map on U. Since E is a submersion, E' is also.

Let p be the natural projection of F on L_0 and consider the map $p\times T\colon U\times F\to L_0\times\mathcal{S}''$ that sends (x,e) to $(p(e),E'(e))$. Then $p\times T$ is transverse to the submanifolds of $L_0\times\mathcal{S}''$ defined by the Segré types. It follows that the sets of two-jets of Lorentz metrics whose energy-momentum tensors have the various Segré types form a submanifold complex $B_0',\ldots,B_q'$ in $J^2L\,|\,U$. Since the definition of the Segré type for the energy-momentum tensor is coordinate independent this defines a submanifold complex of J^2L, which we also denote by $B_0',\ldots,B_q'$. In fact because of the way these manifolds were constructed they form a natural subbundle complex. The codimension of B_i' in J^2L is equal to that of the corresponding submanifold of self-adjoint matrices in T.

C. The Weyl tensor

Let $V=\mathbb{R}^4$ and let η be as above. Let $\hat{C}$ be the linear subspace of $V\otimes V^*\otimes V^*\otimes V^*$ obtained by raising the first index with η of all members of $V^*\otimes V^*\otimes V^*\otimes V^*$ satisfying conditions (3.1)–(3.3). Let C_η be the subspace of $\hat{C}$ for which contraction on the first and third indices gives zero. Thus if p is a point of space-time and we choose a coordinate system in a neighborhood of p so that the matrix of components of the metric at p is η, then the components of the Weyl tensor at p belong to C_η. If Z is the set of 3×3 complex symmetric matrices then it is well known[13] that there is an isomorphism $C_\eta\to Z$. From Sec. II C the sets of matrices in Z of the different Segré types form a submanifold complex in Z. This gives rise to a submanifold complex in C_η via the above isomorphism. In this case the various bundles (in the sense of Sec. II) correspond to the Petrov types.

Again let L_0 be the set of symmetric 4×4 matrices of Lorentz signature. Given $g\in L_0$ construct a vector space C_g in the same way as C_η was constructed from η. Then there is a naturally defined vector bundle over L_0 whose fiber is C_g. To demonstrate this it suffices to construct trivializations for it. As in Sec. V B define a function X on some open neighborhood Y of $g_0\in L_0$ with the property that $X(g)$ takes the stan-

dard basis of $\mathbb{R}^4$ to a g-orthonormal one. Then for each g in Y the map $X(g): V \to V$ induces a map from $V \otimes V^* \otimes V^* \otimes V^*$ to itself that restricts to an isomorphism $\phi_g: C_g \to C_\eta$. Define a map

$$\phi: \bigcup_{g \in Y} (\{g\} \times C_g) \to Y \times C_\eta \text{ by } \phi(g,x) = \phi_g(x) .$$

Then ϕ is the required trivialization. The collection of such ϕ's forms a vector bundle atlas for a bundle E over L_0. Moreover if ϕ_1, ϕ_2 are two such maps then $\phi_2^{-1} \circ \phi_1$ preserves Petrov type. Thus there is a submanifold complex of E corresponding to the Petrov types.

Let U be an open subset of M covered by a single coordinate system. As in Sec. V B let F be the typical fiber of J^2L. Given a coordinate system on U, a Lorentz metric on U corresponds to a map $U \to L_0$. The Weyl tensor of such a metric can then be regarded as a map from U to E. The map sending a metric to its Weyl tensor is then induced by a map $C: U \times F \to E$. By a similar argument to those of Sec. III it can be shown that C is a submersion. It then follows that there is a natural subbundle complex of J^2L representing the Petrov types. It will be denoted by $B_0'', \ldots, B_s''$.

VI. CONCLUSIONS

Proposition 6.1: Let M be a four-dimensional manifold, $\Gamma^r L$ the space of C^r Lorentz metrics on M with Whitney C^r topology. Then there is an open dense subset W of $\Gamma^r L$ $(r \geqslant 3)$ such that if $g \in W$ its Riemann tensor has rank 6 everywhere except possibly on a three-dimensional regular submanifold of M where it has rank 5 and a one-dimensional regular submanifold where it has rank 4.

Proof: Let $B_0, \ldots, B_6$ be the natural subbundle complex of J^2L constructed in Sec. V A and let W be the set of C^r metrics whose two-jet prolongations are transverse to $B_0, \ldots, B_6$. Then it follows from Theorem 4.2 that W is open and dense. If $g \in W$ then applying Theorem 4.1 and using the codimensions of $B_0, \ldots, B_6$ gives the desired result.

This proposition has interesting consequences relating to work of Hall[14] concerning the following question. Given a tensor $R^a_{\ bcd}$ that is the Riemann tensor of some Lorentz metric on a manifold M is there any other metric with the same Riemann tensor? In Ref. 14 sufficient conditions on the Riemann tensor are given to ensure that the only such metrics are those related to the original metric by a constant conformal rescaling. In particular, it is shown that if the Riemann tensor has rank of at least 4 at every point of space-time then the metric is determined up to a constant factor. Thus we have the following corollary.

Corollary 1: In $\Gamma^r L$ $(r \geqslant 3)$ it is generic for a metric to be uniquely determined up to a constant conformal factor by its Riemann tensor.

Remark: Hall and Kay[15] have shown how the property of unique determination of the metric is closely related to the holonomy group of space-time. It turns out that Proposition 6.1 also has consequences for the holonomy group, namely that for a generic metric the local holonomy group at each point of space-time is equal to the full identity component of the Lorentz group. More details are given in Ref. 15.

Proposition 6.2: Let M and $\Gamma^r L$ be as in Proposition 6.1. Then there is an open dense subset W' of $\Gamma^r L$ $(r \geqslant 3)$ such that if $g \in W'$: (i) its energy-momentum tensor has Segré type $\{1,111\}$ or $\{z\bar{z}11\}$ on an open dense subset of M; or (ii) if its energy-momentum tensor satisfies the weak energy condition[16] then the Segré type is $\{1,111\}$ at all points of M except possibly on some two-dimensional submanifold where it is $\{1,1(11)\}$.

Proof: Let $B_0', \ldots, B_q'$ be the natural subbundle complex of J^2L constructed in Sec. V B and let W' be the set of C^r metrics whose two-jet prolongations are transverse to $B_0', \ldots, B_q'$. Then it follows from Theorem 4.2 that W' is open and dense.

Suppose now that $g \in W'$. The only Segré types for which the corresponding bundles of matrices are of maximal dimension are $\{1,111\}$ and $\{z\bar{z}11\}$. Thus B_q' consists precisely of those two-jets of metrics whose energy-momentum tensors are of one of these types. Part (i) of the proposition now follows from Theorem 4.1. Now suppose that the energy-momentum tensor of g satisfies the weak energy condition. Consider a Segré type whose corresponding bundle B of self-adjoint matrices lies in the closure of the set of matrices whose eigenvalues are not all real. Then the energy-momentum tensor of g never takes on that Segré type. For if it did so at a point $p \in M$ the components of the energy-momentum tensor in a chart about p would give a map into the self-adjoint matrices whose image intersected B. Since $g \in W'$ this map would be transverse to B and it would then follow from the local triviality of the stratification by bundles that the image of this map also contained matrices with nonreal eigenvalues. Thus at some point close to x the energy-momentum tensor would have Segré type $\{z\bar{z}11\}$ or $\{z\bar{z}(11)\}$ and would necessarily fail the weak energy condition.[16] The only Segré types that do not lie in the closure of the set of matrices whose eigenvalues are not all real are $\{1,111\}$, $\{1,1(11)\}$, and $\{1,(111)\}$. The dimensions of the corresponding bundles of self-adjoint matrices were calculated in Sec. II B. Part (ii) of the proposition now follows from Theorem 4.1.

Proposition 6.3: Let M and $\Gamma^r L$ be as in Proposition 6.1. Then there is an open dense subset W'' of $\Gamma^r L$ $(r \geqslant 3)$ such that if $g \in W''$ its Weyl tensor has Petrov type I at all points of M except possibly on a two-dimensional regular submanifold of M where the type is II and isolated points where it is III or D.

Proof: Let $B_0'', \ldots, B_s''$ be the natural subbundle complex of J^2L constructed in Sec. V C and let W'' be the set of C^r metrics whose two-jet prolongations are transverse to $B_0'', \ldots, B_s''$. Then from Theorem 4.2 the set W'' is open and dense. Now apply Theorem 4.1 and use the dimensions stated in Sec. II C.

Like Proposition 6.1, Proposition 6.3 has consequences for the unique determination of the metric by its curvature. This time, however, the curvature in question is not the Riemann tensor but the sectional curvature function. Recall that if a non-null two-space is spanned by vectors x^a, y^a then its sectional curvature is defined to be

$$\frac{R_{abcd}\, x^a\, y^b\, x^c\, y^d}{(x_a\, x^a)\,(y_c\, y^c) - (x_a\, y^a)^2} .$$

Hall[17] has shown that if two metrics have the same sectional

curvature function and do not have constant curvature on any nonempty open set then they are conformally related. Moreover, if the conformed factor is not everywhere equal to 1 then there is a nonempty open set where the Weyl tensors of both metrics vanish and other special conditions are satisfied. On the other hand, if a metric has constant curvature on some open set then its Weyl tensor vanishes there. Thus if any metric is not uniquely determined by its sectional curvature function then its Weyl tensor must vanish on some nonempty open subset of space-time and the Petrov type is 0. Comparing with Proposition 6.3 we obtain the following corollary.

Corollary 2: In $\Gamma^r L$ $(r \geqslant 3)$ it is generic for a metric to be uniquely determined by its sectional curvature function.

In view of the present interest among physicists in higher-dimensional space-times it is interesting to ask whether the corollary to Proposition 6.1 is valid for Lorentz metrics on manifolds of dimension greater than 4. The author has shown that in fact this is true but the techniques used in this paper are not directly applicable in that case. The proof will be given in a future publication.

ACKNOWLEDGMENTS

This work was done under an SERC studentship. I would like to thank my supervisor, Dr. G. S. Hall, for his support and for his reading of the manuscript. I would also like to thank Dr. J. R. Pulham in Aberdeen for useful discussions.

[1]S. W. Hawking, Gen. Relativ. Gravit. **1**, 393 (1971).
[2]D. E. Lerner, Commun. Math. Phys. **32**, 19 (1973).
[3]A. D. Rendall, Ph.D. thesis, University of Aberdeen, 1987.
[4]A submanifold Y of a manifold X is called regular if the manifold topology on Y coincides with the subspace topology inherited from X.
[5]V. I. Arnol'd, "Selected papers," *L. M. S. Lecture Notes*, Vol. 53 (Cambridge U.P., Cambridge, 1981), p. 46ff.
[6]G. S. Hall, J. Phys. A **9**, 541 (1976).
[7]Segré types and their relation to the canonical forms (2.1)–(2.4) are discussed in Ref. 6.
[8]An alternative approach that avoids these calculations is given in Ref. 3.
[9]For an introduction to jet bundles see J. R. Mather, Ann. Math. **89**, 254 (1969) and for their applications to relativity see Ref. 2.
[10]Cf. R. Penrose, Ann. Phys. (NY) **10**, 171 (1960).
[11]M. Spivak, *Differential Geometry Vol. II* (Publish or Perish, Boston, 1970), pp. 4B-177, 4D-7.
[12]M. W. Hirsch, *Differential Topology* (Springer, New York, 1976), Chap. 3.
[13]J. L. Synge, *The Petrov Classification of Gravitational Fields* (Comm. Dublin Institute for Advanced Studies A15, Dublin, 1964).
[14]G. S. Hall, Gen. Relativ. Gravit. **15**, 581 (1983).
[15]G. S. Hall and W. Kay, J. Math. Phys. **29**, 428 (1988).
[16]S. W. Hawking and G. F. R. Ellis, *The Large-Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1973), pp. 89 and 90.
[17]G. S. Hall, Gen. Relativ. Gravit. **16**, 79 (1984).

GH^∞ supermanifolds

Peter Bryant
Department of Pure Mathematics and Mathematical Statistics, Cambridge University, 16 Mill Lane, Cambridge, United Kingdom, CB2 1SB

(Received 17 November 1987; accepted for publication 24 February 1988)

GH^∞-de Witt supermanifolds were recently introduced by Rogers [Commun. Math. Phys. **105**, 375 (1986)]. It is shown that these supermanifolds correspond to objects in a specified category of extended graded manifolds (subject only to the condition $L > 2q$), and some conclusions are drawn about their structure. In particular, it is shown that a GH^∞-de Witt supermanifold is always a vector bundle over its body. Finally it is shown that extended graded manifolds are always trivial extensions of ordinary graded manifolds. This means that all GH^∞-de Witt supermanifolds are in fact H^∞.

I. INTRODUCTION

This is the first of two papers describing the structure of de Witt supermanifolds, i.e., Rogers supermanifolds with ground ring B_L $(L < \infty)$ and trivial topology in odd directions.[1,2]

The machinery described in Ref. 3 is used to relate GH^∞-(de Witt) supermanifolds to "extended" graded manifolds (these are essentially graded manifolds with coefficients in a certain Grassmann algebra $B_{L'}$). As predicted in Ref. 1, this relationship can be put on a formal footing as an equivalence of categories. The points made here are that (i) the definition of an "extended graded manifold" should incorporate a $B_{L'}$-module structure; (ii) since extended graded manifolds of dimension (p,q) are (ordinary) graded manifolds of dimension $(p,q+L')$, the geometrical structure of a GH^∞-supermanifold S (in particular, the existence of a *vector* bundle structure) can be derived by embedding it in an associated H^∞ supermanifold $\hat{S}$; (iii) S itself is H^∞ as soon as it corresponds to a trivially extended graded manifold; and (iv) Batchelor's theorem[4] can be adapted to show that extended graded manifold transition functions can be continuously deformed into ordinary graded manifold transition functions; i.e., every extended graded manifold is the trivial extension of some graded manifold having the same dimension.

II. THE GH^∞ CONDITION

The GH^∞-supersmoothness condition was introduced by Rogers[1] in order to give a simple resolution of some difficulties of a technical nature that arise in G^∞-theory. The GH^∞ condition is defined as follows.

Let L denote a finite integer (considered to be sufficiently large that accidental cancellations in practical calculations are avoided) and $L' = [L/2]$ the largest integer no bigger than $L/2$. Let B_L (resp. $B_{L'}$) denote the Grassmann algebra on L (resp. L') real variables: $B_L = \Lambda R^L$ with its fine (= vector space = Banach algebra) topology. Here B_L is a $B_{L'}$ algebra through $i_{L',L}: B_{L'} \to B_L$ given by extending a set of generators for $B_{L'}$ to a set of generators for B_L.

Rogers observed that provided $2q < L$ (which will invariably be assumed in what follows) the z expansion

$$Z: C^\infty(R^p) \otimes \Lambda R^q \otimes B_{L'} \to G^\infty(B_L^{p,q})$$

is an *isomorphism* onto its image, which we denote $GH^\infty(B_L^{p,q})$.

Here Z is the restriction of the map that realizes the full algebra $C^\infty(R^p) \otimes \Lambda R^q \otimes B_L$ as an algebra of functions on $B_L^{p,q}$ and $G^\infty(B_L^{p,q})$ is the algebra of functions $B_L^{p,q} \to B_L$ that are supersmooth (i.e., G^∞).[5] Loosely speaking, G^∞ means smooth as a map between super-Banach spaces.

Here Z is defined[5,6] by sending the standard projections $x_1, x_2, \ldots, x_p: R^p \to R$ to the corresponding even projections $B_L^{p,q} \to (B_L)_0$, elements $\pi_1, \pi_2, \ldots, \pi_q$ of a basis for $\Lambda^1 R^q$ to odd projections $B_L^{p,q} \to (B_L)_1$ and simple multiplication by elements of $B_{L'}$.

Assuming $2q < L$, L' is the largest integer for which this trick works. To be more precise, Z is injective whenever $q + L' < L$ and this condition is equivalent to the condition $2q < L$ as a simple check reveals (do even and odd L separately).

Since Z is injective, Der GH^∞ is a *free* GH^∞ module that therefore resolves one difficulty that we encounter in G^∞ theory; see Ref. 7 for an alternative approach.

Replacing $L' = [L/2]$ by $L' = 0$ in the above discussion defines the H^∞ class discussed in detail in Ref. 3. The problem now is that constant B_L-valued functions fail to be supersmooth in the H^∞ sense. In GH^∞ this problem is circumvented by allowing some supersmooth constant functions (but not too many); $GH^\infty(B_L^{p,q})$ is a $B_{L'}$ algebra whereas $H^\infty(B_L^{p,q})$ is just an R algebra. Hence GH^∞ has the practical advantages of G^∞ over H^∞ (for modeling of supersymmetry transformations) without its technical drawbacks.

In Ref. 1 Rogers remarks that it should be possible to put the (conjectured) relationship between GH^∞-(de Witt) supermanifolds and a category of extended graded manifolds—yet to be defined—on a formal footing as an equivalence of categories. This is the first aim of the present paper.

III. EXTENDED GRADED MANIFOLDS

Locally, extended graded manifolds are just graded manifolds sheaves $\mathscr{A}$ with a factor $B_{L'}$ tensored on; the global conditions were not spelled out in Ref. 1. To motivate our definition of extended graded manifold we make the following remark.

 0022-2488/88/071575-05$02.50

Let S be a GH^∞-(de Witt) supermanifold: that is, S is a de Witt supermanifold whose transition functions all have GH^∞ components. Here S has a body X (see Refs. 5, 6, and 8) and the structure sheaf of S (i.e., the sheaf GH^∞_S of $B_{L'}$-valued GH^∞-supersmooth functions on S) is a $B_{L'}$ module. Then the direct image $\varepsilon_* GH^\infty_S$ of the structure sheaf along the body map $\varepsilon: S \to X$ is also a $B_{L'}$ module. By decree this is to be an extended graded manifold and hence we demand from the outset that extended graded manifolds are Z_2-ringed spaces over $B_{L'}$, i.e., sheaves of Z_2-graded-commutative $B_{L'}$ algebras.

Suppose L is finite and $L' = [L/2]$ as usual.

Definition 3.1: An *extended graded manifold* (w.r.t. L) of dimension (p,q) is a Z_2-ringed space over $B_{L'}$, $(X,\mathscr{B})$, consisting of a nice (Hausdorff, smooth, paracompact) manifold X of dimension p and a sheaf $\mathscr{B}$ of graded-commutative $B_{L'}$ algebras together with (i) a map of sheaves of algebras $\varepsilon: \mathscr{B} \to C^\infty_X$ (augmentation), and (ii) an open cover $\{U_i\}$ of X and isomorphisms of sheaves of graded $B_{L'}$ algebras (local trivializations)

$$T_i: \mathscr{B}_{|U_i} \to C^\infty_X \otimes \Lambda R^q \otimes B_{L'|U_i}.$$

For convenience as part of the definition we demand that $L > 2q$.

Examples 3.2: (i) If $(X,\mathscr{A})$ is an ordinary graded manifold[9] and L is an integer exceeding twice its odd dimension, one may easily extend the data for the graded manifold (GM) $(X,\mathscr{A})$ to the data for an extended graded manifold over X with structure sheaf $\mathscr{B} = \mathscr{A} \otimes B_{L'}$. Such examples of extended graded manifolds will be called trivially extended graded manifolds. As a concrete example let $\mathscr{A}$ be the sheaf of smooth differential forms on the nice manifold X.

(ii) If S is a GH^∞ supermanifold of dimension (p,q) over B_L $(L > 2q)$ with body map $\varepsilon: S \to X$, then $\varepsilon_* GH^\infty_S$ is an extended graded manifold of dimension (p,q) w.r.t. L.

It will be shown eventually that these two examples (simultaneously) cover all possible extended graded manifolds. That is, every extended graded manifold $(X,\mathscr{B})$ is trivially extended $\mathscr{B} = \mathscr{A} \otimes B_{L'}$ and to $(X,\mathscr{B})$ one can associate a GH^∞ supermanifold on which the algebra $\mathscr{B}(X)$ is represented as the algebra of globally supersmooth functions.

The category $\mathbf{EGM(L)}$ of extended graded manifolds w.r.t. L is the category whose objects are defined as in Definition 3.1 and whose morphisms are morphisms as Z_2-ringed spaces[10] over $B_{L'}$. The justification for such a choice is to be found in Proposition 3.3 (iv) below.

Trivial extension is evidently a functor $\mathbf{GM}_{(2q<L)} \to \mathbf{EGM(L)}$.

On the other hand, $\Lambda R^q \otimes B_{L'} = \Lambda R^{q+L'}$ so that one can reinterpret each extended graded manifold $(X,\mathscr{B})$ of dimension (p,q) as an ordinary graded manifold of dimension $(p,q,+L')$. This gives a forgetful functor $\mathbf{EGM(L)}^{(p,q)} \to \mathbf{GM}^{(p,q+L')}$ (the superscript refers to graded dimension). The point to make immediately is that if $\mathscr{B}(X)$ is the $B_{L'}$ algebra of global sections of some extended graded manifold then it is also the algebra of global sections of some graded manifold and thus enjoys all of the algebraic properties noted in Ref. 10. In particular, note the following proposition.

Proposition 3.3: If $(X,\mathscr{B})$ is an extended graded manifold then (i) $\mathscr{B}$ is a fine sheaf (i.e., it admits partitions of unity); (ii) the augmentation $\varepsilon: \mathscr{B}(X) \to C^\infty(X)$ is surjective; (iii) $\mathscr{B}(X)$ is a graded-commutative $B_{L'}$ algebra whose augmentation induces a bijection between X and $\mathrm{Alg}(\mathscr{B}(X),R)$; (iv) $(X,\mathscr{B})$ is a Z_2-locally ringed space and we can identify the stalk $\mathscr{B}_x$ at a point x in X with the local algebra $B(X)/O_x$, where $O_x = \{a \in \mathscr{B}(X)$: *there is b not in* m_x *with* $ab = 0\}$; (v) whenever (C,ε_0) is an augmented Z_2-graded algebra and $\alpha: \mathscr{B}(X) \to C$ is a graded algebra map, α factors through the stalk map $\rho_x: \mathscr{B}(X) \to \mathscr{B}_x$ where x is the point corresponding to $\varepsilon_0 \cdot \alpha$ under the identification of (ii); and (vi) there is a one-to-one correspondence between maps $(X,\mathscr{B}) \to (X',\mathscr{B}')$ of extended graded manifolds and maps $\mathscr{B}'(X') \to \mathscr{B}(X)$ of graded $B_{L'}$ algebras.

Proof (sketch): The augmentation $\varepsilon: \mathscr{B} \to C^\infty_X$ has local sections and it follows that a smooth partition of unity over X can be lifted to a partition of unity for $\mathscr{B}$ (i) and hence that ε is globally surjective (ii). Next $\ker \varepsilon = \mathscr{B}_1^2 \otimes \mathscr{B}_1$, so every algebra map $\alpha: \mathscr{B}(X) \to R$ kills nilpotent elements and thus factors through ε. But every algebra map $\overline{\alpha}: C^\infty(X) \to R$ is an evaluation map, $\overline{\alpha}(f) = f(x)$ for some x (iii). Write m_x for the maximal ideal that is the kernel of the algebra map $b \to \varepsilon(b)(x)$. Since $\rho_x: \mathscr{B}(X) \to \mathscr{B}_x$ is surjective the stalk $\mathscr{B}_x$ is determined by $\ker \rho_x$ and this is O_x (iv). If $\alpha: \mathscr{B}(X) \to C$ is a graded algebra map, α vanishes on O_x precisely when $\alpha(b)$ is invertible in C [i.e., $\varepsilon_0(\alpha(b)) \neq 0$] for each b not in m_x. This is the case when x is related to α as indicated (v). Starting from a map of graded $B_{L'}$-algebras $\psi: \mathscr{B}'(X') \to \mathscr{B}(X)$ we use the algebraic description of stalks in (iv) to define an associated map on stalks. This is enough to determine a map of Z_2-ringed spaces from ψ (vi).

IV. THE EQUIVALENCE THEOREM

Graded manifolds and supermanifolds are related[3,6] using sets of graded algebra maps with the weak topology.[3] If $(X,\mathscr{A})$ is an ordinary graded manifold of dimension (p,q) and $L > q$ then the corresponding H^∞-de Witt supermanifold $z(X,\mathscr{A})$ is given by $\mathrm{Alg}(\mathscr{A}(X),B_L)$. The weak topology is given by pulling back the topology on B_L (coarse or fine) through evaluation maps $\mathrm{Alg}(\mathscr{A}(X),B_L) \to B_L$. This is a more convenient description of the topology than will be found in the original presentation[6] of the coarse case (in which H^∞ was described as M^∞).

We shall concentrate on Hausdorff supermanifolds (fine B_L) although completely analogous results may be obtained in the coarse case.[6,11]

The algebraic properties mentioned in the previous sections suffice to reduce the problem to a local one where the following observation is crucial. Recall that there is a natural homeomorphism[6] $\eta: B_L^{p,q} \to \mathrm{Alg}(C^\infty(R^p) \otimes \Lambda R^q, B_L)$ that is a local model for the z functor. In the present case η evidently extends to give a natural homeomorphism $\eta: B_L^{p,q} \to \mathrm{Alg}_{B_{L'}}(C^\infty(R^p) \otimes \Lambda R^q \otimes B_{L'}, B_L)$ (the latter denotes the set of maps of graded $B_{L'}$ algebras). Here η is given by

$\langle \eta s, f\rangle = \langle Zf, s\rangle$ whenever $s \in B_L^{p,q}$ and $f \in C^\infty(R^p) \otimes \Lambda R^q$ or $C^\infty(R^p) \otimes \Lambda R^q \otimes B_{L'}$.

Let $\mathbf{GH}^\infty\ \mathbf{dW}(B_L)$ denote the category of GH^∞-de Witt supermanifolds whose odd dimension q satisfies $2q < L$. If S is such a supermanifold let $\varepsilon: S \to X$ be its body map.

Our first result is the following theorem.

Theorem 4.1: There is an equivalence of categories Ez: $\mathbf{EGM(L)} \to \mathbf{GH}^\infty\ \mathbf{dW}(B_L)$ given by

$$(X, \mathscr{B}) \to \mathrm{Alg}_{B_{L'}}(\mathscr{B}(X), B_L),$$

$$(S, GH_s^\infty) \to (X, \varepsilon_* GH_s^\infty).$$

Proof: Each $\alpha \in \mathrm{Alg}_{B_{L'}}(\mathscr{B}(X), B_L)$ lies inside some $\mathrm{Alg}_{B_{L'}}(\mathscr{B}(U_i), B_L)$ by 3.3 (v). Then η and T_i combine to give a local modeling of $Ez(X, \mathscr{B})$ on some super-Euclidean space. This gives $Ez(X, \mathscr{B})$ a GH^∞-de Witt supermanifold structure. Conversely, if (S, GH_s^∞) is a GH^∞-de Witt supermanifold, the evaluation map

$$S \to \mathrm{Alg}_{B_{L'}}(GH^\infty(S), B_L)$$

has an inverse found by reducing to the local case through 3.3 again. It follows that the indicated evaluation map is a natural homeomorphism.

The functor property depends on the fact that GH^∞-supersmooth maps $S \to S'$ are precisely maps of Z_2-ringed spaces $(S, GH_s^\infty) \to (S', GH_{s'}^\infty)$ over $B_{L'}$.

Corollary 4.2: The body of S has the algebraic representation $X = \mathrm{Alg}(GH^\infty(S), R)$.

We begin to draw some conclusions about the structure of GH^∞-de Witt supermanifolds from Theorem 4.1.

If we have an extended graded manifold $(X, \mathscr{B})$ w.r.t. L of dimension (p,q) then

$$Ez(X, \mathscr{B}) = \mathrm{Alg}_{B_{L'}}(\mathscr{B}(X), B_L) \subset \mathrm{Alg}_R(\mathscr{B}(X), B_L).$$

It follows that the GH^∞ supermanifold corresponding to $(X, \mathscr{B}) \in \mathbf{EGM(L)}$ sits inside the H^∞ supermanifold corresponding to $(X, \mathscr{B}) \in \mathbf{GM}^{(p, q+L')}$ [(2.15) of Ref. 3 applies because $q + L' < L$ whenever $2q < L$)]. This gives the following proposition.

Proposition 4.3: Every $S \in \mathbf{GH}^\infty\ \mathbf{dW}(B_L)$ sits as a subvector bundle of some $\hat{S} \in \mathbf{H}^\infty \mathbf{dW}(B_L)$ that has the same body.

Proof: Here S and $\hat{S}$ [defined by considering $(X, \mathscr{B})$ as an extended graded manifold and as a graded manifold of higher odd dimension, respectively] have the same body X (and thus the same even dimension). Also $\hat{S}$ is known to be a strict vector bundle[12] but the GH^∞-transition functions for S are just restrictions of the H^∞-transition functions for $\hat{S}$. It follows that the action of transition functions for S on the fiber over $x \in X$ say is linear provided that this fiber is a subvectorspace of the fiber of $\hat{S}$ over x. This is a simple check.

One might be tempted also to conclude that the (H^∞)-transition functions on $\hat{S}$ restrict to H^∞-transition functions on S. The following example warns against this temptation.

Example 4.4: Assume $L > 2$ in the following brief discussion. Let S be the GH^∞-supermanifold $B_L^{1,1}$ so that $\hat{S} = B_L^{1,1+L'}$ considered as an H^∞ supermanifold. Then the GH^∞ map of $B_L^{1,1}$ (ε some odd Grassmann number)

$$x \to x + \varepsilon\theta, \quad \theta \to \theta + \varepsilon,$$

may be interpreted as the restriction to $B_L^{1,1}$ of the H^∞ map on $B_L^{1,1+L'}$ given by adding trivial relations $\varepsilon_2 \to \varepsilon_2$, etc. to the above.

The indicated map (a primitive analog of a supersymmetry transformation) is not H^∞ since it is parametrized by the odd Grassmann number ε. The rather trivial observation that S sits inside some $\hat{S}$ is the global analog of a familiar trick from supersymmetry; namely, one can regard the parameters of a supersymmetry transformation as additional odd coordinates.

One consequence of the fact that graded manifolds and exterior vector bundles are equivalent (Batchelor's theorem) is that every H^∞-de Witt supermanifold with body X can be reconstructed as a vector bundle over X pulled back via $zX \to X$ (Refs. 3 and 12). Here zX is the z extension zX (Ref. 12) of X, the supermanifold of dimension $(\dim X, 0)$ associated with the graded manifold (X, C_X^∞), namely $zX = \mathrm{Alg}(C^\infty(X), B_L)$.

Although a similar statement is not possible for GH^∞-de Witt supermanifolds (or rather those that are not trivially extended) we do have the following proposition.

Proposition 4.5: If S is a GH^∞-de Witt supermanifold its body map $\varepsilon: S \to X$ factors through the natural projection $zX \to X$.

Proof: The algebra $GH^\infty(S)$ is the algebra of global sections not only of some extended graded manifold [dimension (p,q) say] but also of some graded manifold of dimension $(p, q+L')$. Recalling that S has body $X = \mathrm{Alg}(GH^\infty(S), R)$, consider the map $Q: GH^\infty(S) \to C^\infty(X)$ simply given by $\langle Qg, x\rangle = \langle x, g\rangle$. This is the graded manifold augmentation map and Batchelor's theorem ensures that Q has a section j: $C^\infty(X) \to GH^\infty(S)$.

Here j gives rise to a map j^*: $\mathrm{Alg}_{B_{L'}}(GH^\infty(S), B_L) \to \mathrm{Alg}(C^\infty(X), B_L)$ through which the body map ε factors. (Despite the fact that there is no sense in which a canonical choice for j can be made, j^* is a completely natural map since it can be identified with the projection arising from the θ foliation[13] whenever this is regular.[8])

Let $(X, \mathscr{B})$ be the extended graded manifold associated to the GH^∞-supermanifold S (3.2ii). Then $z(X, \mathscr{B})$ (i.e., the H^∞-supermanifold $\hat{S}$ of 4.2) is constructed as the vector bundle $E(B)^* \otimes B_1$ pulled back to zX where $E(B)$ denotes the vector bundle over X corresponding to the C_X^∞-module N/N^2, N the sheaf of nilpotents in $\mathscr{B}$. It can be shown that (unless B is trivially extended) the image of $Ez(X, \mathscr{B})$ under the natural projection h: $z(X, \mathscr{B}) \to E(B) \otimes B_1^*$ is *not* a subbundle of $E(B)^* \otimes B_1$.

Suppose that $\mathscr{B}$ is trivially extended, $\mathscr{B} = \mathscr{A} \otimes B_{L'}$. Then if $E(A)$ is the vector bundle over X constructed from the nilpotent sheaf $N(A)$ of $\mathscr{A}$ in the same way as above, the image under h of $Ez(X, \mathscr{B})$ is $E(A)^* \otimes B_1$ and $Ez(X, \mathscr{B})$ is indeed the pullback of this bundle under $zX \to X$. This is a consequence of the following proposition.

Proposition 4.6: Suppose S is a GH^∞-de Witt supermanifold whose associated extended graded manifold $(X, \mathscr{B})$ is trivially extended. Then S is H^∞.

Proof:

$$S = Ez(X,\mathscr{B}) = \mathrm{Alg}_{B_{L'}}(\mathscr{B}(X),B_L)$$
$$= \mathrm{Alg}_{B_{L'}}(\mathscr{A}\otimes B_{L'},B_L)$$
$$= \mathrm{Alg}(\mathscr{A},B_L) = z(X\mathscr{A})$$

and this is H^∞.

It follows from this easy proposition that in order to find GH^∞-de Witt supermanifolds that are not H^∞ we must look for extended graded manifolds that are not trivially extended. We show that this is not possible in the next section.

V. TRIVIALITY OF EXTENDED GRADED MANIFOLDS

In this section we will consider the structure group of extended graded manifolds over X (w.r.t. L) of odd dimension q and compare it with the structure group of graded manifolds over X of the same odd dimension.

There is a one-to-one correspondence between isomorphism classes of graded manifolds over X with odd dimension q and elements of the first Cech cohomology set $H^1(X;\mathrm{AUT}[C_X^\infty\otimes\Lambda R^q])$. That is the structure group for ordinary graded manifolds is $\mathrm{AUT}[C_X^\infty\otimes\Lambda R^q]$, the sheaf of automorphisms of $C_X^\infty\otimes\Lambda R^q$ as a sheaf of Z_2-graded R algebras. We recall the main result of Ref. 4.

Theorem 5.1 (Batchelor's theorem): If $\mathrm{Gl}[q;C_X^\infty]$ is identified with the sheaf of automorphisms of $C_X^\infty\otimes\Lambda R^q$ that respect both the C_X^∞-module structure and the Z grading then we have a bijection of cohomology sets

$$H^1(X;\mathrm{AUT}[C_X^\infty\otimes\Lambda R^q]) = H^1(X;\mathrm{Gl}[q;C_X^\infty]).$$

Corollary 5.2: Every isomorphism class of graded manifolds contains an exterior vector bundle. In particular, for every graded manifold $(X,\mathscr{A})$, the GM augmentation $\varepsilon\colon \mathscr{A}(X)\to C^\infty(X)$ has a section j.

Turning to extended graded manifolds we find that the relevant structure group is $\mathrm{AUT}_{B_{L'}}[C_X^\infty\otimes\Lambda R^q\otimes B_{L'}]$ (the set of automorphisms that respect the $B_{L'}$-module structure). Clearly isomorphism classes of extended graded manifolds are in one-to-one correspondence with elements of the Cech cohomology set $H^1(X;\mathrm{AUT}_{B_{L'}}[C_X^\infty\otimes\Lambda R^q\otimes B_{L'}])$. The problem is to identify this set with each of the cohomology sets in 5.1.

The unit map j and the exterior augmentation $\varepsilon_0'\colon B_{L'}\to R$ induce group homomorphisms

$$\mathrm{AUT}[C^\infty(X)\otimes\Lambda R^q]$$
$$\rightleftarrows \mathrm{AUT}_{B_{L'}}[C^\infty(X)\otimes\Lambda R^q\otimes B_{L'}].$$

The claim is that these maps j and ε_0' are quasi-inverses.

Theorem 5.3: Here j and ε_0' induce inverse bijections on cohomology

$$H^1(X;\mathrm{AUT}[C_X\otimes\Lambda R^q])$$
$$\rightleftarrows H^1(X;\mathrm{AUT}_{B_{L'}}[C_X^\infty\otimes\Lambda R^q\otimes B_{L'}]).$$

Proof: Since $\varepsilon_0'\cdot j = \mathrm{id}$ on automorphism groups we have $H^1\varepsilon_0'\cdot H^1 j = \mathrm{id}$ on cohomology sets. It follows that

$$H^1 j\colon H^1(X;\mathrm{AUT}[C_X^\infty\otimes\Lambda R^q])$$
$$\to H^1(X;\mathrm{AUT}_{B_{L'}}[C_X^\infty\otimes\Lambda R^q\otimes B_{L'}])$$

is injective.

The idea is to use Batchelor's theorem (for odd dimension $q+L'$) to see that $H^1 j$ is also surjective.

Define $\mathrm{Bl}[q;L';C_X^\infty]$ to be the set of invertible $(q+L')\times(q+L')$ matrices with values in C_X^∞ that has the following block form:

$$\begin{bmatrix} X & Y \\ 0 & I \end{bmatrix},$$

where X is an invertible $q\times q$ matrix, $X\in\mathrm{Gl}[q;C_X^\infty]$ and Y is some $q\times L'$ matrix.

Observe that

$$\mathrm{Bl}[q;L';C_X^\infty] = \mathrm{Gl}[q+L';C_X^\infty]$$
$$\cap\mathrm{AUT}_{B_{L'}}[C_X^\infty\otimes\Lambda R^q\otimes B_{L'}].$$

More explicitly let $\pi_1,\ldots,\pi_q$ be generators of ΛR^q and $f_1,\ldots,f_{L'}$ generators of $B_{L'}$. Then the automorphism of $C_X^\infty\otimes\Lambda R^q\otimes B_{L'}$ associated with the matrix $\left[\begin{smallmatrix} X & Y \\ 0 & I \end{smallmatrix}\right]$ is generated (in matrix notation) by

$$\pi\to X\pi + Yf,\quad f\to f.$$

These are the most general ($B_{L'}$ linear) automorphisms of $C_X^\infty\otimes\Lambda R^q\otimes B_{L'}$ that are automorphisms of $C_X^\infty\otimes\Lambda R^{q+L'}$ and respect both the C_X^∞ module structure and the Z grading.

Batchelor's theorem applied in odd dimension $q+L'$ implies that we can make the identification

$$H^1(X;\mathrm{AUT}_{B_{L'}}[C_X^\infty\otimes\Lambda R^q\otimes B_{L'}])$$
$$= H^1(X;\mathrm{Bl}[q;L';C_X^\infty]).$$

This is the content of Proposition 4.2.

What remains is to show that the Y part of the automorphism group is contractible. As far as the manifold X is concerned the generators $f_1,\ldots,f_{L'}$ of $B_{L'}$ are constants and we can premultiply them by any scalar factor. The effect of the change of $B_{L'}$ generators is to absorb a scalar factor into the definition of Y. It follows that we can make this as small as we like by choosing a suitable scalar factor. By smoothly varying the scalar factor one shows that indeed the Y part is contractible. To be slightly more rigorous one needs to consider the kernel K of the sheaf epimorphism $\mathrm{Bl}(q;L';C_X^\infty)\to\mathrm{Gl}(q;C_X^\infty)$ and observe that K (the matrices $\left[\begin{smallmatrix} I & Y \\ 0 & I \end{smallmatrix}\right]$ with Y an *arbitrary* $q\times L'$ matrix) is a fine sheaf. Also conjugation by any element of $\mathrm{Bl}(q;L';C_X^\infty)$ maps K onto itself so 3.8 of Ref. 4 completes the proof.

We have shown that there are identifications

$$H^1(X;\mathrm{Gl}[q;C_X^\infty]) = H^1(X;\mathrm{Bl}[q;L';C_X^\infty])$$
$$= H^1(X;\mathrm{AUT}_{B_{L'}}[C_X^\infty\otimes\Lambda R^q\otimes B_{L'}]).$$

The result now follows from a direct application of Batchelor's theorem (in odd dimension q).

The implication for the structure of GH^∞-de Witt supermanifolds is the following corollary.

Corollary 5.4: Every GH^∞-de Witt supermanifold over B_L (whose odd dimension q satisfies $2q<L$) is H^∞.

Proof: See 4.6.

It is known that there is no way of finding a *canonical* choice of vector bundle in the isomorphism class of a given graded manifold—or section j of ε—in Corollary 5.2. Similarly there is no *canonical* way of choosing new coordinates on a GH^∞-de Witt supermanifold that exhibit its H^∞ structure.

If we choose to work in the holomorphic category there will of course be the usual cohomological obstructions to results in this section [even when $(p,q) = (1,1)$].

In fact the super Riemann surfaces[14] of Crane and Rabin provide examples of GH^∞-de Witt supermanifolds (with complex coefficients). The canonical super-Riemann surfaces arising from (trivially extended) graded Riemann surfaces[15] are the only super-Riemann surfaces that are holomorphic vector bundles over their bodies.

To each super-Riemann surface M with body X we can associate an extended holomorphic graded manifold $(X,\mathscr{B})$ using the sheaf of superanalytic functions on M [see example 3.2 (ii)]. This is a trivial extension only when M is canonical.

VI. CONCLUSIONS

We end with a few remarks assessing the possible significance of the results in this paper.

The GH^∞ concept is not redundant. First, not all supermanifolds are of de Witt type. Second, the special features of GH^∞ that led to its original introduction are still important whether we are talking about de Witt supermanifolds or not.

Our results do not just tell us about the structure of GH^∞-de Witt supermanifolds over a fixed B_L. It is now possible to compute the direct limit of these categories as L goes to infinity: it was precisely this (formal) limit that Rogers introduced in Ref. 1. The result is the following theorem.

Theorem 6.1: The Rogers direct limit of the categories $\mathbf{GH^\infty dW}(B_L)$ is **GM** (or any category equivalent to it).

Here **GM** really does mean the category of all graded manifolds: the condition $2q < L$ becomes vacuous as L tends to infinity. Especially interesting is the observation[3] that **GM** is equivalent to the category of H^∞-de Witt supermanifolds over the formal Grassmann algebra $B_\infty = \Lambda R^\infty$. This means that the limit category in 6.1 can be represented as a category of supermanifolds over an infinite-dimensional ground ring.

We emphasize that GH^∞ makes no sense when L is taken to be ∞ (for then L' is ill defined). However in this case the need to use GH^∞ is lost because the technical blemishes of the full G^∞ theory disappear. Indeed the category of G^∞-de Witt supermanifolds over B_∞ can be shown to be equivalent to a category of "fully extended" graded manifolds (defined by 3.1 with $L' = \infty$).

Theorem 5.3 is an important step in analyzing the structure of G^∞-de Witt supermanifolds over B_L since the computation of the cohomology sets in 5.3 is actually independent of the specific choice $L' = [L/2]$. This suggests that *all* de Witt supermanifolds over B_L $(L < \infty)$ are H^∞. We have recently shown[16] that this is indeed the case: the crucial observation being that we can identify $\mathrm{AUT}[G^\infty(B_L^{p,q})]$ with $\mathrm{AUT}[C^\infty(R^p) \otimes \Lambda R^q \otimes B_L]$ despite the lack of injectivity of the G^∞ z expansion.

Because in the G^∞ case we can no longer use an embedding $S \to \hat{S}$ of the form employed here (see 4.3) some way of bypassing this step needs to be found (see Ref. 16 for details). This means that we can even omit the minor restriction $2q < L$ from results in this paper.

ACKNOWLEDGMENTS

Marjorie Batchelor is thanked for suggesting improvements in the layout of the text and for pointing out an error in an earlier draft of this paper.

This research was supported by the Science and Engineering Research Council of Great Britain.

[1] A. Rogers, "Graded manifolds, supermanifolds and infinite dimensional Grassmann algebras," Commun. Math. Phys. **105**, 375 (1986).

[2] J. Rabin and L. Crane, "How different are the supermanifolds of Rogers and de Witt?," Commun. Math. Phys. **102**, 123 (1985).

[3] P. Bryant, "De Witt supermanifolds and infinite-dimensional ground rings," to appear in J. London Math. Soc.

[4] M. Batchelor, "The structure of supermanifolds," Trans. Am. Math. Soc. **253**, 329 (1979).

[5] A. Rogers, "A global topology for supermanifolds," J. Math. Phys. **21**, 1352 (1979).

[6] M. Batchelor, "Two approaches to supermanifolds," Trans. Am. Math. Soc. **258**, 257 (1979).

[7] M. Rothstein, "The axioms of supermanifolds and a new structure arising from them," Washington University preprint, 1985.

[8] R. Catenacci, C. Reina, and P. Teofilatto, "On the body of supermanifolds," J. Math. Phys. **26**, 671 (1985).

[9] B. Kostant, "Graded manifolds, graded Lie groups and prequantisation," in *Differential Geometric Methods in Mathematical Physics, Springer Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977), pp. 177–306.

[10] P. Bryant, "Sheaf representations and graded manifolds," J. Pure Appl. Alg. **47**,213 (1987).

[11] B. de Witt, *Supermanifolds* (Cambridge U.P., Cambridge, 1984).

[12] M. Batchelor, "Graded manifolds and supermanifolds," in *Mathematical Aspects of Superspace* edited by C. J. S. Clarke, A. Rosenblum, and H. J. Seifert (Reidel, Dordrecht, 1981).

[13] J. Rabin, "Berezin integration on general fermionic supermanifolds," Commun. Math. Phys. **103**, 431 (1986).

[14] L. Crane and J. Rabin, "Super Riemann surfaces: Uniformisation and Teichmuller theory," Commun. Math. Phys. **113**, 601 (1988).

[15] M. Batchelor and P. Bryant," Graded Riemann surfaces," Commun. Math. Phys. **114**, 243 (1988).

[16] P. Bryant, "The structure of de Witt supermanifolds," to appear in Riv. Mat. Pura Appl.

Invariants for the time-dependent harmonic oscillator

R. K. Colegrave and M. A. Mannan
Department of Mathematics, King's College, Strand, London WC2R 2LS, England

(Received 29 October 1987; accepted for publication 16 March 1988)

A systematic method is presented for the construction of invariants for the damped oscillator under the action of a driving force and for the N-dimensional isotropic or anisotropic oscillator. Invariants for time-dependent oscillators are obtained by canonical transformation. The treatment holds in both classical and quantum mechanics.

I. INTRODUCTION

Variable-mass and variable-frequency oscillators[1,2] are important in quantum optics,[3–6] as well as in many other fields, such as quantum chemistry. First integrals (referred to as invariants) are needed for systems such as the damped weakly pulsating oscillator described by a Kanai–Caldirola Hamiltonian with mass

$$m(t) = m_0 \exp(2\gamma t + 2\epsilon \sin \nu t), \quad \epsilon \ll 1. \tag{1.1}$$

Second-degree energylike invariants can effectively replace the Hamiltonian in the seeking of solutions of the Schrödinger equation.[7]

Any time-dependent oscillator (TDO) may be reduced to a simple harmonic oscillator (SHO) with Hamiltonian

$$H(q,p) = \tfrac{1}{2}p^2/m_0 + \tfrac{1}{2}m_0\omega_0^2 q^2, \tag{1.2}$$

by a generalized canonical transformation.[1,8] In principle, therefore, all possible first integrals of TDO are known in terms of the fundamental SHO invariants,

$$I(q,\bar{p},t) = q \cos \omega_0 t - \bar{p} \sin \omega_0 t \quad (\bar{p} = p/m_0\omega_0), \tag{1.3a}$$

$$J(q,\bar{p},t) = q \sin \omega_0 t + \bar{p} \cos \omega_0 t, \tag{1.3b}$$

or, alternatively, in terms of the complex invariant

$$C(q,\bar{p},t) = I + iJ = (q + i\bar{p})e^{i\omega_0 t}. \tag{1.3c}$$

Here C and C^* ($C^\dagger$ in quantum mechanics) are Dirac variables with evolution in negative time, which evaluates them at a previous time $t = 0$. In the present paper we shall extend this simple idea in a number of ways. We shall consider the N-dimensional oscillator and extend (1.3a) and (1.3b) to the case when a periodic driving force is applied to a damped oscillator. Bohlin's integral[9] will be found in a natural way. Invariants to first order in ϵ will be found for the system with mass given by (1.1). We shall concentrate on finding explicit forms for invariants.

Some previous work on the damped harmonic oscillator[4] is extended. The hierarchy of invariants introduced by Colegrave *et al.*[10] are seen very simply for the SHO (or for the damped oscillator) in terms of the complex invariants (1.3c) in Sec. III, where we construct all possible first integrals for the N-dimensional simple harmonic oscillator. Constants of the motion, consisting of angular momenta and elements of the Fradkin tensor,[11–14] are obtained as special time-independent combinations of complex invariants. Again, damping is easily included. The anisotropic oscillator is discussed in Sec. IV.

First integrals derived from the SHO by transformation[1,8] are expressed in general in terms of an auxiliary function ρ that satisfies an equation of the type, cf. (2.18) or (2.21) in Ref. 4,

$$\ddot{\rho} + \omega^2(t)\rho = \rho^{-3}. \tag{1.4}$$

However, in addition to the case of the damped oscillator, explicit forms for invariants can be found for the strongly pulsating oscillator[15,16] as shown in Sec. V. Some further cases listed in Ref. 1 would also lead to explicit invariants. The Lewis invariant,[17–19] which is not in general explicit, is obtained in Sec. VI and agreement is obtained with Ref. 4 in the case of variable-mass and variable-frequency systems. Finally, in Sec. VII, we consider invariants to order ϵ for the important case of variable mass given by (1.1).

The connection of invariants with symmetries of the Hamiltonian is discussed by many authors, including Eliezer and Gray,[20] Leach,[14,21–25] Lutzky,[26] and Prince and Eliezer.[27] A group-theoretic treatment of invariants for the damped harmonic oscillator is given by Cervero and Villarroel.[28]

Our treatment holds equally well in classical mechanics and in the Heisenberg picture of quantum mechanics. In Secs. II–IV to avoid cumbersome equations, we give a classical discussion, indicating the translation to quantum mechanics only when necessary. In Secs. V and VI we employ quantum-mechanical notation.

II. FUNDAMENTAL INVARIANTS FOR THE DRIVEN DAMPED OSCILLATOR

It is more instructive to begin not with the SHO, but with the more realistic problem of a damped oscillator with a driving force. We shall obtain the fundamental first-degree invariants for this system, together with a complex combination analogous to (1.3c).

An oscillator of mass m_0 with damping γ ($< \omega_0$) acted on by a force $f(t)$ per unit mass is described by the Kanai–Caldirola Hamiltonian,[16,29–32]

$$H(q,p,t) = \tfrac{1}{2}p^2/m(t) + \tfrac{1}{2}m(t)\omega_0^2 q^2 - m(t)f(t)q, \tag{2.1a}$$

where the mass is increased according to

$$m(t) = m_0 \exp(2\gamma t) \quad (\gamma > 0). \tag{2.1b}$$

The elimination of p from the Hamiltonian (or Heisenberg) equations

 0022-2488/88/071580-08$02.50

$$\dot{q} = \frac{\partial H}{\partial p} = \frac{p}{m(t)}, \tag{2.2a}$$

$$\dot{p} = -\frac{\partial H}{\partial q} = -m(t)\omega_0^2 q + m(t)f(t) \tag{2.2b}$$

leads to the correct equation of motion,

$$\ddot{q} + 2\gamma\dot{q} + \omega_0^2 q = f(t). \tag{2.3}$$

We write, as in Refs. 4 and 10,

$$\omega^2 = \omega_0^2 - \gamma^2, \tag{2.4a}$$

$$\bar{p} = p/(m_0\omega_0) = \dot{q}e^{2\gamma t}/\omega_0 \tag{2.4b}$$

and apply the periodic driving force

$$f(t) = f_0 \cos(\lambda t + \phi). \tag{2.5}$$

Then, solving (2.3) with (2.2a) and (2.4b) gives

$$q = e^{-\gamma t}(I\cos\omega t + J\sin\omega t) + (f_0/z^2)\cos(\lambda t + \phi - \phi'), \tag{2.6a}$$

$$\bar{p} = (e^{\gamma t}/\omega_0)[(\omega J - \gamma I)\cos\omega t - (\omega I + \gamma J)\sin\omega t] - [\lambda f_0/(\omega_0 z^2)]\sin(\lambda t + \phi - \phi'), \tag{2.6b}$$

where I and J are arbitrary, expressible in terms of $q(0)$ and $p(0)$, and

$$\phi' = \arctan(2\gamma\lambda/\tilde{\omega}^2), \quad \tilde{\omega}^2 = \omega_0^2 - \lambda^2, \quad z^4 = \tilde{\omega}^4 + 4\gamma^2\lambda^2. \tag{2.6c}$$

Our present purpose is to solve (2.6a) and (2.6b) for I and J as fundamental invariants:

$$\begin{aligned}\omega I(q,\bar{p},t) = {} & (\omega\cos\omega t - \gamma\sin\omega t)qe^{\gamma t} - \omega_0\sin\omega t\bar{p}e^{-\gamma t} \\ & - (f_0/z^2)e^{\gamma t}[\omega\cos\omega t\cos(\lambda t + \phi - \phi') \\ & - (\gamma^2 + \lambda^2)^{1/2}\sin\omega t \\ & \times\cos(\lambda t - \phi - \phi' + \phi'')],\end{aligned} \tag{2.7a}$$

$$\begin{aligned}\omega J(q,\bar{p},t) = {} & (\omega\sin\omega t + \gamma\cos\omega t)qe^{\gamma t} + \omega_0\cos\omega t\bar{p}e^{-\gamma t} \\ & - (f_0/z^2)e^{\gamma t}[\omega\sin\omega t\cos(\lambda t + \phi - \phi') \\ & + (\gamma^2 + \lambda^2)^{1/2}\cos\omega t \\ & \times\cos(\lambda t - \phi - \phi' + \phi'')],\end{aligned} \tag{2.7b}$$

where the phase shift ϕ'' is given by

$$\phi'' = \arctan(\lambda/\gamma). \tag{2.7c}$$

A complex invariant $C(q,\bar{p},t)$ is given by

$$\begin{aligned}\omega C &= \omega(I + iJ) \\ &= e^{i\omega t}\{(\omega + i\gamma)qe^{\gamma t} + i\omega_0\bar{p}e^{-\gamma t} \\ &\quad - (f_0/z^2)e^{\gamma t}[\omega\cos(\lambda t + \phi - \phi') \\ &\quad + i(\gamma^2 + \lambda^2)^{1/2}\cos(\lambda t + \phi - \phi' + \phi'')]\}.\end{aligned} \tag{2.7d}$$

A second-degree energylike invariant is

$$\begin{aligned}\tfrac{1}{2}m_0\omega^2|C|^2 = {} & \tfrac{1}{2}m_0\omega_0^2 q^2 e^{2\gamma t} + \tfrac{1}{2}(p^2/m_0)e^{-2\gamma t} + \gamma qp - (f_0/z^2)e^{\gamma t}\{m_0[\omega^2\cos(\lambda t + \phi - \phi') \\ & + \gamma(\gamma^2 + \lambda^2)^{1/2}\cos(\lambda t + \phi - \phi' + \phi'')]qe^{\gamma t} + (\lambda^2 + \gamma^2)^{1/2}\cos(\lambda t + \phi - \phi' + \phi'')pe^{-\gamma t}\} \\ & + \tfrac{1}{2}m_0(f_0/z^2)^2 e^{2\gamma t}\{\omega^2\cos^2(\lambda t + \phi - \phi') + (\gamma^2 + \lambda^2)\cos^2(\lambda t + \phi - \phi' + \phi'')\},\end{aligned} \tag{2.8}$$

which reduces to the invariant K (Refs. 4 and 33) when we put $f_0 = 0$. An interesting special case is obtained when $\gamma \ll \omega_0$, $\lambda \gg \omega_0$ ($\gamma\lambda \approx \omega_0^2$). Then $\phi'' \approx \pi/2$ and, if we choose $t = 0$ so that $\phi = \phi'$, we find that

$$\begin{aligned}&\tfrac{1}{2}m_0\omega_0^2 q^2 e^{2\gamma t} + \tfrac{1}{2}(p^2/m_0)e^{-2\gamma t} + \gamma qp + (f_0/\lambda)p\sin\lambda t \\ &\quad \approx \tfrac{1}{2}m_0\omega_0^2 q^2(0) + \tfrac{1}{2}p^2(0)/m_0 + \gamma q(0)p(0) \\ &\qquad - \tfrac{1}{2}m_0(f_0/\lambda)^2 e^{2\gamma t}\sin^2\lambda t.\end{aligned} \tag{2.9}$$

The quantum theory of the damped driven oscillator is discussed by Um *et al.*[32]

Bohlin's integrals: Gettys *et al.*[9] write (2.3) with $f(t) = 0$ in the form

$$\ddot{q} + (\lambda_1 + \lambda_2)\dot{q} + \lambda_1\lambda_2 q = 0 \tag{2.10}$$

and discuss the invariants

$$D_1 = e^{\lambda_2 t}(\dot{q} + \lambda_1 q), \quad D_2 = e^{\lambda_1 t}(\dot{q} + \lambda_2 q), \tag{2.11a}$$

$$F = D_1 D_2, \tag{2.11b}$$

$$B = D_1^{\lambda_1}/D_2^{\lambda_2}. \tag{2.11c}$$

Here B is Bohlin's integral,[34] rediscovered by Gordon,[35] and D_1 and D_2 are discussed from a group-theoretic point of view by Cervero and Villarroel.[28] When $\gamma < \omega_0$, λ_1 and λ_2 are complex, so that D_1, D_2, and B are complex invariants. Obviously D_1, D_2, and F can be identified (up to constant multipliers) with the invariants C, C^*, and $|C|^2$ given by (2.7d) and (2.8) with $f_0 = 0$. When $f_0 \neq 0$ the complex invariants C and C^* given by (2.7d) are generalizations of D_1 and D_2.

We return to the damped harmonic oscillator in Sec. V.

III. COMPLEX INVARIANTS FOR THE *N*-DIMENSIONAL SHO

The invariants (2.7a)–(2.7c) reduce to I, J, and C given by (1.3a)–(1.3c) when $\gamma = 0$ and $f_0 = 0$. Here $\gamma \neq 0$ can be retrieved by making a simple canonical transformation. All the invariants found in Ref. 10 can be expressed in terms of I and J (or in terms of C and C^*). For instance, the invariants given by Eqs. (28d) and (28g) of Ref. 10 may be written, respectively, as

$$\begin{aligned}I_3^A(\phi) &= (q^2 + \bar{p}^2)[q\sin(\omega_0 t + \phi) + \bar{p}\cos(\omega_0 t + \phi)] \\ &= |C|^2 I(q,\bar{p},t + (\phi - \tfrac{1}{2}\pi)/\omega_0),\end{aligned} \tag{3.1a}$$

$$I_4^B(\phi) = \mathrm{Re}\, C^4(q,\bar{p},t + \tfrac{1}{4}(\phi - \tfrac{1}{2}\pi)/\omega_0). \tag{3.1b}$$

We employ classical notation for simplicity. To translate to

quantum mechanics it is necessary to symmetrize all products of q and $\bar{p}$.[10,36]

For extension to the N-dimensional SHO it is convenient to adopt the following notation for the one-dimensional SHO ($n = 1,2,...$)

$$C(n,\phi_n) = (q + i\bar{p})^n \exp[i(n\omega_0 t + \phi_n)], \tag{3.2a}$$

$$I_n(q,\bar{p},t;\phi_n) = \mathrm{Re}[C(n,\phi_n)]. \tag{3.2b}$$

We note the relations

$$I_n(q,\bar{p},t;\phi_n + \pi/2) = -\mathrm{Im}[C(n,\phi_n)], \tag{3.3a}$$

$$I_n(q,\bar{p},t;\phi_n) = I_n(q,\bar{p},t;0)\cos\phi_n + I_n(q,\bar{p},t;\pi/2)\sin\phi_n. \tag{3.3b}$$

The invariants I and J given by (1.3a) and (1.3b) are, respectively, $I_1(q,\bar{p},t;0)$ and $-I_1(q,\bar{p},t;\pi/2)$, and are related by

$$J(q,\bar{p},t) = -I(q,\bar{p},t + \pi/2), \tag{3.4}$$

showing that there is essentially only *one* fundamental invariant for the SHO.

Equations (2.7a) and (2.7b) show that the symmetry (3.4) is preserved when $\gamma = 0$ and a driving force is present, but is lost when damping is included. However, under the canonical transformation $Q = qe^{\gamma t}$, $\bar{P} = \bar{p}e^{-\gamma t}$ (considered in Sec. V) (3.4) is regained provided $f_0 = 0$.

We turn our attention now to the N-dimensional isotropic SHO represented by the Hamiltonian

$$H = \frac{1}{2} m_0\omega_0^2 \sum_{k=1}^{N} (q_k^2 + \bar{p}_k^2), \quad \bar{p}_k = \frac{p_k}{(m_0\omega_0)}. \tag{3.5}$$

The linear invariants are simply the sums of one-dimensional SHO invariants:

$$J_1(q_1,...,q_N,\bar{p}_1,...,\bar{p}_N,t) = \sum_{k=1}^{N} I_1^{(k)}(q_k,\bar{p}_k,t). \tag{3.6}$$

Mixed terms enter with the quadratic invariants. For $N \geqslant 2$,

$$J_2(q_1,...,q_N,\bar{p}_1,...,\bar{p}_N,t) = \sum_{k=1}^{N} I_2^{(k)}(q_k,\bar{p}_k,t) + \sum_{k,l(>k)=1}^{N} I_2^{(kl)}(q_k,q_l,\bar{p}_k,\bar{p}_l,t), \tag{3.7}$$

where the $I_2^{(k)}$ are linear combinations of $I_2^A(q_k,\bar{p}_k,t)$ given by (Eq. 28b) of Ref. 10 with a suitable shift in phase and $I_2^B(q_k,\bar{p}_k) = q_k^2 + \bar{p}_k^2$, just as for the one-dimensional SHO [Ref. 10, Eqs. (7a)–(7c)]. The $I_2^{(kl)}$ consist of second-degree terms constructed from the complex invariants

$$C_r(1,0) = (q_r + i\bar{p}_r)\exp(i\omega_0 t) \quad (r = k,l), \tag{3.8}$$

e.g., for $k = 1$ and $l = 2$, $I_2^{(12)}$ is a linear combination of the real and imaginary parts of C_1C_2 and $C_1^*C_2$. However, as in (3.2b) and (3.3a), the real and imaginary parts differ only in phase. Thus C_1C_2 gives the explicitly time-dependent first integral

$$(q_1q_2 - \bar{p}_1\bar{p}_2)\cos(2\omega_0 t + \phi) - (q_1\bar{p}_2 + q_2\bar{p}_1)\sin(2\omega_0 t + \phi), \tag{3.9a}$$

while the real and imaginary parts of $C_1^*C_2$ give the time-independent integrals

$$A_{12} = q_1q_2 + \bar{p}_1\bar{p}_2, \tag{3.9b}$$

$$L_{12} = q_1\bar{p}_2 - q_2\bar{p}_1. \tag{3.9c}$$

Here A_{12} is a Fradkin invariant[12,13] and L_{12} is an angular momentum component. The energy constants

$$A_{kk} = q_k^2 + \bar{p}_k^2 \quad (k = 1,2), \tag{3.9d}$$

from the one-dimensional parts of (3.7) may be added for completeness. From (3.9b)–(3.9d) we find Eq. (16) of Fradkin,[12]

$$A_{12}^2 + L_{12}^2 = A_{11}A_{22}. \tag{3.10a}$$

With $[q_k,\bar{p}_l] = i\hbar(m_0\omega_0)^{-1}\delta_{kl}$ and $E_k = \frac{1}{2}m_0\omega_0^2 A_{kk}$, the connection between A_{12} and L_{12} in quantum mechanics is

$$A_{12}^2 + L_{12}^2 = (m_0\omega_0^2)^{-2}(4E_1E_2 - \hbar^2\omega_0^2). \tag{3.10b}$$

The third-degree invariants for $N \geqslant 2$ are given by (3.7) with $n = 2$ replaced by $n = 3$. For $N \geqslant 3$ mixing occurs between three degrees of freedom,

$$J_3(q_1,...,q_N,\bar{p}_1,...,\bar{p}_N,t) = \sum_{k=1}^{N} I_3^{(k)}(q_k,\bar{p}_k,t) + \sum_{k,l(>k)=1}^{N} I_3^{(kl)}(q_k,q_l,\bar{p}_k,\bar{p}_l,t) + \sum_{k,l(>k),m(>l)=1}^{N} I_3^{(klm)}(q_k,q_l,q_m,\bar{p}_k,\bar{p}_l,\bar{p}_m,t). \tag{3.11}$$

Similarly we may write the nth-degree invariants in the form

$$J_n = \sum_{k=1}^{N} I_n^{(k)} + \sum_{k,l(>k)=1}^{N} I_n^{(kl)} + \cdots + \sum_{k,l(>k),...,s(>r)=1}^{N} I_n^{(kl,...,s)}, \tag{3.12}$$

where n_s, the number of degrees of freedom mixed in $I_n^{(kl,...,s)}$, is given by

$$n_s = \min(n,N), \quad n = 1,2,...,\infty. \tag{3.13}$$

The nth-degree invariant $I_n^{(1,2,...,n_s)}$ is constructed from all possible products of n complex invariants selected from the set $\{C_1,C_2,...,C_N,C_1^*,C_2^*,...,C_N^*\}$, making sure that none of $k = 1,2,...,n_s$ is omitted when selecting either C_k or C_k^*. Explicitly time-dependent invariants result from the selection of unequal numbers of C's and C^*'s, and time-independent integrals (or *constants of the motion*) are obtained from equal numbers of C's and C^*'s. For instance with $N = 3$ and $n = 6$ we could select for a part of $I_6^{(123)}$ the product of $C_1^2(C_2^*)^3C_3$. The real and imaginary parts of this product give two functionally related constants of the motion. However, since the selected product may be factorized into

$$(C_1C_2^*)^2(C_2^*C_3) = (A_{12} - iL_{12})^2(A_{23} + iL_{23}), \tag{3.14}$$

each invariant is simply a combination of L_{kl} and A_{kl}, found already from the two-dimensional SHO. A decomposition of the form (3.14) occurs with any other selection for $n \geqslant 3$, but possibly also including an energy $C_kC_k^*$ or an unpaired C or C^*. Thus no constants of the motion can be of odd degree and no new constants can arise for $n \geqslant 3$. We reach the conclusion that there are $\frac{1}{2}N(N+1)$ constants $A_{kl} = A_{lk}$ and $\frac{1}{2}N(N-1)$ constants $L_{kl} = -L_{lk}$ connected by $\frac{1}{2}N(N-1)$ relations of the form (3.10). This is in agreement with Refs. 12–14. Furthermore, there are no other independent con-

stants. Thus for the N-dimensional isotropic SHO there appear to be $\frac{1}{2}N(N+1)$ independent constants of the motion. However, when $N \geqslant 4$, only $2N$ of the constants can be determined independently from the initial coordinates and momenta. Hence the number of independent constants is $2N$ except for $N=1$ or 2 when it is one or three.

IV. THE ANISOTROPIC OSCILLATOR

The two-dimensional anisotropic oscillator has Hamiltonian

$$H = \tfrac{1}{2}p_1^2/m_1 + \tfrac{1}{2}m_1\omega_1^2 q_1^2 + \tfrac{1}{2}p_2^2/m_2 + \tfrac{1}{2}m_2\omega_2^2 q_2^2. \tag{4.1}$$

Let us write $\omega_1 = \lambda\omega_0$, $\omega_2 = \omega_0$, $m_1 = \mu^2 m_0$, $m_2 = m_0$. Then with

$$\bar{p}_1 = p_1(\lambda\mu^2 m_0\omega_0)^{-1}, \quad \bar{p}_2 = p_2(m_0\omega_0)^{-1},$$
$$H = \tfrac{1}{2}m_0\omega_0^2\left[(\lambda\mu)^2(q_1^2+\bar{p}_1^2) + (q_2^2+\bar{p}_2^2)\right]. \tag{4.2}$$

Building blocks for the construction of complex invariants are

$$C_1(1,\phi) = (q_1 + i\bar{p}_1)\exp[i(\lambda\mu\omega_0 t + \phi)], \tag{4.3a}$$
$$C_2(1,\psi) = (q_2 + i\bar{p}_2)\exp[i(\omega_0 t + \psi)]. \tag{4.3b}$$

The energy invariants are

$$E_1 = \tfrac{1}{2}m_0(\lambda\mu\omega_0)^2 C_1 C_1^*, \tag{4.4a}$$
$$E_2 = \tfrac{1}{2}m_0\omega_0^2 C_2 C_2^*. \tag{4.4b}$$

If $\lambda\mu$ is rational further time-independent invariants exist. The case $\lambda\mu = 2$ is important in the vibrational motion of molecules such as CO_2.[37] In addition to E_1 and E_2, constants of the motion (symmetrized to apply in quantum mechanics)

$$K = q_1(q_2^2 - \bar{p}_2^2) + \bar{p}_1(q_2\bar{p}_2 + \bar{p}_2 q_2), \tag{4.5a}$$
$$L = \bar{p}_1(q_2^2 - \bar{p}_2^2) - q_1(q_2\bar{p}_2 + \bar{p}_2 q_2), \tag{4.5b}$$

are found from the real and imaginary parts of $C_1(0)[C_2^*(0)]^2$. They are connected by the relation [cf. (3.10b)]

$$K^2 + L^2 = \tfrac{1}{2}(m_0\omega_0^2)^{-3}\left[4E_1E_2^2 + \hbar^2\omega_0^2(3E_1 - 8E_2)\right], \tag{4.6}$$

where we have used $[q_2,\bar{p}_2] = i\hbar(m_0\omega_0)^{-1}$. No further constants of the motion exist.

The N-dimensional anisotropic oscillator is described by the Hamiltonian

$$H = \sum_{k=1}^{N}\left(\frac{1}{2}\frac{p_k^2}{m_k} + \frac{1}{2}m_k\omega_k^2 q_k^2\right). \tag{4.7}$$

We put $\omega_k = \lambda_k\omega_0$, $m_k = \mu_k^2 m_0$, $N_k = \lambda_k\mu_k$, $\bar{p}_k = p_k(N_k\mu_k m_0\omega_0)^{-1}$. Then (4.7) becomes

$$H = \frac{1}{2}m_0\omega_0^2\sum_{k=1}^{N}N_k^2(q_k^2+\bar{p}_k^2). \tag{4.8}$$

All possible integrals of the motion are constructed from the complex invariants

$$C_k = (q_k + i\bar{p}_k)\exp(iN_k\omega_0 t) \quad (k = 1,2,\ldots,N). \tag{4.9}$$

If the values of $m_k\omega_k^2$ are commensurate, then $N_k \in \mathbb{Z}$ and (4.9) can lead to strict constants of the motion.

Let us consider two examples.

(a) $N=3$: $N_1=3$, $N_2=N_3=1$. Besides $C_kC_k^*$ ($k=1,2,3$), constants of the motion appear in the real parts of $C_1(C_2^*)^3$, $C_1(C_2^*)^2C_3^*$, $C_1C_2^*(C_3^*)^2$, and $C_1(C_3^*)^3$. The imaginary parts give further constants that are not independent. No further independent constants exist.

(b) $N=3$: $N_1=3$, $N_2=2$, $N_1=1$. Again, all possible constants of the motion, in addition to the energies, are contained in the real parts of $C_1C_2^*C_3^*$, $C_1(C_3^*)^3$, and $C_1^2(C_2^*)^3$.

With general sets of integers $N_1,N_2,\ldots,N_n$ the construction of constants of the motion becomes a complicated combinatorial problem which we do not consider to be of sufficient interest to pursue here.

V. COMPLEX INVARIANTS FOR THE TIME-DEPENDENT OSCILLATOR

The general TDO may be canonically transformed to the SHO (Refs. 1 and 8) so that our present method produces all possible invariants. As we remarked in the Introduction it is, in general, impossible to express these invariants explicitly and in this section we restrict our discussion to variable-mass oscillators for which explicit transformations are available.

For simplicity we set the driving term in (2.1a) equal to zero, when we are left with

$$H(q,p,t) = \tfrac{1}{2}p^2/m(t) + \tfrac{1}{2}m(t)\omega_0^2 q^2, \tag{5.1}$$

which may be transformed[15,33] via

$$Q = [m(t)/m_0]^{1/2}q, \quad P = [m(t)/m_0]^{-1/2}p, \tag{5.2}$$

to the new canonical Hamiltonian

$$K(Q,P,t) = \tfrac{1}{2}p^2/m_0 + \tfrac{1}{2}m_0\omega_0^2 Q^2 + \tfrac{1}{4}(\dot{m}/m)(QP+PQ). \tag{5.3}$$

Leach[1] lists several masses $m(t)$ for which explicit transformations can be found to reduce (5.1) or (5.3) to the SHO Hamiltonian (see also Wille and Vennik[38]). We shall consider the damped harmonic oscillator discussed in Ref. 4 (and in Sec. II of the present paper) and the strongly pulsating oscillator.

A. Invariants for the damped harmonic oscillator

Damping in the harmonic oscillator may be represented by an exponentially growing mass as in (2.1b) with $\gamma > 0$. Colegrave and Abdalla[3] argued that for application to a decaying field in a Fabry–Perot cavity γ should be taken negative and it should be noted that this led to $\gamma < 0$ being taken in Ref. 4.

With $\bar{P} = P/(m_0\omega_0)$, (5.3) gives the Hamiltonian

$$K = \tfrac{1}{2}m_0\omega\left[(Q_0^2 + \Gamma\bar{P})^2 + (1-\Gamma^2)\bar{P}^2\right], \quad \Gamma = \gamma/\omega_0. \tag{5.4}$$

Making the canonical transformation

$$q_0 = Q + \Gamma\bar{P}, \quad p_0 = P = m_0\omega_0\bar{P} \quad ([q_0,p_0] = [Q,P]), \tag{5.5}$$

the Hamiltonian transforms according to

$$K(Q,P) \to K(q_0,p_0) = \tfrac{1}{2}(1-\Gamma^2)p_0^2/m_0 + \tfrac{1}{2}m_0^2\omega_0^2 q_0. \tag{5.6}$$

We write

$$\omega = (\omega_0^2 - \gamma^2)^{1/2} = (1 - \Gamma^2)^{1/2}\omega_0, \tag{5.7}$$

then with the further canonical transformation

$$q_1 = (\omega_0/\omega)q_0, \quad p_1 = (\omega/\omega_0)p_0 \quad ([q_1,p_1] = [q_0,p_0]) \tag{5.8}$$

the reduction is complete:

$$K(q_0,p_0) \to \overline{K}(q_1,p_1) = \tfrac{1}{2}p_1^2/m_0 + \tfrac{1}{2}m_0\omega^2 q_1^2. \tag{5.9}$$

The classical Poisson bracket, or alternatively the quantum-mechanical commutator, can be used to keep track of canonicity.

From (5.9) the complex invariant is

$$C(q_1,p_1,t) = [q_1 + ip_1/(m_0\omega)]\exp(i\omega t). \tag{5.10}$$

The linear invariant

$$\text{Re } C = (\omega_0/\omega)[Q\cos\omega t - (\gamma\cos\omega t + \omega\sin\omega t)P/(m_0\omega_0^2)] \tag{5.11}$$

can easily be transformed to (2.7) with $f_0 = 0$ and agrees with Ref. 4 [Eq. (2.10)] when we change the sign of γ. Higher-degree invariants may be constructed according to (3.2b). In one dimension the only constant of the motion is (classically)

$$\tfrac{1}{2}m_0\omega^2 CC^* = \tfrac{1}{2}p_1^2/m_0 + \tfrac{1}{2}m_0\omega^2 q_1^2 = K, \tag{5.12a}$$

or quantum mechanically

$$\tfrac{1}{4}m_0\omega^2(CC^\dagger + C^\dagger C) = K. \tag{5.12b}$$

A three-dimensional isotropic oscillator with damping coefficient γ can be similarly transformed to a three-dimensional SHO with frequency $\omega = (\omega_0^2 - \gamma^2)^{1/2}$ and with the following six constants of the motion, forming the Fradkin tensor:

$$A_{kl} = q_1^k q_1^l + p_1^k p_1^l/(m_0\omega)^2 \quad (k,l = 1,2,3). \tag{5.13}$$

Alternatively, the three angular momentum constants

$$L_{kl} = q_1^k p_1^l - q_1^l p_1^k \tag{5.14}$$

may be used to replace the Fradkin invariants A_{kl} $(k \neq l)$.

B. Invariants for the strongly pulsating oscillator

We turn our attention to the oscillator with mass

$$m(t) = m_0\cos^2\nu t, \tag{5.15}$$

for which (5.2) and (5.3) give[15]

$$K(Q,P,t) = \tfrac{1}{2}P^2/m_0 + \tfrac{1}{2}m_0\omega_0^2 Q^2 - \tfrac{1}{2}\nu\tan\nu t(QP + PQ). \tag{5.16}$$

Introducing the dimensionless coordinate and momentum

$$q_0 = (m_0\omega_0/\hbar)^{1/2}Q, \quad p_0 = (m_0\omega_0\hbar)^{-1/2}P, \tag{5.17}$$

followed by the canonical transformation

$$\begin{pmatrix} q_1 \\ p_1 \end{pmatrix} = (\Omega\omega_0)^{-1/2}\begin{pmatrix} \Omega & 0 \\ -\nu\tan\nu t & \omega_0 \end{pmatrix}\begin{pmatrix} q_0 \\ p_0 \end{pmatrix}, \quad \Omega^2 = \omega_0^2 + \nu^2, \tag{5.18}$$

we find that the Hamiltonian

$$K(Q,P,t) \to \overline{K}(q_1,p_1) = \tfrac{1}{2}\hbar\Omega(q_1^2 + p_1^2). \tag{5.19}$$

Hence the complex invariant is

$$C(q_1,p_1,t) = (q_1 + ip_1)\exp(i\Omega t) = (m_0/\hbar\Omega)^{1/2}[(\Omega - i\nu\tan\nu t)Q + iP/m_0]\exp(i\Omega t), \tag{5.20}$$

cf. the Dirac operator $A_1(t)$ given in Ref. 15 by Eq. (6.4a), but note that $A_1(t)$ lacks the factor $\exp(i\Omega t)$ in (5.20). The Hermitian part of $C\exp(i\phi)$ gives the linear invariant

$$[\Omega\cos(\Omega t + \phi) + \nu\tan\nu t\sin(\Omega t + \phi)]Q - \sin(\Omega t + \phi)P/m_0. \tag{5.21}$$

Similarly the Hermitian parts of $C^n\exp(i\phi_n)$ $(n = 2,3,\ldots)$ give all the higher-degree invariants, except for the special integral of the motion

$$\tfrac{1}{4}\hbar\Omega(CC^\dagger + C^\dagger C) = K + \tfrac{1}{2}m_0\nu^2\sec^2\nu t Q^2, \tag{5.22a}$$

which may be expressed in the form

$$K(t) = K(0) - \tfrac{1}{2}m_0\nu^2[Q^2(0) - \sec^2\nu t Q^2(t)], \tag{5.22b}$$

cf. Ref. 16 [Eq. (55)]. We have used quantum-mechanical notation in (5.22a) since $\hbar$ has been introduced. However, this could have been avoided and (5.22b) is equally true in classical mechanics.

A three-dimensional extension of the Hamiltonian (5.1) with mass given by (5.15) may be transformed into the sum of three Hamiltonians of the form (5.19). Besides the hierarchy of invariants of type (5.21) in each dimension, there are six independent integrals of the motion given by the Fradkin tensor

$$A_{kl} = q_1^k q_1^l + p_1^k p_1^l = (\hbar\Omega)^{-1}[m_0(\Omega^2 + \nu^2\tan^2\nu t)Q_k Q_l + P_k P_l/m_0 - \nu\tan\nu t(Q_k P_l + P_l Q_k)] \quad (kl = 1,2,3), \tag{5.23a}$$

where (in quantum mechanics)

$$[Q_k,Q_l] = [P_k,P_l] = 0, \quad [Q_k,P_l] = i\hbar\delta_{kl}. \tag{5.23b}$$

An equation of the form (5.22b) holds in each dimension. Again, the angular momentum integrals of the form (5.15) could be used to replace the off-diagonal A_{kl}.

VI. INVARIANTS FOR THE VARIABLE-FREQUENCY AND VARIABLE-MASS OSCILLATORS

We illustrate the procedure of canonical transformation for the two TDO systems discussed in Ref. 4: the variable-frequency oscillator and the general variable-mass oscillator. Any time-dependent oscillator can be easily expressed as one or the other of these systems. The variable-mass Hamiltonian may be transformed to a variable-frequency form as discussed by Abdalla.[2]

A. The variable-frequency oscillator

The system that originated in the slowly lengthening pendulum[8] and led to the Lewis invariant[7] is represented by the Hamiltonian

$$H(q,p,t) = \tfrac{1}{2}p^2/m_0 + \tfrac{1}{2}m_0\omega^2(t)q^2. \tag{6.1}$$

We make the canonical transformation (cf. Ref. 13)

$$q_1 = m_0^{1/2} q/\rho, \quad p_1 = m_0^{-1/2}\rho p - m_0^{1/2}\dot{\rho} q, \tag{6.2}$$

where $\rho(t)$ is a solution of the auxiliary equation (1.4). Then with the generating function

$$F_3(q_1,p,t) = -m_0^{-1/2}\rho q_1 p + \tfrac{1}{2}\rho\dot{\rho} q_1^2, \tag{6.3}$$

the Hamiltonian (6.1) transforms to

$$K(q_1,p_1,t) = H + \frac{\partial F_3}{\partial t} = \frac{1}{2}\rho^{-2}(q_1^2 + p_1^2), \tag{6.4}$$

giving the equations of motion

$$\dot{q}_1 = p_1/\rho^2, \quad \dot{p}_1 = -q_1/\rho^2. \tag{6.5}$$

All that is necessary now is a change of time scale[1,8] $t \to t_1(t)$, where

$$t_1 = \int_0^t \rho^{-2}(t')dt', \tag{6.6a}$$

which causes

$$K(q_1,p_1,t) \to \bar{K}(q_1,p_1,t_1) = \tfrac{1}{2}(q_1^2 + p_1^2). \tag{6.6b}$$

Denoting differentiation with respect to t_1 by a prime, the equations of motion (6.5) become

$$q_1' = p_1, \quad p_1' = -q_1. \tag{6.7}$$

It follows that the complex invariant is

$$C(q_1,p_1,t_1) = (q_1 + ip_1)\exp[it_1(t)]. \tag{6.8}$$

The Lewis invariant is

$$\tfrac{1}{2}(CC^\dagger + C^\dagger C)/m_0 = (q_1^2 + p_1^2)/m_0 = (q/\rho)^2 + (\dot{\rho} q - \rho p/m_0)^2. \tag{6.9}$$

Additional fundamental invariants are [cf. (3.2b), (5.11), and (5.21)]

$$\tfrac{1}{2}[C^n \exp(i\phi_n) + (C^\dagger)^n \exp(-i\phi_n)], \quad n = 1,2,\ldots. \tag{6.10}$$

B. The variable-mass oscillator

The Hamiltonian for the variable-mass oscillator is given by (5.1), which may be transformed to (5.3):

$$H(q,p,t) = \tfrac{1}{2}p^2/m_0 + \tfrac{1}{2}m_0\omega_0^2 q^2 + \tfrac{1}{2}\epsilon(t)(qp + pq), \quad \epsilon = \tfrac{1}{2}\dot{m}/m. \tag{6.11}$$

In Ref. 4 [Eq. (2.22)] we have shown that there exists a Lewis-type invariant

$$I^{(q)} = q^2/\sigma^2 + [\sigma p/m_0 - (\dot{\sigma} - \epsilon\sigma)q]^2, \tag{6.12}$$

where $\sigma(t)$ satisfies

$$\ddot{\sigma} + (\omega_0^2 - \epsilon^2 - \dot{\epsilon})\sigma = 1/\sigma^3. \tag{6.13}$$

We make a canonical transformation analogous to (6.2),

$$q_1 = m_0^{1/2} q/\sigma, \quad p_1 = m_0^{-1/2}\sigma p - m_0^{1/2}(\dot{\sigma} - \epsilon\sigma)q. \tag{6.14}$$

Then with

$$F_3(q_1,p,t) = -m_0^{-1/2}\sigma q_1 p + \tfrac{1}{2}\sigma(\dot{\sigma} - \epsilon\sigma)q_1^2 \tag{6.15}$$

we find that

$$H(q,p,t) \to K(q_1,p_1,t) = \tfrac{1}{2}\sigma^{-2}(t)(q_1^2 + p_1^2). \tag{6.16}$$

Again we obtain the complex invariant (6.8) with

$$t_1 = \int_0^t \sigma^{-2}(t')dt'.$$

The Lewis invariant is

$$\tfrac{1}{2}(CC^\dagger + C^\dagger C)/m_0 = (q/\sigma)^2 + [(\dot{\sigma} - \epsilon\sigma)q - \sigma p/m_0]^2, \tag{6.17}$$

further invariants being given by (6.10). An equivalent treatment could be based on Ref. 4 [Eqs. (2.18) and (2.19)] (see also Moreira[39]).

C. *N*-dimensional time-dependent oscillators

Invariants for the N-dimensional variable-frequency and variable-mass oscillators may be constructed in the manner we have discussed in Secs. III and IV. In particular the components of the Fradkin tensor, as given in (5.13) and (5.23), are obtained as integrals of the motion (cf. Ref. 13, Sec. D, and Ref. 14).

VII. THE WEAKLY PULSATING DAMPED OSCILLATOR

A TDO of importance in quantum optics[5,6] is described by the Kanai–Caldirola Hamiltonian (5.1) with mass given by (1.1). The equation of motion is

$$\ddot{q} + (\dot{m}/m)\dot{q} + \omega_0^2 q = 0, \tag{7.1a}$$

with the fluctuation function

$$\tfrac{1}{2}(\dot{m}/m) = \gamma + \epsilon\nu\cos\nu t \quad (\epsilon \ll 1). \tag{7.1b}$$

A driving force could be added as in Sec. II, but for simplicity this is excluded. With $\epsilon \ll 1$ the damping acquires a slight modulation, corresponding to periodic supply and removal of energy. We must remember, however, that the energy of a TDO is not represented by the Hamiltonian.

A perturbative solution has been found for (7.1) (cf. Ref. 40) and this may be inverted as in Sec. II to give the two fundamental first-degree invariants to first (or higher) order in ϵ. A second-degree energylike invariant may be constructed from them which effectively replaces the constant energy for a system described by a time-independent Hamiltonian. Eigenvalues and eigenvectors of such an invariant are easily found and may be related to solutions of the Schrödinger equation.[7] The canonical transformation

$$Q = [m(t)/m_0]^{1/2} q = \exp(\gamma t + \epsilon\sin\nu t)q, \tag{7.2a}$$

$$P = [m(t)/m_0]^{-1/2} p = \exp(-\gamma t - \epsilon\sin\nu t)p \tag{7.2b}$$

takes the Hamiltonian (5.1) to the new form

$$K(Q,P,t) = \tfrac{1}{2}P^2/m_0 + \tfrac{1}{2}m_0\omega_0^2 Q^2 + \tfrac{1}{2}(\gamma + \epsilon\nu\cos\nu t)(QP + PQ) \tag{7.3}$$

(cf. Refs. 15 and 33). The equations of motion are now

$$\dot{Q} = (i\hbar)^{-1}[Q,K] = P/m_0 + (\gamma + \epsilon\nu\cos\nu t)Q, \tag{7.4a}$$

$$\dot{P} = (i\hbar)^{-1}[P,K] = -m_0\omega_0^2 Q - (\gamma + \epsilon\nu\cos\nu t)P. \tag{7.4b}$$

We introduce

$$\bar{P} = P/(m_0\omega_0), \quad I = Q(0), \quad J = \bar{P}(0), \quad \omega^2 = \omega_0^2 - \gamma^2. \tag{7.5}$$

Then the first-order solution of Eqs. (7.4) is (cf. Ref. 15) Eqs. (4a) and (4b) with $\gamma \to -\gamma$, and Ref. 1,

$$Q = aI + bJ, \quad \bar{P} = cI + dJ, \tag{7.6a}$$

$$a = a_0 + \epsilon a_1, \quad b = b_0 + \epsilon b_1, \quad c = c_0 + \epsilon c_1, \quad d = d_0 + \epsilon d_1, \tag{7.6b}$$

$$a_0 = \cos\omega t + (\gamma/\omega)\sin\omega t, \quad b_0 = (\omega_0/\omega)\sin\omega t, \quad c_0 = -(\omega_0/\omega)\sin\omega t, \quad d_0 = \cos\omega t - (\gamma/\omega)\sin\omega t, \tag{7.6c}$$

$$\begin{aligned}
a_1 &= (2/\omega)\cos(\nu t/2)(4\omega^2 - \nu^2)^{-1} \\
&\quad \times\{[\gamma(4\omega^2 - \nu^2)\sin\omega t - \omega(4\gamma^2 + \nu^2)\cos\omega t]\sin(\nu t/2) + 2\nu\omega_0^2 \sin\omega t\cos(\nu t/2)\}, \\
b_1 &= (2\omega_0/\omega)(4\omega^2 - \nu^2)^{-1}[\nu\sin(\nu t/2) - 2\gamma\cos(\nu t/2)]\,[2\omega\cos\omega t\sin(\nu t/2) - \nu\sin\omega t\cos(\nu t/2)], \\
c_1 &= (2\omega_0/\omega)(4\omega^2 - \nu^2)^{-1}[\nu\sin(\nu t/2) + 2\gamma\cos(\nu t/2)]\,[2\omega\cos\omega t\sin(\nu t/2) - \nu\sin\omega t\cos(\nu t/2)], \\
d_1 &= (2/\omega)\cos(\nu t/2)(4\omega^2 - \nu^2)^{-1} \\
&\quad \times\{[\gamma(4\omega^2 - \nu^2)\sin\omega t + \omega(4\gamma^2 + \nu^2)\cos\omega t]\sin(\nu t/2) - 2\nu\omega_0\sin\omega t\cos(\nu t/2)\}.
\end{aligned} \tag{7.6d}$$

Equations (7.6) represent a canonical transformation for $Q(0), \bar{P}(0)$ to $Q(t), \bar{P}(t)$ and consequently the determinant

$$\Delta \equiv ad - bc = 1, \tag{7.7}$$

which is easily verified, thus checking the solution to first order in ϵ. Inverting (7.6) we find the fundamental invariants

$$I(Q,\bar{P},t) = d(t)Q - b(t)\bar{P}, \quad J(Q,\bar{P},t) = -c(t)Q + a(t)\bar{P}. \tag{7.8}$$

To obtain an energylike invariant that reduces to K given by (5.4), or by Eq. (5.10b) of Ref. 4 with $\Gamma \to -\gamma$, when $\epsilon \to 0$, we take

$$I_2(Q,\bar{P},t) = I^2 + J^2 + (\gamma/\omega_0)(IJ + JI) = \bar{I}^2 + \bar{J}^2, \tag{7.9a}$$

$$\bar{I} = I + (\gamma/\omega_0)J, \quad \bar{J} = (\omega/\omega_0)J. \tag{7.9b}$$

From (7.6) and (7.9) we find

$$\begin{aligned}
I_2 &= [c^2 + d^2 - 2(\gamma/\omega_0)cd]Q^2 + [a^2 + b^2 - 2(\gamma/\omega_0)ab]\bar{P}^2 - [bd + ac - (\gamma/\omega_0)(ad + bc)](Q\bar{P} + \bar{P}Q) + O(\epsilon^2) \\
&= Q^2 + \bar{P}^2 + (\gamma/\omega_0)(Q\bar{P} + \bar{P}Q) + \epsilon L(Q,\bar{P},t) + O(\epsilon^2),
\end{aligned} \tag{7.10a}$$

$$\begin{aligned}
L(Q,\bar{P},t) &= 2[c_0c_1 + d_0d_1 - (\gamma/\omega_0)(c_0d_1 + c_1d_0)Q^2 + 2[a_0a_1 + b_0b_1 - (\gamma/\omega_0)(a_0b_1 + a_1b_0)]\bar{P}^2 \\
&\quad - \{b_0d_1 + b_1d_0 + a_0c_1 + a_1c_0 - (\gamma/\omega_0)[a_0d_1 + a_1d_0 + b_0c_1 + b_1c_0]\}(Q\bar{P} + \bar{P}Q).
\end{aligned} \tag{7.10b}$$

The result of substituting (7.6c) and (7.6d) into (7.10b) is too cumbersome to display here. However, in the undamped case $\gamma = 0$ it reduces to

$$\begin{aligned}
L &= [2\nu/(4\omega_0^2 - \nu^2)] \\
&\quad \times[(\nu\sin\nu t - 2\omega_0\sin 2\omega_0 t)(Q^2 - \bar{P}^2) \\
&\quad + 2\omega_0(\cos\nu t - \cos 2\omega_0 t)(Q\bar{P} + \bar{P}Q)].
\end{aligned} \tag{7.11}$$

The resonance case $\nu = 2\omega$ ($\gamma \neq 0$) or $\nu = 2\omega_0$ ($\gamma = 0$) leads to secular behavior which is, however, illusory from a physical standpoint.[41]

VIII. CONCLUSION

We have presented a new and unified treatment of the invariants for the SHO and hence for the general time-dependent oscillator. We have shown that the method can be applied when a driving force is present. An invariant complex combination of the solutions of the Hamilton or Heisenberg equations of motion permits the construction of all possible invariants for the harmonic oscillator in N dimensions. Two kinds of invariant may be distinguished for the isotropic oscillator: (a) those explicitly involving the time, formed from an unequal number of C's and C^*'s, like (3.1) or (3.9a); (b) those, like (3.9b)–(3.9d), that contain an equal number of C's and C^*'s. For the SHO these are constants of the motion and for the time-dependent oscillator they may be referred to as integrals of the motion. Examples of the latter are (5.23) and the Lewis invariants (6.9) and (6.17). In the case of the anisotropic oscillator the position is complicated by the physically artificial question of the commensurability of the numbers $m_k^{1/2}\omega_k$, i.e., of the N_k. In principle, however, there is no difficulty in the anisotropic case, as discussed in Sec. IV. Unless some of the N_k are commensurate strict integrals of the motion do not exist.

The invariants of type (a) consist of the homogeneous invariants $I_n(q_k, \bar{p}_k, t; \phi_n)$, $n = 1,2,\ldots,\infty$, as in (3.2), and their N-dimensional extensions, e.g., the two-dimensional extension (3.9a). The physical significance of these time-dependent invariants is somewhat obscure.

Type (b) invariants are necessarily of even degree and, because of factorizations of the form (3.14), those of degree greater than 2 are combinations of the $\frac{1}{2}N(N+1)$ fundamental second-degree Fradkin invariants A_{kl} associated in the TDO case with the diagonalized coordinates and momenta as in (5.23a) (cf. Ref. 14, conclusion of Sec. III). The diagonal elements A_{kk} in the case of the SHO are energy constants.

The damped harmonic oscillator, represented by a Kanai–Caldirola Hamiltonian with $m = m_0\exp(2\gamma t)$, may be reduced to a time-free equivalent SHO. This leads to a particularly simple complex invariant. The strongly pulsating oscillator is an example of a genuine time-dependent system for which the complex invariant may be expressed explicitly. The variable-frequency and general variable-mass

oscillators have been shown to have complex invariants involving an auxiliary function that satisfies an equation of the type (1.4).

ACKNOWLEDGMENTS

We are grateful to Dr. M. S. Abdalla for sending us an advance copy of his paper and to Ursula Ramjit for the solution given by Eqs. (7.6). It is a pleasure to acknowledge a most helpful communication from Professor P. G. L. Leach.

[1]P. G. L. Leach, J. Phys. A: Math. Gen. **16**, 3161 (1983).
[2]M. S. Abdalla, Phys. Rev. A **33**, 2870 (1986).
[3]R. K. Colegrave and M. S. Abdalla, Opt. Acta **28**, 495 (1981).
[4]R. K. Colegrave and M. S. Abdalla, J. Phys. A: Math. Gen. **16**, 3805 (1983).
[5]R. K. Colegrave and A. Vahabpour-Roudsari, Opt. Acta **32**, 485 (1985).
[6]R. K. Colegrave and A. Vahabpour-Roudsari, Opt. Acta **33**, 645 (1986).
[7]H. R. Lewis and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[8]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165 (1982).
[9]W. E. Gettys, J. R. Ray, and E. Breitenberger, Am. J. Phys. **49**, 162 (1981).
[10]R. K. Colegrave, M. S. Abdalla, and M. A. Mannan, J. Phys. A: Math. Gen. **17**, 1567 (1984).
[11]J. M. Jauch and E. H. Hill, Phys. Rev. **57**, 641 (1940).
[12]D. M. Fradkin, Am. J. Phys. **33**, 207 (1965).
[13]N. J. Günther and P. G. L. Leach, J. Math. Phys. **18**, 572 (1977).
[14]P. G. L. Leach, J. Math. Phys. **21**, 32 (1980).
[15]R. K. Colegrave and M. S. Abdalla, J. Phys. A: Math. Gen. **15**, 1549 (1982).
[16]M. S. Abdalla and R. K. Colegrave, Phys. Rev. A **32**, 1958 (1985).
[17]H. R. Lewis, Phys. Rev. Lett. **18**, 510, 636 (1967).
[18]H. R. Lewis, J. Math. Phys. **9**, 1976 (1968).
[19]H. R. Lewis, Phys. Rev. **172**, 1313 (1968).
[20]C. J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463 (1976).
[21]P. G. L. Leach, J. Math. Phys. **18**, 810 (1977).
[22]P. G. L. Leach, J. Math. Phys. **18**, 1902 (1977).
[23]P. G. L. Leach, SIAM J. Appl. Math. **34**, 496 (1978).
[24]P. G. L. Leach, J. Math. Phys. **19**, 446 (1978).
[25]P. G. L. Leach, J. Math. Phys. **21**, 300 (1980).
[26]M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[27]G. E. Prince and C. J. Eliezer, J. Phys. A: Math. Gen. **13**, 815 (1980).
[28]J. M. Cervero and J. Villaroel, J. Phys. A: Math. Gen. **17**, 2963 (1984).
[29]E. Kanai, Prog. Theor. Phys. **3**, 440 (1948).
[30]A. Tartaglia, Lett. Nuovo Cimento **19**, 205 (1977).
[31]P. Caldirola, Nuovo Cimento B **77**, 241 (1983).
[32]C. I. Um, K. H. Yeon, and W. H. Kahng, J. Phys. A: Math. Gen. **20**, 611 (1987).
[33]R. K. Colegrave and M. S. Abdalla, J. Phys. A: Math. Gen. **14**, 2269 (1981).
[34]K. Bohlin, Astron, Lakttag. Undersokn. Stockholms Observ. **9**, 3 (1908).
[35]T. J. Gordon, J. Phys. A: Math. Gen. **19**, 183 (1986).
[36]R. K. Colegrave and P. Bala, J. Phys. A: Math. Gen. **18**, 779 (1985).
[37]T. P. Grosdanov, Saini Subhash, and H. S. Taylor, J. Phys. A: Math. Gen. **19**, 691 (1986).
[38]L. T. Wille and J. Vennik, J. Phys. A: Math. Gen. **18**, 1985 (1985).
[39]I. C. Moreira, J. Phys. A: Math. Gen. **18**, 899 (1985).
[40]M. S. Abdalla,R. K. Colegrave, and A. Khosravi, Nuovo Cimento B **93**, 195 (1986).
[41]R. K. Colegrave and A. Khosravi, J. Sound Vibration **109**, 473 (1986).

On the quantization of constrained generalized dynamics

Carlos A. P. Galvão
Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud, 150, Urca, 22290, Rio de Janeiro, Rio de Janeiro, Brazil

Nivaldo A. Lemos
Instituto de Física, Universidade Federal Fluminense, 24210, Niterói, Rio de Janeiro, Brazil

(Received 19 November 1987; accepted for publication 2 March 1988)

A special class of degenerate second-order Lagrangians, those that differ from a nondegenerate first-order Lagrangian by a total time derivative (or a four-divergence) of a function of both the coordinates and velocities, is studied in detail. Using Dirac's theory of constrained systems, it is shown that the canonical quantization starting from the second-order Lagrangian leads to the same physical results as those obtained from the nondegenerate first-order Lagrangian. Then some incorrect results and misleading arguments encountered in the literature on the subject are clarified.

I. INTRODUCTION

In spite of the fact that most physical systems can be described by Lagrangians that depend at most on the first derivatives of the dynamical variables, there is a continuing interest in the so-called generalized dynamics, that is, the study of physical systems described by Lagrangians containing derivatives of order higher than the first.[1]

Besides its mathematical interest connected with general problems in the calculus of variations as first investigated by Ostrogradskii,[2] higher-order terms were used in the past as intended corrections to first-order Lagrangians associated with certain physical theories. The attempts were to generalize them or to get rid of bad properties of those theories. To the best of our knowledge the earliest attempts in this direction were those by Weyl and Eddington[3] who added curvature squared terms to the Einstein–Hilbert Lagrangian so as to extend the theory of general relativity. Modifications to Maxwell's electromagnetic theory have been put forward by Bopp[4] and Podolsky[5] with the goal of avoiding divergences such as the infinite self-energy of a point charge (which, in a certain sense, they succeeded in doing). Stimulated by those findings, Pais and Uhlenbeck[6] investigated whether the use of higher-order field equations might lead to the disappearance of the divergent quantities that plague quantum field theory. Their general conclusion was that it is impossible to reconcile finiteness, positivity of free field energy, and causality. In other words, ghost states with negative norm and possibly unitarity violation are inherent in those theories and these facts turned out to be serious arguments responsible for the bad reputation of higher-order theories.

However, higher-order Lagrangians are endowed with nice properties, too, and they have been the subject of recent interest for many reasons. It has been shown,[7] for instance, that curvature squared terms show up as small corrections in the effective action of superstring theories in the limit of zero slope. The same kind of corrections have been proposed in the quantum theory of gravitation to improve the ultraviolet behavior of the Einstein–Hilbert action,[8] as higher-order derivative terms are known to improve the convergence of Feynman diagrams. As a mechanism for regularizing the ultraviolet divergences of gauge invariant supersymmetric theories it is the only available method that preserves both gauge invariance and supersymmetry.[9] Just to mention one more example, higher-order Lagrangians come forth naturally when one looks for a Hamiltonian description of certain nonlinear physical systems like those described by the equations associated with the names of Boussinesq or Korteweg and de Vries.[10] Last but not least, the Lagrangian of one of the most outstanding physical theories of our times, the Einstein theory of relativity, does in fact contain second-order derivatives of the metric field. It is then clear that such theories deserve a deeper investigation.

On the other hand, the kinetic term of any first-order Lagrangian can be transformed into two terms, one of which is linear in the "accelerations" and the other is a divergence, thus generating a second-order Lagrangian. Both Lagrangians are naively expected to describe the very same physical system and should not lead to different results even at the quantum level. But in the process of passing to the second-order formulation one necessarily ends up with a degenerate (or singular) Lagrangian. It was precisely this fact that gave rise to some controversy in the literature concerning the quantization of higher-order mechanical systems.

Hayes and Jankowski[11] analyzed a second-order Lagrangian that generates the correct equation of motion for a harmonic oscillator but, they claimed, yields an energy spectrum different from the usual one upon quantization. Subsequently Hayes[12] proposed an unorthodox and peculiar quantization prescription to circumvent the difficulties encountered in his previous work. His quantization procedure was immediately criticized by Ryan[13] and Anderson,[14] who, surprisingly enough, put the blame on the Lagrangian chosen by Hayes because it was singular. As a matter of fact they entirely missed the point, for the fundamental shortcoming in Hayes's approach was his lack of recognition that he was dealing with a constrained dynamical system to which Dirac's formalism[15] must be applied. This was perceived by Tesser[16] and also by Cognola, Vanzo, and Zerbini.[17] However, having disregarded the need to substitute Dirac brackets for Poisson brackets before quantizing, Tesser did not apply in a fully transparent and systematic fashion Dirac's theory as developed for systems with second class constraints. This was partly done by Cognola *et al.*, who consid-

 0022-2488/88/071588-05$02.50

ered only a very particular class of one-dimensional Lagrangians. Therefore they did not investigate the drawbacks of Hayes's treatment in their complete generality, and it is also worth mentioning that their reasoning did not furnish the thoroughly reduced phase space as witnessed by the fact that their Hamiltonian and fundamental Dirac brackets retained a dependence on an arbitrary parameter. Furthermore, in a recent paper Tapia[18] does not employ correctly Dirac's formalism to constrained generalized mechanics and in his study of the quantum theory of the harmonic oscillator he arrives at an energy spectrum that again does not coincide with the usual one. His case is even worse than that of Hayes and Jankowski, because the spectrum he obtained is unbounded below, not to mention the misleading arguments that led to the above mentioned spectrum.

It is our purpose in the present paper to provide a general explanation why results such as those found by Hayes and Jankowski[11] or Tapia[18] for the harmonic oscillator are wrong, whereas those obtained by Barcelos-Neto and Braga[19] for the Klein–Gordon field are correct. It is also our aim to dismiss as unnecessary and groundless odd quantization procedures such as the one advanced by Hayes, and finally to characterize as misleading those arguments[13,14] to the effect that degenerate Lagrangians should be avoided in generalized dynamics.

The paper is organized as follows. In Sec. II we state and prove our main result for systems with a finite number of degrees of freedom, while its extension to field theory is briefly sketched in the Appendix. Section III is devoted to a few examples and a general conclusion.

II. THE MAIN RESULT

The generalization of Hamilton's least action principle and of the Hamiltonian formulation to nondegenerate Lagrangians depending on higher-order derivatives was first achieved by Ostrogradskii,[2] and a more modern presentation of the canonical formalism is available in Whittaker's classical treatise.[20] With an eye to physical applications, and for the sake of simplicity, we shall consider only the second-order case, although our reasoning may be extended in a very direct way to Lagrangians involving derivatives up to an arbitrarily high order.

Let $\bar{L}(x,\dot{x},\ddot{x},t)$ be a second-order Lagrangian, where we are using the notation $x=(x_1,...,x_N)$, $\dot{x}=(\dot{x}_1,...,\dot{x}_N)$, etc. The action principle

$$\delta S \equiv \delta \int_{t_1}^{t_2} \bar{L}(x,\dot{x},\ddot{x},t)dt = 0, \tag{2.1}$$

where the variation is performed under the condition that the end points remain fixed, leads to the equations of motion

$$\frac{\partial \bar{L}}{\partial x_i} - \frac{d}{dt}\left(\frac{\partial \bar{L}}{\partial \dot{x}_i}\right) + \frac{d^2}{dt^2}\left(\frac{\partial \bar{L}}{\partial \ddot{x}_i}\right) = 0. \tag{2.2}$$

For future convenience it will be useful to introduce the notation

$$y_i \equiv \dot{x}_i\ . \tag{2.3}$$

The canonical momenta are defined as

$$p_1^i = \frac{\partial \bar{L}}{\partial y_i} - \frac{d}{dt}\left(\frac{\partial \bar{L}}{\partial \dot{y}_i}\right) \tag{2.4}$$

and

$$p_2^i = \frac{\partial \bar{L}}{\partial \dot{y}_i}\ . \tag{2.5}$$

Equations (2.5) can be solved for the $\dot{y}_i$ if and only if the Hessian matrix $\bar{W}$, whose elements are

$$\bar{W}_{ij} = \frac{\partial^2 \bar{L}}{\partial \dot{y}_i\, \partial \dot{y}_j}, \tag{2.6}$$

is nonsingular. Assuming this is the situation, the Hamiltonian defined as (Einstein's summation convention over repeated indices is understood from now on)

$$\bar{H}(x,p_1;y,p_2;t) = y_i p_1^i + \dot{y}_i p_2^i - \bar{L}(x,y,\dot{y},t) \tag{2.7}$$

generates Hamilton's equations of motion

$$\dot{x}_i = \frac{\partial \bar{H}}{\partial p_1^i}, \quad \dot{p}_1^i = -\frac{\partial \bar{H}}{\partial x_i}, \tag{2.8a}$$

$$\dot{y}_i = \frac{\partial \bar{H}}{\partial p_2^i}, \quad \dot{p}_2^i = -\frac{\partial \bar{H}}{\partial y_i}. \tag{2.8b}$$

The pairs of canonically conjugate variables are (x,p_1) and (y,p_2), the Poisson brackets being defined as follows:

$$\{F,G\} = \frac{\partial F}{\partial x_i}\frac{\partial G}{\partial p_1^i} - \frac{\partial F}{\partial p_1^i}\frac{\partial G}{\partial x_i} + \frac{\partial F}{\partial y_i}\frac{\partial G}{\partial p_2^i} - \frac{\partial F}{\partial p_2^i}\frac{\partial G}{\partial y_i}. \tag{2.9}$$

In terms of these brackets the equation of motion of any dynamical variable F becomes simply

$$\frac{dF}{dt} = \{F,\bar{H}\} + \frac{\partial F}{\partial t}. \tag{2.10}$$

Suppose now $\bar{L}$ is of the form

$$\bar{L}(x,y,\dot{y},t) = L(x,y,t) + \frac{d}{dt}f(x,y), \tag{2.11}$$

where f stands for an arbitrary function and L is a nondegenerate first-order Lagrangian, that is, its Hessian matrix W, whose elements are

$$W_{ij} = \frac{\partial^2 L}{\partial y_i\, \partial y_j} \equiv \frac{\partial^2 L}{\partial \dot{x}_i\, \partial \dot{x}_j}, \tag{2.12}$$

is nonsingular. Obviously, $\bar{L}$ and L generate the same equations of motion. The explicit form of $\bar{L}$ is

$$\bar{L}(x,y,\dot{y},t) = L(x,y,t) + \frac{\partial f}{\partial x_i}y_i + \frac{\partial f}{\partial y_i}\dot{y}_i\ , \tag{2.13}$$

from which it follows immediately that $\bar{W}=0$ according to Eq. (2.6), so that $\bar{L}$ is singular. As a consequence, we are sure to meet relations of functional dependence among the canonical variables.

The canonical momenta are easily found to be

$$p_1^i = \frac{\partial \bar{L}}{\partial y_i} - \frac{d}{dt}\frac{\partial \bar{L}}{\partial \dot{y}_i} = \frac{\partial L}{\partial y_i} + \frac{\partial f}{\partial x_i}, \tag{2.14}$$

$$p_2^i = \frac{\partial \bar{L}}{\partial \dot{y}_i} = \frac{\partial f}{\partial y_i}. \tag{2.15}$$

Define the functions

$$g_i(x,y,t) = \frac{\partial L(x,y,t)}{\partial y_i} + \frac{\partial f(x,y)}{\partial x_i} \tag{2.16}$$

and

$$h_i(x,y) = \frac{\partial f(x,y)}{\partial y_i}. \tag{2.17}$$

Then we have the following primary constraints:

$$\phi_i = p_1^i - g_i(x,y,t) \approx 0, \tag{2.18}$$

$$\psi_i = p_2^i - h_i(x,y,t) \approx 0. \tag{2.19}$$

The Poisson brackets of the constraints are easily calculated from definition (2.9). They are

$$\{\phi_i,\phi_j\} = \frac{\partial g_j}{\partial x_i} - \frac{\partial g_i}{\partial x_j} = \frac{\partial^2 L}{\partial x_i\,\partial y_j} - \frac{\partial^2 L}{\partial x_j\,\partial y_i} \equiv Y_{ij}, \tag{2.20}$$

$$\{\phi_i,\psi_j\} = \frac{\partial h_j}{\partial x_i} - \frac{\partial g_i}{\partial y_j} = -\frac{\partial^2 L}{\partial y_i\,\partial y_j} = -W_{ij}, \tag{2.21}$$

$$\{\psi_i,\psi_j\} = \frac{\partial h_j}{\partial y_i} - \frac{\partial h_i}{\partial y_j} = 0. \tag{2.22}$$

It is convenient to define

$$(\chi_1,\ldots,\chi_{2N}) = (\phi_1,\ldots,\phi_N,\psi_1,\ldots,\psi_N) \tag{2.23}$$

and agree that lowercase Latin indices from the beginning of the alphabet always run from 1 to $2N$. The $2N \times 2N$ matrix built up with the Poisson brackets of the constraints is, therefore,

$$\mathbb{C} = \|\{\chi_a,\chi_b\}\| = \begin{bmatrix} Y & -W \\ W & 0 \end{bmatrix}, \tag{2.24}$$

where W and Y are the $N \times N$ matrices defined by Eqs. (2.12) and (2.20), respectively. It is readily shown that $\mathbb{C}$ is nonsingular. In fact,

$$\det\|\{\chi_a,\chi_b\}\| = (\det W)^2 \neq 0, \tag{2.25}$$

since L is nonsingular by hypothesis. Thus all of our constraints are of the second class, and no linear combination of the χ_a can become a first class constraint, so that the extended Hamiltonian is nothing but the usual one given by Eq. (2.7).

According to Dirac's fomalism,[15] one must replace the original Poisson brackets by the new Dirac brackets defined by

$$\{F,G\}^* = \{F,G\} - \{F,\chi_a\}\,(\mathbb{C}^{-1})^{ab}\,\{\chi_b,G\}, \tag{2.26}$$

where $\mathbb{C}^{-1}$ is the inverse of the matrix given by Eq. (2.24). It is an easy task to construct $\mathbb{C}^{-1}$ and check that it can be put in the form

$$\mathbb{C}^{-1} = \begin{bmatrix} 0 & W^{-1} \\ -W^{-1} & W^{-1}YW^{-1} \end{bmatrix}. \tag{2.27}$$

The Hamiltonian (2.7) may be written as

$$\bar{H} = y_i p_1^i + \dot{y}_i p_2^i - L - \frac{\partial f}{\partial x_i} y_i - \frac{\partial f}{\partial y_i}\dot{y}_i, \tag{2.28}$$

where use has been made of Eq. (2.13). Once we are working with Dirac brackets, we are allowed to regard the constraints (2.18) and (2.19) as strong equations. Therefore setting

$$p_1^i = \frac{\partial L}{\partial y_i} + \frac{\partial f}{\partial x_i}, \quad p_2^i = \frac{\partial f}{\partial y_i}, \tag{2.29}$$

and inserting these equations into Eq. (2.28), we are left with

$$\bar{H} = \dot{x}_i \frac{\partial L}{\partial \dot{x}_i} - L. \tag{2.30}$$

By defining

$$p_i = \frac{\partial L}{\partial \dot{x}_i}, \tag{2.31}$$

these equations can be uniquely solved for the velocities $\dot{x}_i$ and Eq. (2.30) ensures that

$$\bar{H} = H(x,p,t), \tag{2.32}$$

where H is the ordinary Hamiltonian corresponding to the nonsingular first-order Lagrangian L. Notice that the momenta p_1 and p_2 have been wholly removed from the theory. It remains to examine the fundamental Dirac brackets of the new presumably canonical pair (x,p). Performing a few straightforward computations one finds successively

$$\begin{aligned} \{x_i,p_j\}^* &= \left\{x_i, \frac{\partial L}{\partial y_j}\right\} - \{x_i,\chi_a\}\,(\mathbb{C}^{-1})^{ab}\left\{\chi_b, \frac{\partial L}{\partial y_j}\right\} \\ &= -\{x_i,\phi_k\}(\mathbb{C}^{-1})^{kb}\left\{\chi_b, \frac{\partial L}{\partial y_j}\right\} \\ &= -\delta_{ik}(\mathbb{C}^{-1})^{kb}\left\{\chi_b, \frac{\partial L}{\partial y_j}\right\} \\ &= -(W^{-1})^{il}\left\{\psi_l, \frac{\partial L}{\partial y_j}\right\}, \end{aligned} \tag{2.33}$$

where we have made use of the explicit form of $\mathbb{C}^{-1}$ given by Eq. (2.27). From Eqs. (2.9) and (2.19) it follows at once that

$$\left\{\psi_l, \frac{\partial L}{\partial y_j}\right\} = -\frac{\partial^2 L}{\partial y_l\,\partial y_j} = -W_{lj}, \tag{2.34}$$

whence, after its insertion into Eq. (2.33), one finally gets

$$\{x_i,p_j\}^* = \delta_{ij}. \tag{2.35}$$

By using a little more effort, a similar kind of calculation allows us to show that

$$\{x_i,x_j\}^* = \{p_i,p_j\}^* = 0. \tag{2.36}$$

This establishes unequivocally that at the level of Dirac brackets (x,p) indeed constitute a pair of canonically conjugate variables.

So long as the canonical quantization then proceeds in a standard manner by requiring that the fundamental commutators be made equal to $i\hbar$ times the corresponding fundamental Dirac brackets, we have proved that the correct way of quantizing the classical theory associated with the singular second-order Lagrangian $\bar{L}$ conduces to the same physical results as the quantum theory based upon the nonsingular first-order Lagrangian L.

III. EXAMPLES AND CONCLUSION

The Lagrangian introduced by Hayes and Jankowski[11,12] is

$$\bar{L} = -(m/2)x\ddot{x} - \tfrac{1}{2}kx^2, \tag{3.1}$$

which gives rise to the equation of motion of a harmonic oscillator:

$$m\ddot{x} + kx = 0. \tag{3.2}$$

The above Langrangian is clearly of the form (2.11) with

$$L = (m/2)\dot{x}^2 - \tfrac{1}{2}kx^2 \tag{3.3}$$

and

$$f(x,\dot{x}) = -(m/2)x\dot{x}\,. \tag{3.4}$$

Thus the correct quantum theory of the harmonic oscillator based on the second-order Lagrangian (3.1) is the same as the usual one whose starting point is the first-order Lagrangian (3.3). The ambiguities encountered in Ref. 11 are just a consequence of an inaccurate threatment of the constraints.

Tapia's approach[18] corresponds to taking $\bar{L}$ of the form (2.11) with L the same as the one given by the above Eq. (3.3) and $f = 0$. His mistake stems from an incomplete treatment of the second class constraints in Sec. 5 of his paper. To go ahead to the quantum theory it would have been indispensable first to introduce Dirac brackets, which he did not. In the Appendix he falls into another error because he treats the canonical variables as independent, paying no attention to the constraints. Finally, it is not true that his energy spectrum differs from the standard one merely by an additive constant. From his method of obtaining the energy eigenvalues $E_n = n\hbar\omega$, it is plain to see that n may be any negative or positive integer, so that his spectrum is unbounded *below*, and this is unacceptable on physical grounds.

The results obtained for the Klein–Gordon field by Barcelos-Neto and Braga[19] can also be easily explained as a particular example of our general result (see the Appendix). Their Lagrangian is

$$\bar{\mathscr{L}} = -\tfrac{1}{2}\phi\Box\phi + V(\phi), \tag{3.5}$$

which is of the form (A1) with

$$\mathscr{L} = \tfrac{1}{2}\partial_\mu\phi\,\partial^\mu\phi + V(\phi) \tag{3.6}$$

being the usual Klein–Gordon Lagrangian, and

$$\Omega_\mu = -\tfrac{1}{2}\phi\,\partial_\mu\phi. \tag{3.7}$$

Since they treat the constraints correctly,[19] it is no surprise that they recover the well-known quantum theory of the Klein–Gordon field.

As a conclusion, let us emphasize that degenerate Lagrangians are not to be rejected in generalized mechanics: all one has to do is apply correctly the formalism designed by Dirac to cope with constrained systems. In particular, Dirac's method can be applied to quantum gravity taking as the starting point the usual Einstein–Hilbert Lagrangian, it being apparent, then, that it is not necessary to change the standard action for the gravitational field through the addition of a surface term, as is commonly done in quantum cosmology.[21]

ACKNOWLEDGMENT

N.A.L was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, Brazil.

APPENDIX: EXTENSION TO FIELD THEORY

The extension of the results of Sec. II to field theory is quite straightforward and for this reason we shall only briefly sketch the Hamiltonian approach. We denote a set of fields on Minkowski space-time by $\psi(x) \equiv \{\psi^A(x)\}$, $A = 1,2,\ldots,N$, where $x \equiv \{x^\mu\} \equiv \{x^0, x^i\}$, $i = 1,2,3$, are the space-time coordinates. Derivatives of the fields will be denoted by $\partial\psi \equiv \{\partial_\mu\psi^A(x)\}$, $\partial^2\psi \equiv \{\partial_\mu\,\partial_\nu\,\psi^A(x)\}$, etc., and, in particular, $\partial_0\psi^A \equiv \dot{\psi}^A \equiv \phi^A$.

The analog of the Lagrangian (2.11) in field theory is

$$\bar{\mathscr{L}} = \mathscr{L}(\psi,\partial\psi) + \partial_\mu\Omega^\mu(\psi,\partial\psi)$$

$$= \mathscr{L}(\psi,\partial\psi) + \frac{\partial\Omega^\mu}{\partial\psi^A}\,\partial_\mu\psi^A + \frac{\partial\Omega^\mu}{\partial(\partial_\alpha\psi^A)}\,\partial_\mu\,\partial_\alpha\,\psi^A\,. \tag{A1}$$

We shall assume that the Hessian matrix associated with $\mathscr{L}$ is nonsingular:

$$W_{AB} = \frac{\partial^2\mathscr{L}}{\partial\dot{\psi}^A\,\partial\dot{\psi}^B} \equiv \frac{\partial^2\mathscr{L}}{\partial\phi^A\,\partial\phi^B}, \quad \det W \neq 0. \tag{A2}$$

It follows from (A1) that

$$\bar{W}_{AB} = \frac{\partial^2\bar{\mathscr{L}}}{\partial\dot{\phi}^A\,\partial\dot{\phi}^B} = 0,$$

so that $\bar{\mathscr{L}}$ is degenerate.

The canonical variables of the theory are $(\psi^A, \overset{(1)}{\Pi}_A)$ and $(\phi^A \equiv \dot{\psi}^A, \overset{(2)}{\Pi}_A)$, with the momenta defined[22] by

$$\overset{(1)}{\Pi}_A = \frac{\partial\bar{\mathscr{L}}}{\partial\phi^A} - 2\,\partial_k\left(\frac{\partial\bar{\mathscr{L}}}{\partial(\partial_0\,\partial_k\,\psi^A)}\right) - \partial_0\left(\frac{\partial\bar{\mathscr{L}}}{\partial\phi^A}\right),$$

$$\overset{(2)}{\Pi}_A = \frac{\partial\bar{\mathscr{L}}}{\partial\dot{\phi}^A}\,.$$

Using (A1) we obtain

$$\overset{(1)}{\Pi}_A = \frac{\partial\mathscr{L}}{\partial\phi^A} + \frac{\partial\Omega^0}{\partial\psi^A} - \partial_k\left(\frac{\partial\Omega^0}{\partial(\partial_k\psi^A)}\right), \tag{A3}$$

$$\overset{(2)}{\Pi}_A = \frac{\partial\Omega^0}{\partial\phi^A}\,. \tag{A4}$$

It is worth remarking that only $\Omega^0(\psi,\partial\psi)$ shows up in the above expressions. This is to be expected, as one can be easily convinced by looking at the action functional constructed with the Lagrangian (A1).

Expressions (A3) and (A4) give us the primary constraints

$$\eta_A = \overset{(1)}{\Pi}_A - F_A(\psi,\partial_k\psi,\phi,\partial_k\phi) \approx 0, \tag{A5}$$

$$\lambda_A = \overset{(2)}{\Pi}_A - G_A(\psi,\partial_k\psi,\phi,\partial_k\phi) \approx 0. \tag{A6}$$

We have, after a straightforward calculation,

$$\{\eta_A,\eta_B\} = \frac{\partial F_B}{\partial\psi^A} - \frac{\partial F_A}{\partial\psi^B} \equiv \xi_{AB}\,, \tag{A7a}$$

$$\{\eta_A,\lambda_B\} = -\frac{\partial^2\mathscr{L}}{\partial\phi^A\,\partial\phi^B} = -W_{AB}\,, \tag{A7b}$$

$$\{\lambda_A,\lambda_B\} = 0, \tag{A7c}$$

so that, as in the case of Sec. II, the constraints are second class. Since the matrix C constructed with the above Poisson brackets is nonsingular, using its inverse, $\mathbb{C}^{-1}$, the Dirac brackets are now defined as

$$\{A(x),B(z)\}^* = \{A(x),B(z)\} - \int dw\,dy\{A(x),\chi_a(w)\} \times (\mathbb{C}^{-1})^{ab}(w,y)\,\{\chi_b(y),B(z)\}, \quad \text{(A8)}$$

where $\chi_a \equiv (\eta_A,\lambda_A)$.

In analogy with Sec. II, the constraints can be regarded as strong equations, thus removing from the theory the momenta $\overset{(1)}{\Pi}_A$ and $\overset{(2)}{\Pi}_A$. Defining

$$\Pi_A = \frac{\partial \mathscr{L}}{\partial \dot{\psi}^A}, \quad \text{(A9)}$$

we find that the Hamiltonian $\overline{\mathscr{H}}$ reduces to the ordinary Hamiltonian $\mathscr{H}$ associated with the first-order Lagrangian $\mathscr{L}$, and it may be easily verified that (ψ,Π) constitute a canonical pair in terms of the Dirac brackets (A8).

[1]V. V. Nesterenko, "The singular Lagrangians with higher derivatives," JINR preprint, 1987; S. W. Hawking, "Who's afraid of (higher derivative) ghosts?," DAMTP preprint, 1985.
[2]M. Ostrogradskii, Mem. Acad. St. Pétersburg **6**, 385 (1850).
[3]See P. Havas, Gen. Relativ. Gravit. **8**, 631 (1977), and references therein.
[4]F. Bopp, Ann. Phys. (Leipzig) **38**, 345 (1940).
[5]B Podolsky, Phys. Rev. **62**, 68 (1942); B. Podolsky and C. Kikuchi, *ibid.* **65**, 228 (1944); **67**, 184 (1945); B. Podolsky and P. Schwed, Rev. Mod. Phys. **20**, 40 (1948). See also Carlos A. P. Galvão and B. M. Pimentel, "The canonical structure of Podolsky generalized electrodynamics," Can. J. Phys. (in press).
[6]A. Pais and G. Uhlenbeck, Phys. Rev. **79**, 145 (1950).
[7]D. G. Boulware and S. Deser, Phys. Rev. Lett. **55**, 2656 (1984).
[8]K. S. Stelle, Phys. Rev. D **16**, 953 (1977).
[9]P. West, Nucl. Phys. B **268**, 113 (1986); S. Thomas, "The β-function in higher covariant derivative regularization," King's College preprint, 1986.
[10]Y. Nutku, J. Math. Phys. **25**, 2007 (1984); F. Neyzi and Y. Nutku, *ibid.* **28**, 1499 (1987).
[11]C. F. Hayes and J. M. Jankowski, Nuovo Cimento B **58**, 494 (1968).
[12]C. F. Hayes, J. Math. Phys. **10**, 1555 (1969).
[13]C. Ryan, J. Math. Phys. **13**, 283 (1972).
[14]D. Anderson, J. Math. Phys. **14**, 934 (1973).
[15]P. A. M. Dirac, *Lectures on Quantum Mechanics, Belfer Graduate School of Science Monograph Series No.* 2 (Yeshiva University, New York, 1964); K. Sundermeyer, *Constrained dynamics, Lecture Notes in Physics*, Vol. 169 (Springer, New York, 1982).
[16]H. Tesser, J. Math. Phys. **13**, 796 (1972).
[17]C. Cognola, L. Vanzo and S. Zerbini, Lett. Nuovo Cimento **38**, 533 (1982).
[18]V. Tapia, Nuovo Cimento B **90**, 15 (1985).
[19]J. Barcelos-Neto and N. R. F. Braga, "Higher order canonical formalism in field theory and constraints," Universidade Federal do Rio de Janeiro preprint IF/UFRJ/87/20, 1987.
[20]E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., Cambridge, 1959), 4th ed., p. 266.
[21]S. W. Hawking, "Lectures on quantum cosmology," in *Quantum Gravity and Cosmology*, edited by H. Sako and T. Inani (World Scientific, Singapore, 1985); J. Louko, Phys. Rev. D **35**, 3760 (1987); W. F. Blyth and C. J. Isham, Phys. Rev. D **11**, 768 (1975).
[22]R. Weiss, Proc. R. Soc. London Ser. A **129**, 102 (1938); J. S. de Wet, Proc. Cambridge Philos. Soc. **43**, 511 (1947).

Recursion operator and bi-Hamiltonian structure for integrable multidimensional lattices

O. Ragnisco and P. M. Santini
Dipartimento di Fisica, Università degli Studi "La Sapienza," 00184 Roma, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Roma, Italy

(Received 21 August 1987; accepted for publication 10 February 1988)

The recursion operator and the bi-Hamiltonian structure for an integrable two-dimensional version of the Toda chain are algorithmically derived from the corresponding linear problem. The intimate relation between two-dimensional theories and non-Abelian one-dimensional theories is emphasized. The analogies and differences between discrete and continuous integrable two-dimensional systems are pointed out and discussed.

I. INTRODUCTION

In the rich literature on integrable nonlinear evolution equations, two different, though related, aspects have been extensively and fruitfully investigated: (i) methods of solution, like the inverse spectral (scattering) transform for solving the Cauchy problem or, for instance, the Bäcklund transformation approach to get large classes of special solutions; (ii) algebraic and geometrical properties like the existence of a recursion operator, bi-Hamiltonian structure, and infinitely many integrals of motion in involution.

While for integrable systems in $1+1$ dimensions both aspects have been satisfactorily understood for a few years,[1,2] only recently a similar program has been accomplished for $(2+1)$-dimensional systems, with a nontrivial extension of the ideas and techniques previously used. Indeed, in $2+1$ dimensions the solution of the Cauchy problem requires the introduction of the $\bar\partial$ method, a natural extension of the Riemann–Hilbert method,[3] and the description of the geometric and algebraic structure is achieved via a suitable extension of the configuration space and consequently of the notions of gradients, symmetries, and recursion and Hamiltonian operators.[4–8] In a previous paper,[9] we have shown that the $\bar\partial$ method can be successfully used to solve multidimensional discrete spectral problems; specifically we have investigated the finite-difference equation

$$\psi(n-1,m)+\beta(n,m)\psi(n,m)+\alpha(n,m)\psi(n+1,m) = \lambda\psi(n,m+1), \tag{1.1}$$

first introduced in Ref. 10, which is a natural two-dimensional extension of the well-known linear problem associated with the Toda lattice and, moreover,[9] in a suitable continuous limit, yields the equation $\sigma\psi_y + Q\psi + \psi_{xx} = 0$, $\sigma\in\mathbb{R}$. The solution of the spectral problem (1.1) obviously implies the solution of the Cauchy problem for the associated class of nonlinear evolution equations in two discrete dimensions, whose simplest members are

$$\beta_t(n,m) = \prod_{j=0}^{\infty}\frac{\alpha(n-1+j,m-j)}{\alpha(n+j,m-j)} - \prod_{j=0}^{\infty}\frac{\alpha(n+j,m+1-j)}{\alpha(n+1+j,m+1-j)}, \tag{1.2a}$$

$$\alpha_t(n,m) = \beta(n,m)\prod_{j=0}^{\infty}\frac{\alpha(n+j,m-j)}{\alpha(n+1+j,m-j)} - \beta(n+1,m)\prod_{j=0}^{\infty}\frac{\alpha(n+j,m+1-j)}{\alpha(n+1+j,m+1-j)}, \tag{1.2b}$$

$$\beta_t(n,m) = 0, \tag{1.3a}$$

$$\alpha_t(n,m) = \alpha(n,m)\left[\prod_{j=0}^{\infty}\frac{\alpha(n+1+j,m-j)}{\alpha(n+2+j,m-j)} - \prod_{j=0}^{\infty}\frac{\alpha(n-1-j,m-j)}{\alpha(n+j,m-j)}\right]. \tag{1.3b}$$

Here we consider the second aspect of the theory of integrable systems and investigate the algebraic properties of the flows associated with the problem (1.1). Namely, we show that the notions of recursion operator and bi-Hamiltonian structure, first introduced for continuous systems in two space dimensions by Fokas and Santini,[4,5] have a natural counterpart in the corresponding discrete case. Our approach is based on the following two main points.

(i) The first point is the crucial role played by spectral problems for deriving, through an algorithmic procedure, all the relevant algebraic properties of the associated class of evolution equations. This algorithmic derivation, in many respects similar to the analogous one holding in the continuous case, presents, however, relevant and intriguing complications, which are peculiar to finite-difference problems.

(ii) The second point is the intimate connection between $(2+1)$-dimensional theories and non-Abelian one-dimensional ones, already established in the continuous case.[6,7] Indeed, we will show that the evolution equations associated with (1.1) can be obtained as "reductions" of those associated with an appropriate (singular) limit of a matrix one-dimensional spectral problem.

In Sec. II, we show that our discrete non-Abelian spectral problem yields the eigenfunctions of the recursion operator and, algorithmically, the operator itself; moreover, the knowledge of such eigenfunctions leads to a definite prescription for identifying the bi-Hamiltonian structure from the compatibility condition.

In Sec. III, we perform the reduction procedure to get the class of evolution equations associated with (1.1) together with their recursion and Hamiltonian operators. Throughout this reduction, the notion of "extended" sym-

 0022-2488/88/071593-11$02.50

metries and gradients, as well as the proper operations on them, are naturally introduced.

II. SPECTRAL PROBLEM, RECURSION OPERATOR, AND HAMILTONIAN STRUCTURE

A. The spectral problem

We consider the following linear problem:

$$\epsilon^{-1}\psi = (Q + \lambda a)\psi, \tag{2.1a}$$

$$Q = \begin{pmatrix} -B & -A \\ I & 0 \end{pmatrix}, \quad a = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.1b}$$

where ϵ is the shift operator

$$(\epsilon^k F)(n) := F(n+k) \quad (n,k \in Z), \tag{2.2}$$

λ is the spectral parameter, the fields A and B depend on the integer variable n (and possibly on the continuous variable t) and take values in some associative algebra $\mathscr{A}$ with unit element I, and σ is a constant invertible element in the algebra. The fields A and B fulfill the following boundary conditions:

$$\lim_{|n|\to\infty} A(n) = I, \quad \lim_{|n|\to\infty} B(n) = 0. \tag{2.3}$$

Note that one recovers the well-known spectral problem for the matrix Toda lattice[11] if the algebra is $\mathrm{gl}(N,C)$ and $\sigma = I$; the spectral problem (1.1) for the two-dimensional Toda lattice is instead obtained by considering matrices of infinite rank of the form

$$A(n;m_1,m_2) = \delta_{m_1,m_2}\alpha(n,m_1), \tag{2.4a}$$

$$B(n;m_1,m_2) = \delta_{m_1,m_2}\beta(n,m_1), \tag{2.4b}$$

$$\sigma(m_1,m_2) = \delta_{m_1+1,m_2}, \tag{2.4c}$$

where $\delta_{r,s}$ is the Kronecker symbol and m_1 and m_2, the indices of the matrices, in this infinite rank limit behave as lattice variables. Here the connection between one-dimensional matrix problems and two-dimensional scalar ones is particularly simple and direct.

B. Gradients and compatibility: Discrete versus continuous

For our purposes, it is convenient to introduce also the linear problem

$$\epsilon\tilde{\Psi} = \tilde{\Psi}(Q + \lambda a), \tag{2.5}$$

which is the adjoint of (2.1a) with respect to the natural bilinear form (we *assume* the algebra $\mathscr{A}$ to be endowed with a trace form)

$$\langle F,G\rangle = \mathrm{Tr}\sum_{n=-\infty}^{+\infty} F(n)G(n). \tag{2.6}$$

An elementary variational calculation yields the following expression for the gradient $\nabla_Q\lambda$ of the spectral parameter λ w.r.t. the field Q:

$$\delta\lambda = -\langle \Psi\tilde{\Psi}, \delta Q\rangle, \tag{2.7}$$

whence

$$\nabla_{Q_{ij}}\lambda = -(\Psi\tilde{\Psi})_{ji}. \tag{2.8}$$

The evolution equations associated with (2.1) are obtained, in a standard way, as compatibility conditions between (2.1) and

$$\Psi_t(n) = V(n+1)\Psi(n). \tag{2.9}$$

Then they are embedded in the equation

$$Q_t(n) = V(n)Q(n) - Q(n)V(n+1) + \lambda\,[V(n)a - aV(n+1)], \tag{2.10}$$

and arise by expanding V in powers of λ:

$$V = \sum_{\kappa\geq 0} \lambda^\kappa V^{(\kappa)}. \tag{2.11}$$

We can assert that the evolution equation (2.10) is a Hamiltonian system if it can be cast in the form

$$Q_t(n) = (\theta\gamma)(n), \tag{2.12}$$

when θ is a Hamiltonian (or Poisson, or cosymplectic) operator and γ is the gradient of some functional of the field Q, namely,

$$\theta^* = -\theta \quad \text{(skew symmetry)}, \tag{2.13a}$$

$$\langle \gamma_1, \theta'[\theta\gamma_2]\gamma_3\rangle + \text{cyclic permutation} = 0 \quad \text{(closure)}, \tag{2.13b}$$

$$\langle \gamma'[F], G\rangle = \langle \gamma'[G], F\rangle \quad \text{(irrotational property)}. \tag{2.13c}$$

In the above formulas, and everywhere in the following, an asterisk denotes the adjoint w.r.t. (2.6) and a prime denotes the directional derivative:

$$F'[G] = \frac{\partial}{\partial t} F(Q + tG)\Big|_{t=0}. \tag{2.13d}$$

In order to recognize the Hamiltonian structure of (2.11) we have to relate V to the gradient (2.8). To this aim, we notice that a simple calculation shows that the quantity

$$\Phi(n) = \Psi(n-1)\tilde{\Psi}(n) \tag{2.14}$$

fulfills the following "eigenvalue" equation:

$$0 = \Phi(n)Q(n) - Q(n)\Phi(n+1) + \lambda[\Phi(n)a - a\Phi(n+1)]. \tag{2.15}$$

The function Φ, though not itself a gradient, is, however, related to $\nabla_Q\lambda$ by the simple formulas

$$\Phi(n) = (Q(n) + \lambda a)(\nabla_Q\lambda)(n) = (\nabla_Q\lambda)(n-1)(Q(n-1) + \lambda a). \tag{2.16}$$

Moreover, a comparison between (2.10) and (2.15), once both Φ and V are expanded in powers of λ,

$$\Phi = \sum_{k\geq 0} \Phi^{(k)}\lambda^k, \tag{2.17a}$$

$$V = \sum_{k\geq 0} V^{(k)}\lambda^k, \tag{2.17b}$$

shows that, up to boundary terms, $\Phi^{(k)} = V^{(k)}$ for k strictly greater than zero.

It is interesting to compare this situation with the analogous but simpler one, characteristic of the continuous case:

$$\Psi_x = (Q + \lambda a)\Psi, \quad -\tilde{\Psi}_x = \tilde{\Psi}(Q + \lambda a), \tag{2.18a}$$

$$\Psi_t = V(\lambda)\Psi. \tag{2.18b}$$

Here we have

$$\nabla_Q \lambda = -\Psi\tilde{\Psi},$$

and moreover (the symbol $[\cdot,\cdot]$ denotes the commutator)

$$Q_t = V_x + [Q,V] + \lambda\,[a,V], \tag{2.19a}$$

$$0 = (\Psi\tilde{\Psi})_x + [Q,\Psi\tilde{\Psi}] + \lambda\,[a,\Psi\tilde{\Psi}]. \tag{2.19b}$$

Formulas (2.19) imply that the coefficients of the expansions in powers of λ of $\Psi\tilde{\Psi}$ (the "squared" eigenfunction) coincide, for $k>0$, with the homologous coefficients of V, which are themselves gradients: accordingly, Eq. (2.19a) is already written in bi-Hamiltonian form, the two Hamiltonian operators being

$$\theta_1 = \partial_x + [Q,\cdot], \quad \theta_2 = [a,\cdot]. \tag{2.20}$$

They are easily seen to be well coupled (or compatible), i.e., $\theta(\lambda) := \theta_1 + \lambda\theta_2$ is a one-parameter family of Hamiltonian operators.

Unlike the continuous case, we have seen before that in the discrete case the compatibility conditions do not immediately yield the evolution equations in bi-Hamiltonian form: The deep group-theoretical reason for this different behavior will be investigated elsewhere. Concretely, in the discrete case, although it is not evident how to identify *a priori* the basic Hamiltonian structures of the theory, nevertheless, the definite relation between the function V taking part in the compatibility (2.10) and the gradient (2.8) allows us to construct, for the choice (2.1b), the proper recursion and Hamiltonian operators of the theory.

C. Recursion operator and bi-Hamiltonian structure

We first observe that, for the choice (2.1b), the fundamental solution $\Psi,\tilde{\Psi}$ of the spectral problems (2.1a) and (2.5) can be parametrized as follows:

$$\Psi(n) = \begin{pmatrix} \psi(n) & \varphi(n) \\ \psi(n+1) & \varphi(n+1) \end{pmatrix},$$
$$\tilde{\Psi}(n) = \begin{pmatrix} \tilde{\psi}(n) & -\tilde{\psi}(n-1)A(n-1) \\ \tilde{\varphi}(n) & -\tilde{\varphi}(n-1)A(n-1) \end{pmatrix}. \tag{2.21}$$

Thus the relevant entries of the gradient matrix (2.8) take the form

$$(\nabla_B\lambda)(n) = (\Psi\tilde{\Psi})_{11} = \psi(n)\tilde{\psi}(n) + \varphi(n)\tilde{\varphi}(n), \tag{2.22a}$$

$$(\nabla_A\lambda)(n) = (\Psi\tilde{\Psi})_{21} = \psi(n+1)\tilde{\psi}(n) + \varphi(n+1)\tilde{\varphi}(n), \tag{2.22b}$$

whence it follows that the "squared eigenfunction" Φ (2.14) can be written as

$$\Phi(n) = \begin{pmatrix} \Phi_{11}(n) & -(\nabla_B\lambda)(n-1)A(n-1) \\ \nabla_B\lambda(n) & -(\nabla_A\lambda)(n-1)A(n-1) \end{pmatrix}. \tag{2.23}$$

This form of Φ, plugged into the "eigenvalue" equation (2.15), yields the following system:

$$\Phi_{11}(n)B(n) + (\nabla_B\lambda)(n-1) - B(n)\Phi_{11}(n+1) - A(n)(\nabla_B\lambda)(n+1) = \lambda\,[\sigma\Phi_{11}(n) - \Phi_{11}(n+1)\sigma], \tag{2.24a}$$

$$(\nabla_B\lambda)(n)B(n) + (\nabla_A\lambda)(n-1)A(n-1) + \Phi_{11}(n+1) = \lambda(\nabla_B\lambda)(n)\sigma, \tag{2.24b}$$

$$\Phi_{11}(n) + B(n)(\nabla_B\lambda)(n) + A(n)(\nabla_A\lambda)(n) = \lambda\sigma(\nabla_B\lambda)(n). \tag{2.24c}$$

Equations (2.24) enable us to express $\Phi_{11}(n)$ in terms of the gradients $\nabla_B\lambda$ and $\nabla_A\lambda$ through the formula

$$\Phi_{11} = P_{21}(\nabla_B\lambda) + P_{22}(\nabla_A\lambda), \tag{2.25a}$$

where

$$(P_{21}F)(n) = \sum \epsilon^j\sigma^j[\sigma(FB)(n) - (BF)(n)\sigma] \times\sigma^{-(j+1)}, \tag{2.25b}$$

$$(P_{22}F)(n) = \sum \epsilon^j\sigma^j[\sigma(\epsilon^{-1}FA)(n) - (AF)(n)\sigma] \times\sigma^{-(j+1)}, \quad \sum = \sum_{j=0}^{\infty}. \tag{2.25c}$$

Introducing this information into Eqs. (2.24a) and (2.24b), one gets, after a little algebra, the explicit expression of the "squared eigenfunction" operator N^*:

$$N^*\begin{bmatrix} \nabla_B\lambda \\ \nabla_A\lambda \end{bmatrix} = \lambda\begin{pmatrix} \nabla_B\lambda \\ \nabla_A\lambda \end{pmatrix}, \tag{2.26a}$$

$$\left(N^*\begin{bmatrix} g_1 \\ g \end{bmatrix}\right)(n) = \begin{pmatrix} \sum_{j=0}^{\infty}\sigma^j[(\epsilon^{j+1}Bg_1 - \epsilon^j g_1 B)(n) + (\epsilon^{j+1}Ag_2 - \epsilon^{j-1}g_2 A)(n)]\sigma^{-(j+1)} \\ \prod(n+1)\left\{\sum_{j=0}^{\infty}\epsilon^j\left[\left(\sigma\prod(n+1)\right)^{-1}(g_1(n) + B(n+1)g_2(n))\sigma\prod(n)\right]\right. \\ \left. -\epsilon^{j+1}\left[\left(\prod(n+1)\right)^{-1}(g_1(n+1) + g_2(n)B(n))\prod(n)\right]\right\}\left(\sigma\prod(n)\right)^{-1} \end{pmatrix}, \tag{2.26b}$$

where

$$\prod(n) := \prod_{j=n}^{\infty}\sigma^{-1}A(j). \tag{2.26c}$$

The operator N^* is the adjoint of the recursion operator N w.r.t. the bilinear form

$$\langle f,g\rangle := \mathrm{Tr}\sum_{n=-\infty}^{+\infty}(f_1(n)g_1(n)+f_2(n)g_2(n)), \tag{2.27}$$

where $f(n)=(f_1(n),f_2(n))$ [resp. $g(n)=(g_1(n),g_2(n))$] is an arbitrary element of the tangent (resp. cotangent) space to our manifold at the point $(B(n),A(n))$. Of course, the bilinear form (2.27) is induced by (2.6) through the reduction (2.1b). Taking into account (2.27), one gets the explicit form of N, which reads

$$\left(N\begin{bmatrix}f_1\\f_2\end{bmatrix}\right)(n) := \begin{bmatrix}f_1'(n)\\f_2'(n)\end{bmatrix},$$

$$\begin{aligned} f_1'(n) &= B(n)\sum_{j=0}^{\infty}\sigma^j(\epsilon^{j+1}f_1)(n)\sigma^{-(j+1)} - \sum_{j=0}^{\infty}\sigma^j(\epsilon^j f_1)(n)\sigma^{-(j+1)}B(n) + \sigma\prod(n)\sum_{j=0}^{\infty}\epsilon^{j+1}\left(\left(\sigma\prod\right)^{-1}f_2\epsilon\prod\right)(n)\\ &\quad\times\left(\sigma\prod(n+1)\right)^{-1} - \prod(n-1)\sum_{j=0}^{\infty}\epsilon^{j-1}\left(\left(\sigma\prod\right)^{-1}f_2\epsilon\prod\right)(n)\left(\prod(n)\right)^{-1}, \end{aligned} \tag{2.28}$$

$$\begin{aligned} f_2'(n) &= A(n)\epsilon^2\sum_{j=0}^{\infty}\sigma^j(\epsilon^j f_1)(n)\sigma^{-(j+1)} - \sum_{j=0}^{\infty}\sigma^j(\epsilon^j f_1)(n)\sigma^{-(j+1)}A(n) + \sigma\prod(n)\sum_{j=0}^{\infty}\epsilon^{j+1}\left(\left(\sigma\prod\right)^{-1}f_2\epsilon\prod\right)(n)\\ &\quad\times\left(\sigma\prod(n+1)\right)^{-1}B(n+1) - B(n)\prod(n)\sum_{j=0}^{\infty}\epsilon^j\left(\left(\sigma\prod\right)^{-1}f_2\epsilon\prod\right)(n)\left(\prod(n+1)\right)^{-1}. \end{aligned}$$

To derive the explicit form of the class of evolution equations associated with the spectral problem (2.1), one has to go back to the compatibility equation (2.10). By setting

$$V(n)=\begin{pmatrix}C(n) & E(n)\\ F(n) & G(n)\end{pmatrix}, \tag{2.29}$$

one easily gets

$$B_t(n) = C(n)B(n+1) - B(n)C(n+1) - E(n) - A(n)F(n+1) + \lambda\,[\sigma C(n+1) - C(n)\sigma], \tag{2.30a}$$

$$A_t(n) = C(n)A(n) - A(n)G(n+1) - B(n)E(n+1) + \lambda\sigma E(n+1), \tag{2.30b}$$

$$0 = F(n)B(n) - G(n) + C(n+1) - \lambda F(n)\sigma, \tag{2.30c}$$

$$0 = E(n+1) + F(n)A(n). \tag{2.30d}$$

By solving the constraints (2.30c) and (2.30d) w.r.t. $C(n)$, $F(n)$, Eqs. (2.30a) and (2.30b) can be recast in the form

$$\begin{bmatrix}B_t\\A_t\end{bmatrix} = (\vartheta_2+\lambda\vartheta_1)\begin{bmatrix}F\\C\end{bmatrix}, \tag{2.31}$$

where

$$\left(\vartheta_1\begin{bmatrix}F\\C\end{bmatrix}\right)(n) = \begin{pmatrix}\sigma C(n+1) - C(n)\sigma\\ A(n)F(n+1)\sigma - \sigma F(n)A(n)\end{pmatrix}, \tag{2.32a}$$

$$\left(\vartheta_2\begin{bmatrix}F\\C\end{bmatrix}\right)(n) = \begin{pmatrix}C(n)B(n) - B(n)C(n+1) + F(n-1)A(n-1) - A(n)F(n+1)\\ C(n)A(n) - A(n)C(n+2) + B(n)F(n)A(n) - A(n)F(n+1)B(n+1)\end{pmatrix}. \tag{2.32b}$$

By expanding C and F in powers of λ,

$$\begin{bmatrix}F\\C\end{bmatrix} = \sum_{j=0}^{N}\lambda^j\begin{bmatrix}F^{(N-j-1)}\\C^{(N-j-1)}\end{bmatrix}, \tag{2.33}$$

and equating the coefficients of equal powers of λ, we obtain (i) the starting point of the iteration (power λ^N),

$$\begin{bmatrix}F^{(-1)}(n)\\C^{(-1)}(n)\end{bmatrix} = \begin{bmatrix}\prod(n)H_1\sigma\left(\sigma\prod(n)\right)^{-1}\\ \sigma^{-n}H_2\sigma^n\end{bmatrix}, \tag{2.34}$$

where H_1,H_2 are arbitrary constant elements of the algebra $\mathscr{A}$; (ii) the recursion relation

$$\begin{bmatrix}F^{(K+1)}\\C^{(K+1)}\end{bmatrix} = M\begin{bmatrix}F^{(K)}\\C^{(K)}\end{bmatrix}, \tag{2.35}$$

where

$$M\begin{bmatrix}F\\C\end{bmatrix}=\vartheta_1^{-1}\vartheta_2\begin{bmatrix}F\\C\end{bmatrix}$$

$$=\begin{bmatrix}\prod(n)\left\{\sum_{j=0}^{\infty}\epsilon^j\left[\left(\sigma\prod\right)^{-1}(n)(CA-A\epsilon^2C+BFA-A\epsilon FB)(n)\left(\epsilon\prod\right)(n)\right]\right\}\left(\sigma\prod(n)\right)^{-1}\\ \sum_{j=0}^{\infty}\{\epsilon^j\sigma^j(B\epsilon C-CB+A\epsilon F-\epsilon^{-1}FA)(n)\sigma^{-(j+1)}\}\end{bmatrix};\tag{2.36}$$

and (iii) the class of evolution equations

$$\begin{bmatrix}B_t\\A_t\end{bmatrix}=\vartheta_2\begin{bmatrix}F^{(N-1)}\\C^{(N-1)}\end{bmatrix}=\vartheta_1M^{N+1}\begin{bmatrix}F^{(-1)}\\C^{(-1)}\end{bmatrix}.\tag{2.37}$$

However, the class (2.37) is not written in Hamiltonian form, since $\left[\begin{smallmatrix}F\\C\end{smallmatrix}\right]$ are not gradients (and, correspondingly, ϑ_1 and ϑ_2 are not Hamiltonian operators) as follows from formulas (2.14), (2.16), and (2.17). On the other hand, the same formulas, together with (2.23) imply that the proper gradients $\gamma^{(k)}$ are given by

$$\gamma^{(k)}=P^{-1}\begin{bmatrix}F^{(K)}\\C^{(K)}\end{bmatrix},\tag{2.38a}$$

$$\gamma^{(-1)}=\begin{bmatrix}\prod(n)H_1\sigma\left(\sigma\prod(n)\right)^{-1}\\ \prod(n+1)H_2\sigma\left(\sigma\prod(n)\right)^{-1}+\prod(n+1)\left\{\sum_{j=n+1}^{\infty}\left[\left(\sigma\prod(j)\right)^{-1}B(j)\prod(j),H_1\sigma\right]\right\}\left(\sigma\prod(n)\right)^{-1}\end{bmatrix},\tag{2.38b}$$

where the operator P is defined as

$$P\begin{bmatrix}g_1\\g_2\end{bmatrix}=\begin{bmatrix}I&0\\P_{21}&P_{22}\end{bmatrix}\begin{bmatrix}g_1\\g_2\end{bmatrix},\tag{2.38c}$$

the operators P_{21},P_{22} having been introduced in formulas (2.25). The transformation (2.38a) enables us to cast the class of evolution equations (2.37) in the following bi-Hamiltonian form:

$$\begin{bmatrix}B_t\\A_t\end{bmatrix}=\theta_2\gamma^{(N-1)}=\theta_1\gamma^{(N)},\tag{2.39}$$

where the Hamiltonian operators θ_1,θ_2 are given by

$$\theta_i:=\vartheta_iP\quad(i=1,2),\tag{2.40}$$

that is,

$$\left(\theta_1\begin{bmatrix}g_1\\g_2\end{bmatrix}\right)(n)=\begin{bmatrix}(Bg_1)(n)\sigma-\sigma(g_1B)(n)+(Ag_2)(n)\sigma-\sigma(\epsilon^{-1}g_2A)(n)\\(A\epsilon g_1)(n)\sigma-\sigma(g_1A)(n)\end{bmatrix},\tag{2.41a}$$

$$\left(\theta_2\begin{bmatrix}g_1\\g_2\end{bmatrix}\right)(n)$$

$$=\begin{bmatrix}(\epsilon^{-1}g_1A)(n)-(A\epsilon g_1)(n)+B(n)\sum_{j=0}^{\infty}\sigma^j(\epsilon^{j+1}(Bg_1+Ag_2)(n)\sigma^{-j}-\sigma^{j+1}(\epsilon^{j+1}g_1B+g_2\epsilon^{-1}A))(n)\sigma^{-(j+1)}\\(Bg_1A)(n)-(A\epsilon g_1B)(n)-\left[\sum_{j=0}^{\infty}\sigma^j(\epsilon^j(Bg_1+Ag_2-\sigma g_1B\sigma^{-1}-\sigma g_2\epsilon^{-1}A))(n)\sigma^{-j}\right]A(n)\end{bmatrix}$$

$$+\begin{bmatrix}-\sum_{j=0}^{\infty}[\sigma^j(\epsilon^j(Bg_1+Ag_2))(n)\sigma^{-j}-\sigma^{(j+1)}(\epsilon^j(g_1B+\epsilon^{-1}g_2A))(n)\sigma^{-(j+1)}]B(n)\\A(n)\sum_{j=0}^{\infty}\sigma^j(\epsilon^{j+2}(Bg_1+Ag_2-\sigma g_1B\sigma^{-1}-\sigma g_2\epsilon^{-1}A\sigma^{-1}))(n)\,\sigma^{-j}\end{bmatrix}.\tag{2.41b}$$

Equivalently, the class (2.39) can be written as

$$\begin{bmatrix}B_t\\A_t\end{bmatrix}=\theta_2N^{*N}\gamma^{(-1)}=\theta_1N^{*N+1}\gamma^{(-1)}=N^NK^{(0)}:=K^{(N)},\tag{2.42}$$

where the recursion operator N and its adjoint N^* have been defined in formulas (2.26b) and (2.28), and the starting vector field $K^{(0)}:=\theta_2\gamma^{(-1)}$ is given by

$$K^{(0)}(H):=K_1^{(0)}(H_1)+K_2^{(0)}(H_2), \tag{2.43a}$$

$$(K_1^{(0)}(H_1))(n)=\begin{bmatrix} \prod(n-1)H_1\sigma\left(\prod(n)\right)^{-1}-\sigma\prod(n)H_1\sigma\left(\sigma\prod(n+1)\right)^{-1} \\ B(n)\prod(n)H_1\sigma\left(\prod(n+1)\right)^{-1}-\sigma\prod(n)H_1\sigma\left(\sigma\prod(n+1)\right)^{-1}B(n+1)\end{bmatrix}, \tag{2.43b}$$

$$(K_2^{(0)}(H_2))(n)=\begin{bmatrix}\sigma^{-n}H_2\sigma^nB(n)-B(n)\sigma^{-(n+1)}H_2\sigma^{n+1} \\ \sigma^{-n}H_2\sigma^nA(n)-A(n)\sigma^{-(n+2)}H_2\sigma^{n+2}\end{bmatrix}. \tag{2.43c}$$

The proof that θ_1 and θ_2 are well-coupled (compatible) Hamiltonian operators, or, equivalently, that N is a Nijenhuis (or hereditary) operator, namely, it satisfies

$$N'[Nf]\tilde{f}-N'[N\tilde{f}]f=N(N'[f]\tilde{f}-N'[\tilde{f}]f) \tag{2.44}$$

for any pair of tangent vectors $f,\tilde{f}$, and θ_1 is a Hamiltonian operator well coupled with N, can be achieved by direct calculation. We recall that such results have been already established in Ref. 12 in the special case $\sigma=I$.

D. An interesting subclass of evolution equations

A direct calculation shows that the square of the recursion operator is reducible on the submanifold $B=0$: that is, its restriction $N^2|_{B=0}:=R$ is again a Nijenhuis operator mapping into itself the tangent bundle to this submanifold. Its explicit expression reads

$$\begin{aligned}(Rf)(n)=&-A(n)\left\{\prod(n+1)S(n+1)\left(\sigma\prod(n+2)\right)^{-1}+\sum_{j=0}^{\infty}\sigma^j(\epsilon^{j+2}f)(n)\sigma^{-(j+2)}\right\}\\&+\left\{\prod(n-1)S(n-1)\left(\sigma\prod(n)\right)^{-1}+\sum_{j=0}^{\infty}\sigma^j(\epsilon^jf)(n)\sigma^{-(j+2)}\right\}A(n),\end{aligned} \tag{2.45}$$

where we have set

$$S(n)=\sum\epsilon^j\left(\left(\sigma\prod(n)\right)^{-1}f(n)\prod(n+1)\right). \tag{2.46}$$

The same property is enjoyed by the Hamiltonian operator $\theta^{(2)}$: Namely, its restriction $\theta^{(2)}|_{B=0}:=\theta$ is again a Hamiltonian operator mapping the cotangent bundle into the tangent bundle to the submanifold $B=0$. It has the following form:

$$\begin{aligned}(\theta g)(n)=&-A(n)\left\{\sum_{j=0}^{\infty}\sigma^j[\epsilon^{j+2}(Ag)](n)\sigma^{-j}-\sigma^{j+1}[\epsilon^{j+1}(gA)](n)\sigma^{-(j+1)}\right\}\\&+\left\{\sum_{j=0}^{\infty}\sigma^j[\epsilon^j(Ag)](n)\sigma^{-j}-\sigma^{j+1}[\epsilon^{j-1}(gA)](n)\sigma^{-(j+1)}\right\}.\end{aligned} \tag{2.47}$$

As is well known,[13] this "reduction by restriction" preserves the coupling between Nijenhuis and Hamiltonian operators, so that $\theta':=R\theta$ is again a Hamiltonian operator compatible with θ. Then, observing that the vector fields $K^{(0)}(H_2)$, $K^{(1)}(H_1):=NK^{(0)}(H_1)$ can also be restricted to the same submanifold, giving rise to the reduced vector fields

$$\begin{aligned}\chi_1^{(0)}(H_1):=K_1^{(1)}(H_1)|_{B=0}=&\sigma\prod(n)H_2\sigma\left(\sigma\prod(n+2)\right)^{-1}-\prod(n-1)H_1\sigma\left(\prod(n+1)\right)^{-1}\\&+A(n)\sigma^{-(n+2)}H_1\sigma^{n+1}-\sigma^{-n}H_1\sigma^{n-1}A(n):=\theta\eta_1^{(0)}(H_1),\end{aligned} \tag{2.48}$$

$$\begin{aligned}&\eta_1^{(0)}(H_1)=\prod(n+1)H_1\sigma\left(\prod(n+1)\right)^{-1}\sigma^{-2}+\prod(n+1)\left\{\sum_{j=0}^{\infty}\left[\left(\epsilon^j\left(\sigma\epsilon\prod\right)^{-1}\prod\right)(n),H_1\sigma\right]\right\}\left(\sigma\prod(n)\right)^{-1},\\&\chi_2^{(0)}(H_2):=K_2^{(0)}(H_2)|_{B=0}=\sigma^{-n}H_2\sigma^nA(n)-A(n)\sigma^{-(n+2)}H_2\sigma^{n+2}:=\theta\eta_2^{(0)}(H_2),\\&\eta_2^{(0)}(H_2)=\prod(n+1)H_2\sigma\left(\sigma\prod(n)\right)^{-1},\end{aligned} \tag{2.49}$$

we can assert that the class of evolution equations (2.42) contains the bi-Hamiltonian subclass

$$A_t=R^K(\chi_1^{(0)}(H_1)+\chi_2^{(0)}(H_2)):=\chi^{(K)}(H)=R^K\theta(\eta_1^{(0)}(H_1)+\eta_2^{(0)}(H_2))=\theta R^{*K}(\eta_1^{(0)}(H_1)+\eta_2^{(0)}(H_2)):=\theta\eta^{(K)}(H). \tag{2.50}$$

This subclass has been already derived by Bruschi and Ragnisco,[14] in the special case $\sigma = I$: There it has been shown that its continuum limit is just the non-Abelian KdV class. It could also be shown, but we will not go into any detail, that a similar continuum limit performed on the hierarchy (2.50) in the special choice (2.4) allows recovery of the KP class recently derived by Fokas and Santini.[4,5]

E. Symmetries and integrals of motion

It is well known[2] that the existence of a compatible pair of Hamiltonian operators—the "bi-Hamiltonian structure"—entails the existence of infinitely many commuting symmetries and integrals of motion in involution whenever certain additional algebraic properties are satisfied by the starting members of a given hierarchy of evolution equations.

In our case, one can show that, for the class of evolution equations (2.22), the starting elements of the class, namely $K_i^{(0)}$ $(i = 1,2)$, defined in (2.43b) and (2.43c), enjoy the following Lie-algebraic properties:

$$[K_1^{(0)}(H_1),K_1^{(0)}(H_1')]_L = NK_1^{(0)}([H_1\sigma,H_1'\sigma]\sigma^{-1}), \tag{2.51a}$$

$$[K_2^{(0)}(H_2),K_2^{(0)}(H_2')]_L = K_2^{(0)}([H_2,H_2']), \tag{2.51b}$$

$$[K_1^{(0)}(H_1),K_2^{(0)}(H_2)]_L = 0, \tag{2.51c}$$

where by the symbol $[\cdot,\cdot]_L$ we have denoted the usual Lie bracket between two vector fields:

$$[K_1,K_2]_L = K_1'[K_2] - K_2'[K_1]. \tag{2.52}$$

Moreover, $\gamma^{(-1)}(H_1,H_2)$ is a gradient for any choice of H_1,H_2, since it spans the null space of the Hamiltonian operator $\theta^{(1)}$:

$$\langle f,\gamma^{(-1)\prime}[\tilde{f}]\rangle - \langle \tilde{f},\gamma^{(-1)\prime}[f]\rangle = 0. \tag{2.53}$$

Finally, N is a strong symmetry for $K_{1,2}^{(0)}$, namely,

$$N'[K_i^{(0)}]f + NK_i^{(0)\prime}[f] = K_i^{(0)\prime}[Nf] \quad (i = 1,2), \tag{2.54}$$

for any f.

Remark: Property (2.54) is also referred to as the vanishing of the Lie derivative of N along $K_i^{(0)}$. Actually, (2.54) is a consequence of the Nijenhuis property of N, of the good coupling between N,θ_1 and $\theta_2 = N\theta_1$, and of the definition of $K_i^{(0)} = \theta_2\gamma_i^{(-1)}$.

Due to the additional properties (2.51), (2.53), and (2.54), the following propositions hold true for the class of evolution equations (2.42).

(i) The flows

$$K^{(j)}(H) := N^j K^{(0)}(H) \tag{2.55a}$$

form an infinite set of commuting symmetries

$$[K^{(j)}(H),K^{(l)}(\tilde{H})]_L = 0, \tag{2.55b}$$

provided that

$$[H_1\sigma,\tilde{H}_1\sigma] = [H_2,\tilde{H}_2] = 0. \tag{2.55c}$$

Property (2.55c) is, in particular, guaranteed whenever the constant elements H_i, $\tilde{H}_i$ $(i = 1,2)$ are (analytic) functions of σ.

(ii) Under the same condition (2.55c), the functionals

$$I_H^{(j)}[V] := \int_0^1 d\mu\langle\gamma^{(j)}(\mu V;H),V\rangle, \tag{2.56}$$

where for a moment we have denoted by V the pair (B,A), form an infinite set of integrals of motion in involution, namely,

$$\{I_H^{(j)},I_H^{(l)}\} := \langle\gamma^{(j)}(H),\theta_1\gamma^{(l)}(\tilde{H})\rangle = \langle\gamma^{(j)}(H),\theta_2\gamma^{(l-1)}(\tilde{H})\rangle = 0. \tag{2.57}$$

Analogous results are valid for the subclass (2.50), where, however, the Lie-algebraic properties of the starting elements of the hierarchy are slightly different from (2.51), and agree with those established for the KP class.[5] In fact they read

$$[\chi_1^{(0)}(H_1),\chi_1^{(0)}(\tilde{H}_1)]_L = R\chi_1^{(0)}([H_1\sigma,\tilde{H}_1\sigma]\sigma^{-1}) + \chi_2^{(0)}([H_1\sigma,\tilde{H}_1\sigma]\sigma^{-1}), \tag{2.58a}$$

$$[\chi_1^{(0)}(H_1),\chi_2^{(0)}(H_2)]_L = -\chi_2^{(0)}([H_1\sigma,H_2]), \tag{2.58b}$$

$$[\chi_2^{(0)}(H_2),\chi_2^{(0)}(\tilde{H}_2)]_L = \chi_2^{(0)}([H_2,\tilde{H}_2]). \tag{2.58c}$$

Therefore, we can again assert that, for the subclass (2.50), the flows

$$\chi^{(j)}(H) := R^j\chi^{(0)}(H) \tag{2.59a}$$

form an infinite set of commuting symmetries, and the functionals

$$I_H^{(j)}[A] := \int_0^1 d\mu\langle\eta^{(j)}(\mu A;H),A\rangle \tag{2.59b}$$

provide an infinite set of integrals of motion in involution, once H_1, H_2, etc., are (analytic) functions of σ.

III. RECURSION OPERATOR AND HAMILTONIAN STRUCTURE FOR THE TWO-DIMENSIONAL TODA HIERARCHY

We have already pointed out that, by the choice (2.4), the spectral problem (2.1) becomes the spectral problem (1.1) associated with the two-dimensional Toda lattice. In this section we clarify to what extent, and in what sense, the bi-Hamiltonian hierarchy constructed in the previous section is reducible on the submanifold (2.4). We then derive the main aspects of the theory associated with the recursion and bi-Hamiltonian operators of the two-dimensional Toda hierarchy through this reduction procedure, extending the approach presented in Refs. 6 and 7 to our discrete context.

A. Reduction to the two-dimensional Toda class

We first notice that, with the positions (2.4), the Hamiltonian operators θ_i [(2.41a) and (2.41b)] and the recursion operator N (2.28) become

$$\theta_i := \theta_{12}^{(i)}, \quad N := N_{12} = \theta_{12}^{(2)}[\theta_{12}^{(1)}]^{-1}, \tag{3.1}$$

where

$$\theta_{12}^{(1)} = \begin{bmatrix} B_1 D_2^{-1} - B_2 D_1 & A_1 D_2^{-1} - D_1 \epsilon^{-1} A_2 \\ A_1 D_2^{-1} \epsilon - D_1 A_2 & 0 \end{bmatrix}, \tag{3.2a}$$

$$\theta_{12}^{(2)} = \begin{bmatrix} (B_2 - B_1\epsilon)(1 - \epsilon D_1 D_2)^{-1} D_2 (B_1 D_2^{-1} - B_2 D_1) & (B_2 - B_1\epsilon)(1 - \epsilon D_1 D_2)^{-1} \\ \quad + \epsilon^{-1} A_2 - A_1 \epsilon & \quad \times D_2 (A_1 D_2^{-1} - D_1 \epsilon^{-1} A_2) \\ B_1 A_2 - A_1 \epsilon B_2 + (A_2 - A_1 \epsilon^2)(1 - \epsilon D_1 D_2)^{-1} & (A_2 - A_1\epsilon^2)(1 - \epsilon D_1 D_2)^{-1} \\ \quad \times D_2 (B_1 D_2^{-1} - D_1 B_2) & \quad \times D_2 (A_1 D_2^{-1} - D_1 \epsilon^{-1} A_2) \end{bmatrix}, \tag{3.2b}$$

$$N_{12} = \begin{bmatrix} (B_1\epsilon - B_2) D_2 (1 - \epsilon D_1 D_2)^{-1} & [D_2 \Pi_{12} (D_1 D_2)^{-1} - (\epsilon D_1^{-1} \Pi_{12})] (1 - \epsilon D_1^{-1} D_2^{-1}) \Pi_{12}^{-1} \\ (A_1\epsilon^2 - A_2) D_2 (1 - \epsilon D_1 D_2)^{-1} & [D_2 \Pi_{12} (D_1 D_2)^{-1} \epsilon B_2 - B_1 D_1^{-1} \Pi_{12}] (1 - \epsilon D_1^{-1} D_2^{-1}) \Pi_{12}^{-1} \end{bmatrix}, \tag{3.2c}$$

where $\theta^{(i)}$ and N_{12} are operator-valued matrices acting on arbitrary two-component vectors, and we have introduced the notations

$$\Pi_{12} = \prod (n, m_1, m_2) = \prod_{j=0}^{\infty} [\alpha(n+j, m_1 - j)] / [\alpha(n + j + 1, m_2 - j)], \tag{3.3a}$$

$$(D_1^K f)(n, m_1, m_2) = f(n, m_1 + k, m_2), \tag{3.3b}$$

$$(D_2^K f)(n, m_1, m_2) = f(n, m_1, m_2 + k), \tag{3.3c}$$

$$A_i = \alpha(n, m_i), \quad B_i = \beta(n, m_i), \tag{3.3d}$$

$$[(1 - \epsilon^r (D_1 D_2)^s)^{-1} f](n, m_1, m_2) = \sum_{j=0}^{\infty} f(n + jr, m_1 + js, m_2 + js). \tag{3.3e}$$

The structure of the operators θ_{12}, N_{12} implies the following important commutationlike relations:

$$\theta_{12} h(m_1 - m_2) \cdot = h(m_1 - m_2 + 1) \theta_{12} \cdot, \tag{3.4a}$$

$$\theta_{12} h(m_1 - m_2) \cdot = h(m_1 - m_2) \theta_{12} \cdot, \tag{3.4b}$$

$$N_{12} h(m_1 - m_2) \cdot = h(m_1 - m_2 - 1) N_{12} \cdot, \tag{3.4c}$$

where h is an arbitrary scalar function. Formulas (3.4) mean in particular that, once applied to diagonal matrices $f(m_1, m_2) = \delta_{m_1, m_2} f(m_1)$, θ_{12} yields matrices of the form $f(m_1, m_2) = \delta_{m_1 + 1, m_2} f(m_1)$, θ_{12} yields diagonal matrices, and N_{12} yields matrices of the form $f(m_1, m_2) = \delta_{m_1 - 1, m_2} f(m_1)$.

These properties entail that the Mth vector field of the hierarchy (2.42) is given by a diagonal matrix and thus belongs to the tangent space of the submanifold (2.4), if the matrices H_1, H_2 entering in the definition of the starting vector field $K^{(0)}$ are chosen as

$$H_1(m_1, m_2) = h_1 \delta_{12}^{(M)}, \quad H_2(m_1, m_2) = h_2 \delta_{12}^{(M)}, \tag{3.5}$$

where h_1, h_2 are arbitrary real numbers and $\delta_{12}^{(k)} = \delta_{0, m_1 - m_2 + k}$. In this case the hierarchy (2.42) becomes the following hierarchy of nonlinear evolution equations associated with the spectral problem (1.1):

$$\begin{bmatrix} B_{1,t} \\ A_{1,t} \end{bmatrix} = \sum_{m_2} \delta_{12}^{(0)} N_{12}^M K_{12}^{(0)} (h_1 \delta_{12}^{(M)}, h_2 \delta_{12}^{(M)}) := \sum_{m_2} \delta_{12}^{(0)} N_{12}^M K_{12}^{(0)} (h \delta_{12}^{(M)}), \tag{3.6}$$

where

$$K_{12}^{(0)}(h_{12}) = \begin{bmatrix} B_2 - B_1 \\ A_2 - A_1 \end{bmatrix} h_{12}^{(2)} + \begin{bmatrix} (\epsilon^{-1} - D_1 D_2) \Pi_{12} \\ (B_1 - \epsilon B_2 \epsilon^{-1} D_1 D_2) \Pi_{12} \end{bmatrix} h_{12}^{(1)}; \tag{3.7}$$

the vector

$$h_{12} := \begin{bmatrix} h_{12}^{(1)} \\ h_{12}^{(2)} \end{bmatrix}$$

is hereafter a function of $m_1 - m_2$ and Σ_{m_2} stands for $\Sigma_{m_2 = -\infty}^{+\infty}$.

On the other hand, formulas (3.4) also imply the commutation of θ_{12}, N_{12} with the multiplication by any matrix having equal entries [corresponding to $h_{12} = \bar{h}, \forall (m_1, m_2)$], so that the same class of evolution equations (3.6) can be obtained by the following restriction to the diagonal:

$$\begin{bmatrix} B_{1,t} \\ A_{1,t} \end{bmatrix} = \sum_{m_2} \delta_{12}^{(0)} N_{12}^M K_{12}^{(0)} (\bar{h}) := \sum_{m_2} \delta_{12}^{(0)} K_{12}^{(M)} (\bar{h}) := K_{11}^{(M)} (\bar{h}). \tag{3.8}$$

We notice that the factorization property (2.42) allows us to cast Eqs. (3.8) in the equivalent forms

$$\begin{bmatrix} B_{1,t} \\ A_{1,t} \end{bmatrix} = \sum_{m_2} \delta_{12}^{(0)} \theta_{12}^{(2)} \gamma_{12}^{(M-1)} (\bar{h}) = \sum_{m_2} \delta_{12}^{(0)} \theta_{12}^{(1)} \gamma^{(M)} (\bar{h}), \tag{3.9}$$

where

$$\gamma_{12}^{(r)} = (\theta_{12}^{(1)})^{-1} K_{12}^{(r)} \quad (r = 0, 1, \ldots),$$

$$\gamma_{12}^{(-1)}(h_{12}) = \begin{bmatrix} 0 \\ \Pi_{12} / (A_1 A_2) \end{bmatrix} h^{(2)}(m_1 - m_2 + 1) + \begin{bmatrix} D_1^{-1} \Pi_{12} / A_2 \\ \Pi_{12} / (A_1 A_2) \epsilon (1 - \epsilon)^{-1} (B_1 \Pi_1 - B_2 \Pi_2) \end{bmatrix} \times h^{(1)}(m_1 - m_2 + 1), \tag{3.10}$$

having defined

$$\Pi_i(n) = \prod_{k=0}^{\infty} \frac{\alpha(n + k, m_i - 1 - k)}{\alpha(n + k, m_i - k)}.$$

Moreover, formula (3.4b) allows us to write Eq. (3.9) in the

following form:

$$\begin{bmatrix} B_1 \\ A_1 \end{bmatrix}_t = \theta_1^{(2)} \gamma_{11}^{(M-1)}(\bar{h}), \tag{3.11a}$$

where

$$\gamma_{11}^{(r)} = \gamma_{12}^{(r)}|_{m_2 = m_1} \tag{3.11b}$$

and

$$\theta_1^{(2)} = \begin{bmatrix} B_1(1-\epsilon)(1-\epsilon D_1)^{-1}(1-D_1)B_1 + \epsilon^{-1}A_1 - A_1\epsilon & B_1(1-\epsilon)(1-\epsilon D_1)^{-1}(1-D_1\epsilon^{-1})A_1 \\ A_1(1-\epsilon)B_1 + A_1(1-\epsilon^2)(1-\epsilon D_1)^{-1}(1-D_1)B_1 & A_1(1-\epsilon^2)(1-\epsilon D_1)^{-1}(1-D_1\epsilon^{-1})A_1 \end{bmatrix}. \tag{3.11c}$$

It is important to remark that $\theta_1^{(2)}$, i.e., the restriction to the diagonal of $\theta_{12}^{(2)}$, is again a Hamiltonian operator: indeed, it is skew symmetric with respect to the (restricted) bilinear form

$$\langle f,g \rangle = \sum_{i=1}^{2} \sum_{n} \sum_{m_1} f^{(i)}(n,m_1) g^{(i)}(n,m_1), \tag{3.12a}$$

and fulfills the closure condition w.r.t. the Fréchet derivative

$$\theta^{(2)\prime}[f] := \frac{\partial}{\partial t} \theta^{(2)}[B_1 + tf^{(1)}, A_1 + tf^{(2)}]|_{t=0} \qquad [f^{(i)} = f^{(i)}(n,m_1)]. \tag{3.12b}$$

Remark: Similar considerations apply to the subclass $B = 0$ discussed in Sec. II D. The Hamiltonian operator θ_{12} and the recursion operator R_{12} take the form

$$\theta_{12} = (A_2 - A_1\epsilon^2)(1 - \epsilon D_1 D_2)^{-1} \times (A_2 - \epsilon^{-1} D_1 D_2 A_1), \tag{3.13a}$$

$$R_{12} = (A_2\epsilon^{-1} - A_1\epsilon)\,(D_1^{-1} D_2\, \Pi_{12} D_1 D_2^{-1}) \times (1 - \epsilon D_1^{-1} D_2^{-1}) \Pi_{12}^{-1} + (A_2 - A_1\epsilon^2) D_2^2 (1 - \epsilon D_1 D_2)^{-1}. \tag{3.13b}$$

Of course, being the restriction of $\theta_{12}^{(2)}$ to the submanifold $B = 0$, θ_{12} obeys again the commutation relation (3.4b), while R_{12}, being a restriction of N_{12}^2, will instead obey the following one:

$$R_{12} h(m_1 - m_2)\cdot = h(m_1 - m_2 - 2) R_{12}\cdot. \tag{3.14}$$

It follows that, in order to guarantee that the Mth vector field of the hierarchy (2.50) be represented by a diagonal matrix, one has to choose the matrices H_i, entering in the definition of the starting vector fields, as

$$H_2(m_1,m_2) = h^{(2)} \delta_{12}^{(2M)}, \tag{3.15a}$$

$$H_1(m_1,m_2) = h^{(1)} \delta_{12}^{(2M+1)}. \tag{3.15b}$$

In this case, one gets nonlinear evolution equations of the form

$$A_{1,t} = \sum_{m_2} (R_{12}^M \chi^{(0)}(h\delta_{12}^{(2M)})) \delta_{12}^{(0)}, \tag{3.16}$$

the starting vector field $\chi^{(0)}$ being given by

$$\chi_{12}^{(0)}(h_{12}) = (A_2 - A_1) h^{(2)}(m_1 - m_2) + (A_1 - A_2 + (A_1\epsilon - A_2\epsilon^{-1})\,\Pi_{12}) \times h^{(1)}(m_1 - m_2 + 1). \tag{3.17}$$

The alternative representation, corresponding to formula (3.8), reads

$$A_{1,t} = \sum_{m_2} \delta_{12}^{(0)} R_{12}^M \chi_{12}^{(0)}(\bar{h}) := \sum_{m_2} \delta_{12}^{(0)} \chi_{12}^{(M)}(\bar{h}) = \chi_{11}^{(M)}(\bar{h}), \tag{3.18}$$

which can be also written in the Hamiltonian form

$$A_{1,t} = \sum_{m_2} \delta_{12}^{(0)} \theta_{12} \gamma_{12}^{(M-1)}(\bar{h}) := \theta_1 \gamma_{11}^{(M-1)}(\bar{h}), \tag{3.19}$$

where the Hamiltonian operator θ_1 is the restriction on the submanifold $B = 0$ of $\theta_1^{(2)}$ [formula (3.11c)], namely,

$$\theta = A_1(I - \epsilon^2)(I - \epsilon D_1)^{-1}(I - \epsilon^{-1} D_1) A_1. \tag{3.20}$$

We can thus answer the question raised at the beginning of this section, stating that the class of evolution equations (2.42) [or (2.50)] is *reducible* on the submanifold (2.4), in the sense that the generic vector field of the hierarchy belongs to the tangent bundle of this submanifold, provided that the starting point of the iteration is suitably shifted with respect to the diagonal; moreover, any equation of the hierarchy so obtained is certainly endowed with a Hamiltonian structure. Alternatively, we can say that the same hierarchy of Hamiltonian vector fields, tangent to the submanifold (2.4), can be extracted from the general class (2.42) [or (2.50)], by *restricting to the diagonal* the vector fields generated by a unique starting point, given by $K^{(0)}(\bar{h})$ [or $\chi^{(0)}(\bar{h})$]. To end this subsection, we notice that, by setting $M = 0$, Eq. (3.8) yields Eq. (1.2), while Eq. (3.18) yields Eq. (1.3).

B. Extended directional derivative and bilinear form

In Sec. III A, we have shown that on the manifold (2.4) it is natural to introduce vector and covector fields like K_{12}, γ_{12} and operators like N_{12}, θ_{12}. Such fields and operators, which arise as *special realizations* of the abstract mathematical objects of the non-Abelian theory developed in Sec. II, can be, on the other hand, considered as *multidimensional extensions* of the analogous quantities taking part in the standard one-dimensional Abelian theory. In order to construct a self-consistent theory in terms of such *extended objects*, namely, to prove that the relevant properties derived in Sec. II E are preserved by the "reduction" (2.4), we need first of all an appropriate definition of the directional derivative. To this aim, we notice that the extended vector and covector fields as well as the extended recursion operator are constructed in terms of the operators θ_{12} and (possibly) of γ_{12}. Moreover, the field-dependent entries of θ_{12} contain just

terms of the form

$$\text{(i)}\quad B_1\epsilon^p D_2^{-q} - D_1^q\epsilon^r B_2 =: \hat{B}_{12}(p,q,r), \tag{3.21a}$$

$$\text{(ii)}\quad A_1\epsilon^p D_2^{-q} - D_1^q\epsilon^r A_2 =: \hat{A}_{12}(p,q,r), \tag{3.21b}$$

$$\text{(iii)}\quad B_1A_2 - A_1\epsilon B_2 =: \hat{C}_{12}. \tag{3.21c}$$

Hence, the directional derivative of $\theta_{12}^{(i)}$, and consequently of $N_{12},K_{12},\gamma_{12}$, will be uniquely defined in terms of the directional derivative of $\hat{A}_{12},\hat{B}_{12},\hat{C}_{12}$, which we define here, consistently with the non-Abelian origin of these operators. The directional derivatives of $\hat{B}_{12}$, $\hat{A}_{12}$, and $\hat{C}_{12}$ in an arbitrary direction

$$f_{12} =: f(n,m_1,m_2) =: \begin{bmatrix} f_{12}^{(1)} \\ f_{12}^{(2)} \end{bmatrix}$$

are denoted by $B_{12,d}[f_{12}]$, $A_{12,d}[f_{12}]$, $C_{12,d}[f_{12}]$, respectively, and given by

$$\hat{B}_{12,d}[f_{12}]g_{12}^{(i)} =: \sum_{m_3}[f_{13}^{(1)}\epsilon^p D_2^{-q} g_{32}^{(i)} - D_1^q\epsilon^r g_{13}^{(i)} f_{32}^{(1)}], \tag{3.22a}$$

$$\hat{A}_{12,d}[f_{12}]g_{12}^{(i)} =: \sum_{m_3}[f_{13}^{(2)}\epsilon^p D_2^{-q} g_{32}^{(i)} - D_1^q\epsilon^r g_{13}^{(i)} f_{32}^{(2)}], \tag{3.22b}$$

$$\hat{C}_{12,d}[f_{12}]g_{12}^{(i)} = \sum_{m_3}[f_{13}^{(1)} g_{32}^{(i)} A_2 + B_1 g_{13}^{(i)} f_{32}^{(2)} - f_{13}^{(2)}\epsilon g_{32}^{(i)} B_2 - A_1\epsilon g_{13}^{(i)} f_{32}^{(1)}]. \tag{3.22c}$$

Besides this d derivative (hereafter referred to also as *extended directional derivative*), it is also natural to introduce the *total Fréchet derivative*, denoted by the subscript F: for any operator $\hat{O}$ (or vector field K, or covector field γ) depending on A_i,B_i $(i=1,2)$, this total Fréchet derivative is defined as

$$\hat{O}_F[f]g_{12} = \frac{\partial}{\partial t}\sum_{\substack{i=1 \\ j\neq i}}^{2} \hat{O}(B_i + tf_{ii}^{(1)},A_i + tf_{ii}^{(2)};B_j,A_j)g_{12}|_{t=0}. \tag{3.23}$$

It is easy to check that on operator valued matrices $\hat{O}_{12}$, defined in terms of $\hat{A}_{12},\hat{B}_{12},\hat{C}_{12}$, like our Hamiltonian operators $\theta_{12}^{(i)}$ and recursion operator N_{12}, the following relation between the d derivative and the F derivative holds:

$$\hat{O}_{12,d}[\delta_{12}^{(0)} f_{12}]g_{12} = \hat{O}_{12,F}[f]g_{12}. \tag{3.23'}$$

Definition (2.27) finally motivates the following definition of extended bilinear form:

$$\langle f_{12},g_{12}\rangle =: \sum_{i=1}^{2}\sum_{n}\sum_{m_1,m_2} f_{12}^{(i)} g_{21}^{(i)}. \tag{3.24}$$

Due to the definition of the d derivative and of the extended bilinear form, which are the ones induced by the *ordinary* directional derivative and bilinear form for the non-Abelian original structure, we can also assert, with no need of any further explicit proof, that in terms of these extended operations the following properties hold.

(1) $\theta_{12}^{(1)}$ and $\theta_{12}^{(2)}$ are *extended compatible Hamiltonian operators*, and thus $N_{12} = \theta_{12}^{(2)}[\theta_{12}^{(1)}]^{-1}$ is a *Nijenhuis operator*, well coupled with both $\theta_{12}^{(1)}$ and $\theta_{12}^{(2)}$.

(2) All the *extended* vector fields $K_{12}^{(r)}$ (3.8) constructed via the Nijenhuis operator N_{12} out of the starting vector field (3.7) commute in pairs for any $\hat{h}_{12}$ of the type $\hat{h}_{12} = \Sigma C_k\delta_{12}^{(k)}$, C_k constants, namely,

$$[K_{12}^{(r)}(\hat{h}_{12}),K_{12}^{(s)}(\hat{h}'_{12})]_d = 0. \tag{3.25a}$$

(3) All the quantities $\gamma_{12}^{(r)}$ (3.10) are *extended* gradients, namely,

$$\langle f_{12},\gamma_{12,d}^{(r)}[g_{12}]\rangle = \langle g_{12},\gamma_{12,d}^{(r)}[f_{12}]\rangle. \tag{3.25b}$$

(4) We have

$$\langle\gamma_{12}^{(r)}(\hat{h}_{12}),\theta_{12}^{(i)}\gamma_{12}^{(s)}(\hat{h}'_{12})\rangle = 0 \tag{3.25c}$$

for any $\hat{h}_{12},\hat{h}'_{12}$ and for any non-negative integers r,s.

Remark: Equations (3.25) are direct consequences of Eqs. (2.55) and (2.57) since $\hat{h}_{12}$ is nothing but an arbitrary function of σ in the reduction (2.4).

C. Commuting symmetries and integrals of motion in involution for the two-dimensional Toda hierarchy

In order to show that the two-dimensional Toda hierarchy defined in (3.6) and (3.8) is endowed with infinitely many commuting symmetries and infinitely many integrals of motion in involution, we need to establish certain nontrivial consequences of (3.24) and properties (1)–(4).

We start by introducing the notion of "extended symmetry": An extended vector field σ_{12} is called an *extended symmetry* for the evolution equation (3.8) if it enjoys the following property:

$$\sigma_{12,F}[K^{(n)}] = (\delta_{12}^{(0)}K_{12}^{(n)})_d[\sigma_{12}]. \tag{3.26}$$

It is not difficult to check that the quantities

$$\sigma_{12}^{(m)} =: N_{12}^m K_{12}^{(0)}(\bar{h}) \tag{3.27}$$

are extended symmetries for any evolution equation of the hierarchy (3.8) for arbitrary $\bar{h}$. Indeed, due to formula (3.23') we have

$$\sigma_{12,F}^{(m)}[K^{(n)}] = \sigma_{12,d}^m[\delta_{12}^{(0)}K_{12}^{(n)}], \tag{3.28a}$$

and moreover,

$$\begin{aligned}\sigma_{12,d}^{(m)}[\delta_{12}^{(0)}K_{12}^{(n)}] - (\delta_{12}^{(0)}K_{12}^{(n)})_d[\sigma_{12}^{(m)}] &= \sigma_{12,d}^{(m)}[\delta_{12}^{(0)}K_{12}^{(n)}] - (N_{12}^n K_{12}^{(0)}(\bar{h}'\delta_{12}^{(n)}))_d[\sigma_{12}^{(m)}] \\ &= (N_{12}^{(m)}K_{12}^{(0)}(\bar{h}))_d[N_{12}^n K_{12}^{(0)}(\bar{h}'\delta_{12}^{(n)})] \\ &\quad - (N_{12}^n K_{12}^{(0)}(\bar{h}'\delta_{12}^{(n)}))_d[N_{12}^m K_{12}^{(0)}(\bar{h})] \\ &=: [N_{12}^m K_{12}^{(0)}(\bar{h}),N_{12}^n K_{12}^{(0)}(\bar{h}'\delta_{12}^{(n)})]_d \\ &= N_{12}^{n+m}[K_{12}^{(0)}(\bar{h}),K_{12}^{(0)}(\bar{h}'\delta_{12}^{(n)})] = 0,\end{aligned} \tag{3.28b}$$

since the matrices associated with $(\bar{h}')$ and $(\bar{h}\delta_{12}^{(n)})$ obviously commute for any n.

There are two main consequences of (3.26).

(i) The restriction of σ_{12} to the diagonal,

$$\sigma_{11}^{(m)} = \sum_{m_2}\delta_{12}^{(0)}\sigma_{12}^{(m)}, \tag{3.29a}$$

is a symmetry for each equation of the two-dimensional Toda hierarchy, namely,

$$\sigma_{11}^{(m)\prime}[K_{11}^{(n)}] = K_{11}^{(n)\prime}[\sigma_{11}^{(m)}]. \tag{3.29b}$$

(ii) The equation

$$\sigma_{12}^{(m)} = 0 \tag{3.29c}$$

is an auto-Bäcklund transformation for any equation of the hierarchy.

The way to prove (i) is rather involved and, on the other hand, fairly analogous to the one followed in the continuous case (5); so we will not repeat it in detail, just summarize its main steps. First, we observe that property (3.26) entails

$$\delta_{12}^{(0)}\sigma_{12,t}^{(m)} = \delta_{12}^{(0)}\sigma_{12,d}^{(m)}[\delta_{12}^{(0)}K_{12}^{(n)}] = \delta_{12}^{(0)}(\delta_{12}^{(0)}K_{12}^{(n)})_d[\sigma_{12}^{(m)}]. \tag{3.30a}$$

Moreover, using the "commutationlike" property of N_{12} with an arbitrary function of $(m_1 - m_2)$ (3.4c), one can establish by induction the following *nontrivial* identity:

$$\delta_{12}^{(0)}(\delta_{12}^{(0)}K_{12}^{(n)})_d[f_{12}] = (\delta_{12}^{(0)}K_{12}^{(n)})_d[\delta_{12}^{(0)}f_{12}] \tag{3.30b}$$

(for any vector field f_{12}), which allows us to rewrite (3.30a) in the form

$$\delta_{12}^{(0)}\sigma_{12,t}^{(m)} = (\delta_{12}^{(0)}K_{12}^{(n)})_{,d}[\delta_{12}^{(0)}\sigma_{12}^{(m)}], \tag{3.30c}$$

whence, by summing up w.r.t. m_2,

$$\sigma_{11,t}^{(m)} = K_{11}^{(n)\prime}[\sigma_{11}^{(m)}], \tag{3.30d}$$

or, equivalently,

$$[\sigma_{11}^{(m)},K_{11}^{(n)}]_L = 0. \tag{3.30e}$$

To prove that (3.29c) provides an auto-Bäcklund transformation for (3.8), we first notice that, as a functional difference equation relating the fields (B_1,A_1) to the fields (B_2,A_2), it will provide an auto-Bäcklund transformation for the flow (3.8) iff, assumed to hold at a given time t_0, it will stay valid along the flow for any time. But the time derivative of $\sigma_{12}^{(m)}$ along the flow is just

$$\sigma_{12,t}^{(m)} = \sigma_{12,d}^{(m)}[\delta_{12}^{(0)}K_{12}^{(n)}], \tag{3.31a}$$

so that (3.26) implies

$$\sigma_{12,t}^{(m)} = (\delta_{12}^{(0)}K_{12}^{(n)})_{,d}[\sigma_{12}^{(m)}] = 0. \tag{3.31b}$$

The existence of infinitely many integrals of motion in involution for the hierarchy (3.8) is also a clear consequence of the analogous property of the non-Abelian hierarchy (2.42). Indeed we have the following results.

(i) The functionals $I_H^{(m)}[A,B]$ introduced in the previous section, formula (2.56) *evaluated on the submanifold (2.4)*, are integrals of motion in involution for any flow of the hierarchy (3.8), i.e., for any positive integers n,m, we have

$$\frac{d}{dt}I_{\bar{h}}^{(m)} = \{I_{\bar{h}}^{(m)},I_{\bar{h}'}^{(n)}\}_{\text{ext}} = 0, \tag{3.32}$$

where the *extended Poisson brackets* are defined by

$$\{I_{\bar{h}}^{(m)},I_{\bar{h}'}^{(n)}\} := \langle\gamma_{12}^{(m)}(\bar{h}),\delta_{12}^{(0)}\theta_{12}\gamma_{12}^{(n)}(\bar{h}')\rangle,$$
$$\theta_{12} = \theta_{12}^{(1)} \text{ or } \theta_{12}^{(2)}. \tag{3.33}$$

(ii) The restriction of the extended gradients $\gamma^{(r)}(h_{12})$ to the diagonal

$$\gamma_{11}^{(r)} = \sum_{m_2}\delta_{12}^{(0)}\gamma_{12}^{(r)} \tag{3.34a}$$

yields ordinary gradients, namely,

$$(\gamma^{(r)\prime}[g_{11}],f_{11}) = (\gamma^{(r)\prime}[f_{11}],g_{11}). \tag{3.34b}$$

In order to prove Eq. (3.34b), we use Eq. (3.25b), letting $f_{12} = \delta_{12}^{(0)}f_{11}$ and $g_{12} = \delta_{12}^{(0)}g_{11}$. Equations (3.32) and (3.33) follow from (3.25c) by letting $\hat{h}_{12} = \bar{h}$, and $\bar{h}'_{12} = \bar{h}'\delta_{12}^{(n)}$ if $\theta_{12} = \theta_{12}^{(2)}$ or $\hat{h}'_{12} = \bar{h}'\delta_{12}^{(n+1)}$ if $\theta_{12} = \theta_{12}^{(1)}$.

Remarks: (i) Since the Hamiltonian operator $\theta_{12}^{(2)}$ commutes with $\delta_{12}^{(0)}$, formulas (3.32), (3.33), and (3.34a) entail that the ordinary Poisson bracket

$$\{I_{\bar{h}}^{(m)}I_{\bar{h}'}^{(n)}\} = (\gamma_{11}^{(m)},\theta_{11}^{(2)}\gamma_{11}^{(n-1)}) \tag{3.35}$$

vanishes identically for any positive integers n,m.

(ii) Analogous considerations, both for the symmetries and the integrals of motion, hold, of course, for the reduction $B = 0$, characterized by the hierarchy of evolution equations (3.16) or (3.18) and (3.19).

ACKNOWLEDGMENT

This work has been partially supported by the Italian Ministry of Public Education (M.P.I. 40% grant devoted to the research project, "Metodi analitici e tecniche di calcolo per lo studio delle onde nonlineari").

[1] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981); F. Calogero and A. Degasperis, *Spectral Transform and Solitons I* (North-Holland, Amsterdam, 1982).
[2] V. E. Zakharov and L. D. Faddeev, Funct. Anal. Appl. **5**, 287 (1981); M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974); I. M. Gel'fand and L. A. Dikij, Funct. Anal. Appl. **11**, 93 (1977); I. M. Gel'fand and J. Ya. Dorfman, *ibid.* **13**, 248 (1979); F. Magri, J. Math. Phys. **19**, 1156 (1978); F. Magri and C. Morosi, Quaderno **S19**, Universitá di Milano (1984); P. J. Olver, J. Math. Phys. **18**, 1212 (1977); B. Fuchssteiner, Nonlinear Anal. **3**, 849 (1979); A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980); B. Fuchssteiner and A.S. Fokas, Phys. D **4**, 47 (1981); A. S. Fokas and R. L. Anderson, J. Math. Phys. **23**, 1066 (1982); B. A. Kupershmidt and G. Wilson, Invent. Math. **62**, 403 (1981); Yu. I. Manin, J. Sov. Math. **11**, 1 (1979); M. Adler, Invent. Math. **50**, 219 (1979).
[3] A. S. Fokas and M. J. Ablowitz, "The inverse scattering transform for multi-dimensional (2 + 1) problems," in *Lecture Notes in Physics Vol. 189* (Springer, Berlin, 1983), pp. 139–183.
[4] A. S. Fokas and P. M. Santini, Stud. Appl. Math. **75**, 179 (1986).
[5] P. M. Santini and A. S. Fokas, "Recursion operators and bi-Hamiltonian structures in multidimensions I," Institute for Nonlinear Studies preprint INS 65, 1986.
[6] A. S. Fokas and P. M. Santini, "Recursion operators and bi-Hamiltonian structures in multidimensions II," Institute for Nonlinear Studies preprint INS 67, 1986.
[7] P. M. Santini, "Integrable 2 + 1 dimensional equations," Dipartimento di Fisica, Universita di Roma preprint No. 540, 1987.
[8] F. Magri and C. Morosi, "Old and new results on recursion operators: an algebraic approach to KP equation," Dipartimento di Matematica, Universita di Milano preprint, 1987.
[9] O. Ragnisco, P. M. Santini, M. J. Ablowitz, and S. Chitlaru Briggs, "An example of $\bar{\partial}$ problem arising in a finite difference context: Direct and inverse problem for the discrete analog of the equation $\psi_{xx} + u\psi = \sigma\psi_y'$," J. Math. Phys. **28**, 777 (1987).
[10] D. Levi, L. Pilloni, and P. M. Santini, J. Phys. A: Math. Gen. **14**, 1567 (1981).
[11] M. Bruschi, S. V. Manakov, D. Levi, and O. Ragnisco, J. Math. Phys. **21**, 2749 (1980).
[12] M. Bruschi and O. Ragnisco, J. Math. Phys. **24**, 1414 (1983).
[13] F. Magri, C. Morosi, and O. Ragnisco, Commun. Math. Phys. **99**, 115 (1985).
[14] M. Bruschi, O. Ragnisco, and D. Levi, J. Math. Phys. **22**, 2463 (1981).

Application of the Hamilton–Jacobi method to linear nonconservative vibration theory

B. Vujanovic
Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Yugoslavia

A. M. Strauss
Department of Mechanical and Materials Engineering, Vanderbilt University, Nashville, Tennessee 37235

(Received 18 December 1986; accepted for publication 9 March 1988)

In this paper it is shown how linear, strictly nonconservative, oscillatory systems with one and many degrees of freedom may be analyzed by application of the theory of Hamilton and Jacobi.

I. INTRODUCTION

In this study we illustrate the application of the Hamilton–Jacobi method to the linear single and multidegrees of freedom oscillatory systems that contain nonconservative elements. The motivation for this study is that linear dissipative systems, possessing even one degree of freedom, have not been analyzed from the standpoint of the Hamilton–Jacobi method, despite the fact that it is of interest from practical, theoretical, and pedagogical points of view. The exception in the literature is the paper of Denman and Buch[1] in which a linearly damped harmonic oscillator and an uncoupled linearly damped oscillator in three dimensions are considered by means of the Hamilton–Jacobi theory.

In our analysis the study of the oscillatory problems is analyzed along the sequence:

Newtonian equations → Lagrangian function → Hamiltonian → Hamilton–Jacobi → solution.

It is known that the first step has been known as the inverse problem of the calculus of variations or inverse Lagrangian problem. We demonstrate here that the *form* of the Langragian function of the problem (which is developed in the full accordance of the inverse Lagrangian theory) is of primary importance in our considerations. Namely, in each particular case, we demand that the Lagrangian function is composed of the quadratic time-dependent terms. This requirement considerably simplifies the Hamilton–Jacobi equations.

For the multidegree of freedom nonconservative, linear oscillatory systems, the starting Newtonian differential equations of motion, are non-self-adjoint, i.e., they are not immediately derivable from a Lagrangian function. Thus we initially apply the indirect Lagrangian representation (as suggested in numerous contemporary references of the inverse Lagrangian problem, as for example, Refs. 2–8), i.e., we find a suitably chosen set of the constant multipliers, transforming the Newtonian non-self-adjoint system into a self-adjoint one, whose Lagrangian again has the full quadratic time-dependent structure.

II. FORCED LINEARLY DAMPED OSCILLATOR WITH ONE DEGREE OF FREEDOM

Let us consider the forced, linearly damped oscillator whose differential equation of motion is

$$\ddot{x} + 2k\dot{x} + \omega_0^2 x = h\cos\Omega t, \tag{2.1}$$

where x is the position coordinate, t is the time, k, ω_0, and h are given constants, and an overdot denotes the time derivative. At the same time, consider the time-dependent Lagrangian function

$$L = \{\tfrac{1}{2}[\dot{x} - \dot{f}(t)]^2 - \tfrac{1}{2}[x - f(t)]^2\}e^{2kt}, \tag{2.2}$$

where $f(t)$ is an unknown function of time that should be determined in such a way that Lagrange's equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0 \tag{2.3}$$

generates the correct differential equation (2.1) up to a certain multiplier.

By inserting (2.2) into (2.3) one obtains

$$(\ddot{x} + 2k\dot{x} + \omega_0^2 x)e^{2kt} = (\ddot{f} + 2k\dot{f} + \omega_0^2 f)e^{2kt}.$$

Therefore if $f(t)$ is selected as

$$\ddot{f} + 2k\dot{f} + \omega_0^2 f \equiv h\cos\Omega t,$$

Eq. (2.3) will be reduced to the equation of motion (2.1) multiplied by the factor e^{2kt}. The function $f(t)$ is found to be

$$f = A\cos\Omega t + B\sin\Omega t \tag{2.4}$$

with

$$A = \frac{h(\omega_0^2 - \Omega^2)}{(\omega_0^2 - \Omega^2)^2 + 4k^2\Omega^2}, \quad B = \frac{2kh\Omega}{(\omega_0^2 - \Omega^2)^2 + 4k^2\Omega^2}. \tag{2.5}$$

The Lagrangian function (2.2) is therefore

$$L = \{\tfrac{1}{2}[\dot{x} + \Omega(A\sin\Omega t - B\cos\Omega t)]^2 - \tfrac{1}{2}\omega_0^2[x - (A\cos\Omega t + B\sin\Omega t)]^2\}e^{2kt}. \tag{2.6}$$

We will take the Lagrangian function (2.6) as a basis for constructing the Hamilton–Jacobi partial differential equation.

Hamilton's function is of the form

$$H = \tfrac{1}{2}p^2 e^{-2kt} - p\Omega(A\sin\Omega t - B\cos\Omega t) + \tfrac{1}{2}\omega_0^2[x - (A\cos\Omega t + B\sin\Omega t)]^2 e^{2kt}. \tag{2.7}$$

 0022-2488/88/071604-06$02.50

The canonical differential equations of motion are

$$\dot{x} = \frac{\partial H}{\partial p} = pe^{2kt} - \Omega(A \sin \Omega t - B \cos \Omega t),$$
$$\dot{p} = -\frac{\partial H}{\partial x} = -\omega_0^2 [x - (A \cos \Omega t + B \sin \Omega t)] e^{2kt}. \tag{2.8}$$

It is well known from the Hamilton–Jacobi method that a general solution of these equations can be obtained from a complete solution of the Hamilton–Jacobi differential equation

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 e^{-2kt} - (A \sin \Omega t - B \cos \Omega t)\Omega \frac{\partial S}{\partial x} + \frac{1}{2}\omega_0^2 [x - (A \cos \Omega t + B \sin \Omega t)]^2 e^{2kt} = 0, \tag{2.9}$$

where $S = S(t,x)$ is Hamilton's field function.

Let us try to find a complete solution of this equation in the form

$$S = \tfrac{1}{2}[x - (A \cos \Omega + B \sin \Omega t)]^2 e^{2kt} F(t), \tag{2.10}$$

where $F(t)$ is an unknown function of time. Substituting this into (2.9) we obtain

$$\dot{F} + 2kF + F^2 + \omega_0^2 = 0.$$

By integrating we find

$$F = -[k + \omega \tan(\omega t + C)],$$

where C is a constant and

$$\omega = \sqrt{\omega_0^2 - k^2}. \tag{2.11}$$

Thus a complete solution of (2.9) is

$$S(t,x,C) = -\tfrac{1}{2}[x - (A \cos \Omega t + B \sin \Omega t)]^2 \times [k + \omega \tan(\Omega t + C)] e^{2kt}. \tag{2.12}$$

According to the well-known Hamilton–Jacobi theorem (see, for example, Ref. 9, p. 275) the general solution of the canonical system (2.8) can be obtained from the complete solution (2.12) of the Hamilton–Jacobi equation (2.9) by means of the following rules:

$$p = \frac{\partial S}{\partial x}, \quad \frac{\partial S}{\partial C} = B, \tag{2.13}$$

where B is a new constant.

Hence

$$p = -[x - (A \cos \Omega t + B \sin \Omega t)] \times [k + \omega \tan(\omega t + C)] e^{2kt},$$
$$B = -\tfrac{1}{2}[x - (A \cos \Omega t + B \sin \Omega t)]^2 \omega e^{2kt} \times [\cos^2(\omega t + C)]^{-1}. \tag{2.14}$$

From the last equation one has

$$x = (-2B/\omega) e^{-kt} \cos(\omega t + C) + A \cos \Omega t + B \sin \Omega t$$
$$= (-2B/\omega)^{1/2} e^{-kt} \cos(\omega t + C) + \frac{h(\omega_0^2 - \Omega^2)\cos \Omega t + 2kh\Omega \sin \Omega t}{(\omega_0^2 - \Omega^2)^2 + 4k^2\Omega^2}, \tag{2.15}$$

which is the well-known solution of (2.1). Inserting this result into the first equation (2.14) we find the momentum as a function of time:

$$p = -(-2B/\omega)^{1/2} e^{kt} [\omega \sin(\omega t + C) + k \cos(\omega t + C)]. \tag{2.16}$$

The following special cases are of interest.

A. Harmonic oscillator

If the dissipative and force terms in (2.1) are negligible, i.e., $k = 0$, $h = 0$, the complete solution (2.12) becomes

$$S = -\tfrac{1}{2}\omega_0^2 x^2 \tan(\omega_0 t + C). \tag{2.17}$$

This solution of the Hamilton–Jacobi equation,

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 + \frac{\omega_0^2 x^2}{2} = 0,$$

is not frequently employed in the literature, although it seems that it can be of use as a starting point for the study of nonconservative phenomena.

B. Linearly damped oscillator

For the linearly damped oscillator without the force term $h = 0$ (i.e., $A = 0$, $B = 0$), the Hamilton–Jacobi equation (2.9) and the corresponding complete solution (2.12) are, respectively,

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 e^{-2kt} + \frac{1}{2}\omega_0^2 x^2 e^{2kt} = 0, \tag{2.18}$$

$$S(t,x,C) = -\tfrac{1}{2}x^2 [k + \omega \tan(\omega t + C)] e^{2kt}. \tag{2.19}$$

It is interesting to note that the same problem was treated in Ref. 1 by supposing that a complete solution of (2.18) was of the form

$$S(t,x) = -Q(x)e^{2kt}.$$

For this case, the problem is reduced to the differential equation

$$\left(\frac{dQ}{dx}\right)^2 - 4kQ + \omega_0^2 x^2 = 0, \tag{2.20}$$

which cannot be solved in a closed form. The authors applied an implicit procedure in the application of the Hamilton–Jacobi theorem (2.13), which makes the process of finding the motion rather complicated.

C. Forced oscillator

If the oscillator is not subject to the dissipative force ($k = 0$) but to a periodic external force acting upon it ($h \neq 0$), the Hamilton–Jacobi equation and its complete solution are, respectively,

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 - \frac{h\Omega \sin \Omega t}{\omega_0^2 - \Omega^2}\frac{\partial S}{\partial x} + \frac{1}{2}\omega_0^2\left(x - h\frac{\cos \Omega t}{\omega_0^2 - \Omega^2}\right)^2 = 0, \tag{2.21}$$

$$S(t,x,C) = -\frac{1}{2}\omega_0\left(x - \frac{h \cos \Omega t}{\omega_0^2 - \Omega^2}\right)^2 \tan(\omega_0 t + C). \tag{2.22}$$

It is easy to verify that the application of the Hamilton–Jacobi theorem (2.13) will produce the familiar solution

$$x = (-2B/\omega_0)^{1/2}\cos(\omega_0 t + C) + (h\cos\Omega t)/(\omega_0^2 - \Omega^2)$$

and the corresponding expression for the momentum follows from (2.16) for $k = 0$:

$$p = -(-2B\omega_0)^{1/2}\sin(\omega_0 t + C).$$

D. Resonance

For the case of resonance, the differential equation of motion

$$\ddot{x} + \omega^2 x = h\cos\omega t \tag{2.23}$$

can be derived from a Lagrangian function of the form

$$L = \frac{1}{2}\left(\dot{x} - \frac{1}{2}ht\cos\omega t - \frac{h}{2\omega}\sin\omega t\right)^2 - \frac{1}{2}\omega^2\left(x - \frac{ht}{2\omega}\sin\omega t\right)^2. \tag{2.24}$$

Hamilton's function becomes

$$H = \frac{1}{2}p^2 + \left(\frac{1}{2}ht\cos\omega t + \frac{h}{2\omega}\sin\omega t\right)p + \frac{1}{2}\omega^2\left(x - \frac{ht}{2\omega}\sin\omega t\right)^2. \tag{2.25}$$

Therefore the general solution of the canonical equations of motion

$$\dot{x} = \frac{\partial H}{\partial p} = p + \frac{1}{2}ht\cos\omega t + \frac{1}{2}\frac{h}{\omega}\sin\omega t,$$
$$\dot{p} = -\frac{\partial H}{\partial x} = -\omega^2\left(x - \frac{ht}{2\omega}\sin\omega t\right) \tag{2.26}$$

can be obtained from the Hamilton–Jacobi equation

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{1}{2}ht\cos\omega t + \frac{h}{2\omega}\sin\omega t\right)\frac{\partial S}{\partial x} + \frac{1}{2}\omega^2\left(x - \frac{ht}{2\omega}\sin\omega t\right)^2 = 0. \tag{2.27}$$

A complete solution of this equation is found to be

$$S(t,x,C) = -\tfrac{1}{2}\omega(x - (ht/2\omega)\sin\omega t)^2\tan(\omega t + C). \tag{2.28}$$

It is easy to verify that the equation $\partial S/\partial C = B$ generates the familiar solution for the resonant case

$$x = (-2B/\omega)^{1/2}\cos(\omega t + C) + (ht/2\omega)\sin\omega t.$$

Note that the Lagrangian function (2.6) of the dynamical system (2.1) can serve as a basis for the study of the conservative laws of this system. For more details about these conservation laws see Ref. 10.

III. MULTIDEGREE OF FREEDOM OSCILLATORY SYSTEM WITHOUT RESTRICTIONS ON THE DISSIPATIVE TENSOR

In the following two sections we demonstrate that under some circumstances the Hamilton–Jacobi method can be successfully applied to purely nonconservative linear systems whose original (Newtonian) form of differential equations of motion does not possess a Lagrangian function. In the light of the theory of the inverse Lagrangian problem, such dynamical systems are usually termed as non-self-adjoint (Ref. 2, pp. 54–61). However, we confine ourselves to those dynamical systems with the following two important properties.

(a) There exists a set of the constant integrating factors by means of which we can transform the equations of motion into a self-adjoint form.

(b) The form of the Lagrangian function of the Lagrangian function obtained in this way should be composed of the full quadratic time-dependent terms, suitable (liable) for the Hamilton–Jacobi analysis. Let us consider the linear dissipative oscillatory dynamical system

$$G_i \equiv \ddot{x}_i + b_{ij}\dot{x}_j + \omega^2 x_i = 0 \quad (i,j = 1,\ldots,n), \tag{3.1}$$

where b_{ij} and ω are a set of given constants. In (3.1) and throughout the text, we suppose that the summation convention with respect to the repeated indices is applied. In order to include gyroscopic forces in our considerations (whose tensor, as a rule, is antisymmetric: $b_{ij} = -b_{ji}$) and dissipative forces (whose tensor, as a rule is symmetric: $b_{ij} = b_{ji}$), we will suppose that the coefficients b_{ij} are not symmetric, i.e., $b_{ij} \neq b_{ji}$.

The differential equations (3.1) as they stand are not derivable from a Lagrangian function. To show this, we note that if a system of the differential equations $G_i(t,x_1,\ldots,x_n,\dot{x}_1,\ldots,\dot{x}_n) = 0$ $(i = 1,\ldots,n)$ is derivable from a Lagrangian function $L(t,x^1,\ldots,x_n,x_1,\ldots,x_n)$ the following Helmholtz conditions should hold identically for all $i, k = l,\ldots,n$ (for a modern treatment of this subject see Refs. 2, 3, and 8):

$$\frac{\partial G_i}{\partial \ddot{x}_k} = \frac{\partial G_k}{\partial \ddot{x}_i},$$
$$\frac{\partial G_i}{\partial \dot{x}_k} + \frac{\partial G_k}{\partial \dot{x}_i} = \frac{d}{dt}\left[\frac{\partial G_i}{\partial \ddot{x}_k} + \frac{\partial G_k}{\partial \ddot{x}_i}\right], \tag{3.2}$$
$$2\left[\frac{\partial G_i}{\partial x_k} - \frac{\partial G_k}{\partial x_i}\right] = \frac{d}{dt}\left[\frac{\partial G_i}{\partial \dot{x}_k} - \frac{\partial G_k}{\partial \dot{x}_i}\right].$$

Since these conditions are not satisfied for the Newtonian system (3.1), we turn to the possibility of transforming it into a Lagrangian form by means of the procedure below. As outlined in Refs. 2–8, the introduction of the integrating factors (multipliers) into a Newtonian system can considerably enhance finding a Lagrangian (if it exists). Thus we replace the original set of Newtonian equations $G_i = 0$ by a linear combination of these equations, by introducing a set of constant multipliers f_{ij}, with $\det(f_{ij}) \neq 0$. In what follows, we apply this procedure on the Newtonian nonconservative system.

Let us take the first differential equation of the system (3.1), i.e., $i = 1$ as it stands, multiply the second ($i = 2$) by a constant factor λ_2, the third ($i = 3$) by λ_3, etc., and sum all equations. Thus we obtain

$$(\ddot{x}_1 + \lambda_\alpha\ddot{x}_\alpha) + (b_{11} + \lambda_\alpha b_{\alpha 1})\dot{x}_1 + (b_{12} + \lambda_\alpha b_{\alpha 2})\dot{x}_2 + \cdots + (b_{1n} + \lambda_\alpha b_{\alpha n})\dot{x}_n + \omega^2(x_1 + \lambda_\alpha x_\alpha) = 0 \quad (\alpha = 2,3,\ldots,n) \tag{3.3}$$

or

$$(x_1+\lambda_\alpha x_\alpha)^{\cdot\cdot}+(b_{11}+\lambda_\alpha b_{\alpha 1})\times\left[\dot{x}_1+\frac{b_{12}+\lambda_\alpha b_{\alpha 2}}{b_{11}+\lambda_\alpha b_{\alpha 1}}\dot{x}_2+\cdots+\frac{b_{1n}+\lambda_\alpha b_{\alpha n}}{b_{11}+\lambda_\alpha b_{\alpha 1}}\dot{x}_n\right]+\omega^2(x_1+\lambda_\alpha x_\alpha)=0. \tag{3.4}$$

By introducing the new variable

$$Y=x_1+\lambda_\alpha x_\alpha, \tag{3.5}$$

selecting the multipliers λ_α $(=2,3,...,n)$ in such a way that

$$\frac{b_{12}+\lambda_\alpha b_{\alpha 2}}{b_{11}+\lambda_\alpha b_{\alpha 1}}=\lambda_2,...,\frac{b_{1n}+\lambda_\alpha b_{\alpha n}}{b_{11}+\lambda_\alpha b_{\alpha 1}}=\lambda_n, \tag{3.6}$$

and denoting

$$2k=b_{11}+\lambda_\alpha b_{\alpha 1}, \tag{3.7}$$

we reduce Eq. (3.3) to the form

$$\ddot{Y}+2k\dot{Y}+\omega^2 Y=0, \tag{3.8}$$

which can be expressed in terms of Lagrangian or Hamiltonian mechanics.

Note that employing (3.7) we write the system (3.6) in the form of $(n-1)$ algebraic equations

$$\begin{aligned}&\lambda_2(b_{22}-2k)+\lambda_3 b_{32}+\cdots+\lambda_n b_{n2}=-b_{12},\\&\lambda_2 b_{32}+\lambda_3(b_{33}-2k)+\cdots+\lambda_n b_{n3}=-b_{13},\\&\quad\vdots\\&\lambda_2 b_{2n}+\lambda_3 b_{3n}+\cdots+\lambda_n(b_{nn}-2k)=-b_{1n}.\end{aligned} \tag{3.9}$$

If we solve this system with respect to λ and substitute the solution into (3.7), we obtain the determinantal equation

$$\begin{vmatrix} b_{11}-2k, b_{12},...,b_{1n}\\ b_{21}, b_{22}-2k,...,b_{2n}\\ \vdots\\ b_{n1}, b_{n2},...,b_{nn}-2k\end{vmatrix}=0, \tag{3.10}$$

which is an algebraic equation of the nth order with respect to $(2k)$.

Solving (3.10) by supposing that there are no multiple roots, we obtain $2k_{(1)},...,2k_{(n)}$. Substituting this into (3.9) we find $\lambda_{(2)},...,\lambda_{(n)}$.

Therefore Eqs. (3.5), (3.7), and (3.8) are, in fact,

$$Y_{(i)}=x_1+\lambda_\alpha(i)x_\alpha, \tag{3.11}$$

$$2k_{(i)}=b_{11}+\lambda_\alpha(i)b_{\alpha 1}, \tag{3.12}$$

$$\ddot{Y}_{(i)}+2k_{(i)}\dot{Y}_{(i)}+\omega^2 Y_{(i)}=0. \tag{3.13}$$

Note that in the last three equations and in what follows, the summation convention with respect to the repeated Latin indices in the parentheses (i), is not assumed, while this convention is applied to the repeated Greek indices.

The Lagrangian and Hamiltonian corresponding to (3.13) are, respectively,

$$L=\sum_{(i)=1}^{n}\frac{1}{2}\left[\dot{Y}^2_{(i)}-\omega^2 Y^2_{(i)}\right]e^{2k_{(i)}t}, \tag{3.14}$$

$$H=\sum_{(i)=1}^{n}\frac{1}{2}\left[p^2_{(i)}e^{-2k_{(i)}t}+\omega^2 Y^2_{(i)}e^{2k_{(i)}t}\right], \tag{3.15}$$

and the associated Hamilton–Jacobi equation is

$$\frac{\partial S}{\partial t}+\sum_{(i)}\frac{1}{2}\left[\left(\frac{\partial S}{\partial Y_{(i)}}\right)^2 e^{-2k(i)t}+\omega^2 Y^2_{(i)}e^{2k(i)t}\right]=0. \tag{3.16}$$

To solve this equation we apply the method of separation of variables

$$S(t,Y_{(1)},Y_{(2)},...,Y_{(n)})=S_{(1)}(t,Y_{(1)})+S_{(2)}(t,Y_{(2)})+\cdots+S_{(n)}(t,Y_{(n)}) \tag{3.17}$$

and, substituting this into the previous equation, we obtain n equations for

$$\begin{aligned}&S_{(i)}(t,Y_{(i)}),\quad (i)=1,...,n,\\&\frac{\partial S_{(i)}}{\partial t}+\frac{1}{2}\left[\left(\frac{\partial S_{(i)}}{\partial Y_{(i)}}\right)^2 e^{-2k(i)t}+\omega^2 Y^2_{(i)}e^{2k(i)t}\right]=0,\\&\qquad (i)=1,...,n.\end{aligned} \tag{3.18}$$

Since for each particular (i), (3.18) is identical to (2.18), we have a complete solution of (3.18) in the form

$$S_{(i)}=-\tfrac{1}{2}Y^2_{(i)}\left[k_{(i)}+\omega\tan(\omega t+C_{(i)})\right]e^{2k(i)t}, \tag{3.19}$$

where the $C_{(i)}$ are constant parameters.

Applying the Jacobi theorem we have

$$p_{(i)}=\frac{\partial S_{(i)}}{\partial Y_{(i)}}=-Y_{(i)}\left[k_{(i)}+\omega\tan(\omega t+C_{(i)})\right]e^{2k(i)t}, \tag{3.20}$$

$$\frac{\partial S_{(i)}}{\partial C_{(i)}}=-\frac{1}{2}Y^2_{(i)}\frac{\omega e^{2k(i)t}}{\cos^2(\omega t+C_{(i)})}=B_{(i)}=\text{const.} \tag{3.21}$$

From the last group of equations we find

$$Y_{(i)}=\left[-2B_{(i)}/\omega\right]^{1/2}e^{-k(i)t}\cos(\omega t+C_{(i)}). \tag{3.22}$$

Using the last equation, we express the generalized momentum vector (3.20) as a function of time:

$$p_{(i)}=\left[k_{(i)}(-2B_{(i)}/\omega)^{1/2}\cos(\omega t+C_{(i)})+(-2B_{(i)}\omega)^{1/2}\sin(\omega t+C_{(i)})\right]e^{k_{(i)}t}. \tag{3.23}$$

If the initial conditions of the system (3.1) are given we can find, by using (3.11) and the relation

$$p_{(i)}=\frac{\partial L}{\partial\dot{Y}_{(i)}}=\dot{Y}_{(i)}e^{2k_{(i)}t}, \tag{3.24}$$

for $t=0$, $Y_{(i)}(0)$ and $p_{(i)}(0)$. Therefore from (3.22) and (3.24) we can find $2n$ constants $B_{(i)}$ and $C_{(i)}$ in terms of a_i and b_i. The motion of the original system is given by (3.11). Namely, we suppose that this expression can be inverted to given $x_i=x_i(t)$.

IV. OSCILLATORY SYSTEM WITHOUT RESTRICTIONS ON THE STIFFNESS COEFFICIENTS

The method of Hamilton and Jacobi can also be applied to dynamical systems whose differential equations of motion are of the form

$$G_i\equiv\ddot{x}_i+2k\dot{x}_i+a_{ij}x_j=0\quad(i,j=1,...,n), \tag{4.1}$$

where k and a_{ij} are given constants.

In order to include the nonconservative position forces in consideration, we suppose that the coefficients a_{ij} are not symmetric, i.e.,

$$a_{ij} \neq a_{ji}, \tag{4.2}$$

therefore, on the basis of Eq. (3.2) the dynamical system is not describable by means of a Lagrangian function.

However, repeating the same process as in the previous section, i.e., multiplying the second equation of the system (4.1) by a constant factor μ_2, the third equation by μ_3, etc., and summing all equations, including the first, we obtain

$$\ddot{X} + 2k\dot{X} + (a_{11} + \mu_\alpha a_{\alpha 1})x_1 + (a_{12} + \mu_\alpha a_{\alpha 2})x_2 + \cdots + (a_{1n} + \mu_\alpha a_{\alpha n})x_n = 0, \tag{4.3}$$

where

$$X = x_1 + \mu_\alpha x_\alpha \quad (\alpha = 2,3,...,n). \tag{4.4}$$

By denoting

$$\omega^2 = a_{11} + \mu_\alpha a_{\alpha 1} \tag{4.5}$$

and selecting in such a manner that

$$\frac{a_{12} + \mu_\alpha a_{\alpha 2}}{a_{11} + \mu_\alpha a_{\alpha 1}} = \mu_2, \ldots, \frac{a_{1n} + \mu_\alpha a_{\alpha n}}{a_{11} + \mu_\alpha a_{\alpha 1}} = \mu_n, \tag{4.6}$$

we reduce Eq. (4.3) to

$$\ddot{X} + 2k\dot{X} + \omega^2 X = 0. \tag{4.7}$$

To determine ω^2 and the unknown multipliers μ_α $(\alpha = 2,3,...,n)$, we first, by means of (4.5), express system (4.6) as

$$\begin{aligned} &\mu_2(a_{22} - \omega^2) + \mu_3 a_{32} + \cdots + \mu_n a_{n2} = -a_{12}, \\ &\mu_2 a_{32} + \mu_3(a_{33} - \omega^2) + \cdots + \mu_n a_{n3} = -a_{13}, \\ &\quad \vdots \\ &\mu_2 a_{2n} + \mu_3 a_{3n} + \cdots + \mu_n(a_{nn} - \omega^2) = -a_{1n}. \end{aligned} \tag{4.8}$$

Solving this system with respect to μ_α we substitute the solution into (4.5), and after simple transformations we obtain the determinantal equation

$$\begin{vmatrix} a_{11} - \omega^2, a_{12}, \ldots, a_{1n} \\ a_{21}, a_{22} - \omega^2, \ldots, a_{2n} \\ \vdots \\ a_{n1}, a_{n2}, \ldots, a_{nn} - \omega^2 \end{vmatrix} = 0, \tag{4.9}$$

which is an algebraic equation of the nth order with respect to ω^2. Thus from (4.9) and (4.8) we find $\omega^2_{(1)}, \ldots, \omega^2_{(n)}$ and $\mu_2, \ldots, \mu_n$ and from (4.4), (4.5) and (4.7) we actually have

$$X_{(i)} = x_1 + \mu_{\alpha(i)} x_\alpha, \tag{4.10}$$

$$\omega^2_{(i)} = a_{11} + \mu_{\alpha(i)} a_{\alpha 1}, \tag{4.11}$$

$$\ddot{X}_{(i)} + 2k\dot{X}_{(i)} + \omega^2_{(i)} X_{(i)} = 0 \quad [\alpha = 2,3,...,n, \quad (i) = 1,2,...,n], \tag{4.12}$$

where the repeated indices in the parentheses ( ) are not summed. Since the transformed system (4.12) can be generated by means of the Lagrangian function

$$L = \sum_{(i)=1}^{n} \frac{1}{2} [\dot{X}^2_{(i)} - \omega^2_{(i)} X^2_{(i)}] e^{2kt}. \tag{4.13}$$

It is interesting to note that the dynamic systems (3.1) and (4.1) admit some quadratic conservation laws, which are discussed in Refs. 11 and 12.

V. AN EXAMPLE

As an illustration, consider the dynamic system (Ref. 13, p. 87)

$$\begin{aligned} &\ddot{x} + \dot{y} + 3x = 15e^{-t}, \quad x(0) = 35, \quad \dot{x}(0) = -48, \\ &\ddot{y} - 4\dot{x} + 3y = 15 \sin 2t, \quad y(0) = 27, \quad \dot{y}(0) = -55. \end{aligned} \tag{5.1}$$

By applying the multiplier method, we reduce this system to

$$\begin{aligned} &\ddot{X} + 2i\dot{X} + 3X = 15(e^{-t} - (i/2)\sin 2t), \\ &\ddot{Y} - 2i\dot{Y} + 3Y = 15(e^{-t} + (i/2)\sin 2t), \end{aligned} \tag{5.2}$$

where

$$X = x - (i/2)y, \quad Y = x + (i/2)y, \quad i^2 = -1. \tag{5.3}$$

By repeating the same process as in Sec. II, we find the Lagrangian function of the system (5.2) in the form

$$\begin{aligned} L = &\{\tfrac{1}{2}[\dot{X} + (3 + \tfrac{3}{2}i)e^{-t} + i\cos 2t + 4\sin 2t]^2 \\ &- \tfrac{3}{2}[X - (3 + \tfrac{3}{2}i)e^{-t} + (i/2)\sin 2t - 2\cos 2t]^2\}e^{2it} \\ &+ \{\tfrac{1}{2}[\dot{Y} + (3 - \tfrac{3}{2}i)e^{-t} - i\cos 2t + 4\sin 2t]^2 \\ &- \tfrac{3}{2}[Y - (3 - \tfrac{3}{2}i)e^{-t} \\ &- (i/2)\sin 2t - 2\cos 2t]^2\}e^{-2it}, \end{aligned} \tag{5.4}$$

and the corresponding Hamilton function is

$$\begin{aligned} H = &\tfrac{1}{2}p_1^2 e^{-2it} + \tfrac{1}{2}p_2^2 e^{2it} \\ &- [(3 + \tfrac{3}{2}i)e^{-t} + i\cos 2t + 4\sin 2t]p_1 \\ &- [(3 - \tfrac{3}{2}i)e^{-t} - i\cos 2t + 4\sin 2t]p_2 \\ &+ \tfrac{3}{2}[X - (3 + \tfrac{3}{2}i)e^{-t} + (i/2)\sin 2t - 2\cos 2t]^2 e^{2it} \\ &+ \tfrac{3}{2}[Y - (3 - \tfrac{3}{2}i)e^{-t} \\ &- (i/2)\sin 2t - 2\cos 2t]^2 e^{-2it}, \end{aligned} \tag{5.5}$$

where $p_1 = \partial L/\partial\dot{X}$ and $p_2 = \partial L/\partial\dot{Y}$ are the canonical momenta.

The Hamilton–Jacobi equation is then

$$\begin{aligned} &\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial X}\right)^2 e^{-2it} + \frac{1}{2}\left(\frac{\partial S}{\partial Y}\right)^2 e^{2it} \\ &- \left[\left(3 + \frac{3}{2}i\right)e^{-t} + i\cos 2t + 4\sin 2t\right]\frac{\partial S}{\partial X} \\ &- \left[\left(3 - \frac{3}{2}i\right)e^{-t} - i\cos 2t + 4\sin 2t\right]\frac{\partial S}{\partial Y} \\ &+ \frac{3}{2}\left[X - \left(3 + \frac{3}{2}i\right)e^{-t} + \frac{i}{2}\sin 2t - 2\cos 2t\right]^2 e^{2it} \\ &+ \frac{3}{2}\left[Y - \left(3 - \frac{3}{2}i\right)e^{-t}\right. \\ &\left. - \frac{i}{2}\sin 2t - 2\cos 2t\right]^2 e^{-2it} = 0. \end{aligned} \tag{5.6}$$

Since this equation has the same structure as Eq. (2.9), we take, analogous to (2.10), a complete solution of (5.6) in the form

$$\begin{aligned} S(t,X,Y) = &\tfrac{1}{2}[X - (3 + \tfrac{3}{2}i)e^{-t} + (i/2)\sin 2t \\ &- 2\cos 2t]^2 e^{2it} f_1(t) + \tfrac{1}{2}[Y - (3 - \tfrac{3}{2}i)e^{-t} \\ &- (i/2)\sin 2t - 2\cos 2t]^2 e^{-2it} f_2(t), \end{aligned} \tag{5.7}$$

where $f_1(t)$ and $f_2(t)$ are the unknown functions of time. This expression will satisfy (5.6) identically for every X, Y, and t if

$$\dot{f}_1 + 2if_1 + f_1^2 + 3 = 0, \quad \dot{f}_2 - 2if_2 + f_2^2 + 3 = 0. \tag{5.8}$$

Hence we find

$$\begin{aligned} X &= B_1 e^{-it}\cos(2t + C_1) + (3 + \tfrac{3}{2}i)e^{-t} \\ &\quad - (i/2)\sin 2t + 2\cos 2t, \\ Y &= B_2 e^{it}\cos(2t + C_2) + (3 - \tfrac{3}{2}i)e^{-t} \\ &\quad + (i/2)\sin 2t + 2\cos 2t. \end{aligned} \tag{5.9}$$

Since, in accordance with (5.3)

$$x = \tfrac{1}{2}(X + Y), \quad y = (-X + Y)/i,$$

we find after matching the initial conditions the solution to (5.1) in the form

$$x = 30\cos t - 15\sin 3t + 3e^{-t} + 2\cos 2t,$$

$$y = 30\cos 3t - 60\sin t - 3e^{-t} + \sin 2t,$$

where we have employed the well-known identity $e^{it} = \cos t + i\sin t$.

VI. CONCLUDING REMARKS

As demonstrated in Sec. II, the Hamilton–Jacobi partial differential equation (2.9) that is based on the "full quadratic" Lagrangian function (2.6), can be easily treated. However, if we base our considerations on the more "obvious" Lagrangian function

$$L_1 = (\tfrac{1}{2}\dot{x}^2 - \tfrac{1}{2}\omega_0^2 x^2 + xh\cos\Omega t)e^{2kt}, \tag{6.1}$$

whose Lagrangian equation (2.3) also generates the correct differential equation (2.1), the process of obtaining a complete solution of the associated Hamilton–Jacobi equation,

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial s}{\partial x}\right)^2 e^{-2kt} + \frac{1}{2}\omega_0^2 x^2 e^{2kt} - hxe^{2kt}\cos\Omega t = 0, \tag{6.2}$$

is considerably more complicated.

We can find a complete solution of (6.2) in the form

$$S = \tfrac{1}{2}[Ax^2 + 2xf(t) + F(t)]e^{2kt}, \tag{6.3}$$

where A is the constant and $f(t)$ and $F(t)$ are unknown functions of time. For example, if $k = 0$, i.e., in the case of a pure force oscillator, a complete solution of (6.2) is given in the following form [suitable for applying the Hamilton–Jacobi theorem (2.13)]:

$$\begin{aligned} S &= \frac{1}{2}\omega_0 x^2 i - \frac{1}{2}xhi\left[\frac{e^{\Omega it}}{\omega_0 + \Omega} + \frac{e^{-\Omega it}}{\omega_0 - \Omega}\right] + Cxe^{-i\omega_0 t} \\ &\quad - \frac{Ch}{2(\omega_0^2 - \Omega^2)}\left[e^{(\Omega - \omega_0)it} + e^{-(\Omega + \omega_0)it}\right] \\ &\quad + \frac{C^2}{4\omega_0 i}e^{-2\omega_0 it} \\ &\quad + (\text{a function of } t \text{ free of } x \text{ and } C). \end{aligned} \tag{6.4}$$

It is easy to verify that the Hamilton–Jacobi theorem produces a correct solution to the forced oscillator problem discussed in Sec. II C.

[1] H. H. Denman and L. H. Buch, "Solution of the Hamilton–Jacobi equation for certain dissipative classical mechanical systems," J. Math. Phys. **14**, 326 (1973).

[2] R. M. Santilli, *Foundations of Theoretical Mechanics I, The Inverse Problem in Newtonian Mechanics* (Springer, New York, 1978).

[3] W. Sarlet, "Symmetries, first integrals and the inverse problem of Lagrangian mechanics," J. Phys. A: Math. Gen. **14**, 2227 (1981).

[4] W. Sarlet, "On linear nonconservative systems derivable from a variational principle," Hadronic J. **3**, 765 (1980).

[5] E. Engels, "A method for the computation of a Lagrangian within the context of the inverse problem of Newtonian mechanics," Hadronic J. **1**, 465, 469 (1978).

[6] M. Crampin, "Constants of motion in Lagrangian mechanics," Int. J. Theor. Phys. **16**, 741 (1977).

[7] F. Cantrijn, "Note on nonconservative dynamical systems," in *Modern Developments in Analytical Mechanics*, edited by S. Benenti, M. Francaviglia, and A. Lichnerowicz (Acta Academie Scientiarum Taurinensis, Turin, 1983), Vol. I, pp. 93–95.

[8] P. Havas, "The range of application of the Lagrangian formalism," Nuovo Cimento **3**, 363 (1957).

[9] L. A. Pars, *A Treatise on Analytical Dynamics* (Heinemann, London, 1968).

[10] B. Vujanovic and A. M. Strauss, "Linear and quadratic first integral of a forced linearly damped oscillator with a single degree of freedom," J. Acoust. Soc. Am. **69**, 1213 (1981).

[11] B. Vujanovic, "A study of conservation laws of dynamical systems by means of the variational principles of Jourdain and Gauss," Acta Mech. **65**, 63 (1986).

[12] B. Vujanovic, A. M. Strauss, and S. E. Jones, "On some conservation laws of conservative and nonconservative dynamical systems," Int. J. Nonlinear Mech. **21**, 489 (1986).

[13] M. R. Spiegel, *Laplace Transforms* (Schaum, New York, 1965).

Hamiltonian structures for systems of hyperbolic conservation laws

Peter J. Olver
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

Yavuz Nutku
Department of Mathematics, Bosphorus University, Bebek, Istanbul, Turkey

(Received 17 February 1988; accepted for publication 2 March 1988)

The bi-Hamiltonian structure for a large class of one-dimensional hyberbolic systems of conservation laws in two field variables, including the equations of gas dynamics, shallow water waves, one-dimensional elastic media, and the Born–Infeld equation from nonlinear electrodynamics, is exhibited. For polytropic gas dynamics, these results lead to a quadri-Hamiltonian structure. New higher-order entropy-flux pairs (conservation laws) and higher-order symmetries are exhibited.

I. HAMILTONIAN SYSTEMS OF CONSERVATION LAWS

With Gardner's discovery of the Hamiltonian structure of the Korteweg–de Vries equation,[1] and Arnol'd's determination of the Lie–Poisson structure underlying the Euler equations of fluid flow,[2] the range of applications of the Hamiltonian formalism embraced truly infinite-dimensional systems. Subsequent progress in the field has been rapid, especially after Magri proved his remarkable theorem on the complete integrability of bi-Hamiltonian systems.[3] This has led to a large number of papers on the applications of Hamiltonian methods to the integrable soliton systems arising as models for nonlinear wave phenomena. More recently, a number of researchers have exploited Arnol'd's original idea to vastly broaden the range of Hamiltonian systems in fluids, plasmas, and relativity.[4] Yet a third branch of infinite-dimensional Hamiltonian mechanics was opened with the investigations of Manin,[5] Cavalcante and McKean,[6] Dubrovin and Novikov,[7] and Nutku,[8] into the Hamiltonian structures arising in first-order quasilinear systems of partial differential equations, known as "systems of hydrodynamic type." These are beginning to have applications to the analysis of hyperbolic conservation laws, and the study of discontinuous (shock wave) solutions.[9–11]

This paper represents a further investigation of the last class of equations, with two-component hyperbolic systems of conservation laws in one spatial variable being the type of system under consideration, and with particular attention being paid to gas dynamics and some model equations arising in elasticity. Our results build on earlier work of Sheftel',[12,13] who was primarily concerned with the symmetry structure of these systems. We connect Sheftel's results with the Hamiltonian framework via Magri's theorem, and deduce the bi-Hamiltonian structure of a broad class of systems of conservation laws, as well as the quadri-Hamiltonian structure of the equations of polytropic gas dynamics themselves. New conservation laws (entropy-flux pairs), substantially extending earlier results of Verosky,[14] are also found. Interesting examples of incompatible bi-Hamiltonian systems are found. Indeed, the fundamental message of this paper is that, in the case of two-dimensional Hamiltonian systems, and particularly for polytropic gas dynamics, nature appears to be overly generous in the amount of structure she provides. Why these systems should be this way remains mysterious. We hope to return to this topic, and to applications of these results to the analysis of smooth solutions and shock waves in a subsequent paper. As a prerequisite for studying this paper, we assume that the reader is familiar with the fundamentals of symmetry groups, Hamiltonian systems of evolution equations, conservation laws, and Magri's theorem, as presented, for instance, in Olver.[15]

The general form of a two-component hyperbolic system of conservation laws of Hamiltonian type is as follows. The unknowns $\mathbf{u}(x,t) = \binom{u(x,t)}{v(x,t)}$ depend on the real-valued temporal and spatial variables t and x. The Hamiltonian functional or energy is

$$\mathscr{H}[\mathbf{u}] = \int H(u,v)\,dx,$$

where the integrand or *Hamiltonian density* $H(u,v)$ is a smooth, but otherwise arbitrary function of u and v. The system takes the explicit form

$$\begin{aligned} u_t &= D_x(H_v) = H_{uv}\cdot u_x + H_{vv}\cdot v_x, \\ v_t &= D_x(H_u) = H_{uu}\cdot u_x + H_{uv}\cdot v_x, \end{aligned} \tag{1.1}$$

with the subscripts denoting partial derivatives. We can also write it in the convenient vector format

$$\mathbf{u}_t = \mathbf{H}\cdot\mathbf{u}_x, \tag{1.2}$$

where

$$\mathbf{H} = \sigma_1\cdot D^2 H = \begin{pmatrix} H_{uv} & H_{vv} \\ H_{uu} & H_{uv} \end{pmatrix}, \tag{1.3}$$

and $\sigma_1 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. The system (1.2) is in the elementary Hamiltonian form

$$\mathbf{u}_t = \mathscr{D}_0 E[H]. \tag{1.4}$$

Here $E = E_{\mathbf{u}}$ denotes the Euler operator or variational derivative, and the Hamiltonian operator is the matrix differential operator

$$\mathscr{D}_0 = \sigma_1\cdot D_x = \begin{pmatrix} 0 & D_x \\ D_x & 0 \end{pmatrix}. \tag{1.5}$$

The corresponding Poisson bracket is

 0022-2488/88/071610-10$02.50

$$\{F,H\} = \int E(F)\cdot \mathscr{D}_0 E(H)\,dx$$

$$= \int \{E_u(F)\cdot D_x E_v(H) + E_v(F) D_x E_u(H)\}dx,$$

which arises in a number of physically important systems.

There are three important examples. The equations of gas dynamics[16]

$$u_t + uu_x + f(v)v_x = 0, \quad v_t + (uv)_x = 0, \tag{1.6}$$

are of the form (1.1), with

$$H(u,v) = -(\tfrac{1}{2}u^2 v + F(v)), \quad \text{where } f(v) = F''(v). \tag{1.7}$$

In physical applications, $u(x,t)$ represents the velocity, and $v(x,t)$ the density (more customarily denoted by ρ) of the fluid; the function $f(v)$ is related to the physical pressure $P(v)$ according to the equation $f(v) = P'(v)/v$. The equations of *polytropic gas dynamics* correspond to the choice $F(v) = v^\gamma/\gamma(\gamma-1)$ for some exponent $\gamma \neq 0,1$. In particular, the special case $F(v) = \frac{1}{2}v^2$ also arises in shallow water theory.[6] A second important example is provided by the Hamiltonian

$$H(u,v) = u/v + v/u, \tag{1.8}$$

in which case (1.1) is equivalent to the Born–Infeld equation from nonlinear electrodynamics.[17,18] Finally, we mention the simple model

$$u_t = [\sigma(v)]_x, \quad v_t = u_x, \tag{1.9}$$

for a one-dimensional nonlinear elastic medium.[9,10,19] In this case, the Hamiltonian density is

$$H(u,v) = \tfrac{1}{2}u^2 + F(v), \tag{1.10}$$

where $F'(v) = \sigma(v)$. The case when σ is a monotone function of v corresponds to an ideal fluid or elastic solid; models of phase transitions arise with more general functions σ, the most interesting being that of a van der Waals fluid, where $\sigma(v) = c/(v-b) - a/v^2$. Furthermore, the case $\sigma(v) = (1+v)^{-(1+\gamma)}$ corresponds to the Euler equation arising in nonlinear acoustics.[20]

II. A PROTOTYPE: THE RIEMANN EQUATION

A remarkable feature of the two-component Hamiltonian system (1.1) is that, suitably interpreted, all the structure already manifests itself in the simple scalar nonlinear wave equation

$$u_t = uu_x. \tag{2.1}$$

This equation has many names (Riemann, inviscid Burgers', Hopf, etc.), and serves as a prototype for so many of the phenomena associated with hyperbolic systems, that its role in the Hamiltonian structure and conservation laws will not be surprising. See Whitham[16] for a good review of the classical theory of this equation.

The purpose of this introductory section is to list some basic results for the Riemann equation (2.1), all of which have direct counterparts in the equations for polytropic gas dynamics (1.6) and the model elasticity equations (1.9), and many of which have counterparts for more general two-dimensional hyperbolic systems. The rather easy proofs are left to the reader, as they serve as good warm-up exercises for the much harder calculations in two dimensions.

(1) There is an infinite sequence of zeroth-order conserved densities

$$H_n(u) = u^n, \quad n = 1,2,3,\ldots. \tag{2.2}$$

In fact, any function $F(u)$ is a conserved density for (2.1). [By the *order* of a function $F[u] = F(u,u_x,\ldots)$, we mean the highest order derivative of u on which it depends.]

(2) The equation admits three first-order Hamiltonian operators

$$\mathscr{D}_0 = D_x, \quad \mathscr{D}_1 = 2u\cdot D_x + u_x, \quad \mathscr{D}_2 = u^2\cdot D_x + uu_x, \tag{2.3}$$

and so can be written in Hamiltonian form in three distinct ways:

$$u_t = \mathscr{D}_0\cdot E(\tfrac{1}{6}H_3) = \mathscr{D}_1\cdot E(\tfrac{1}{6}H_2) = \mathscr{D}_2\cdot E(H_1).$$

Moreover, since

$$\mathscr{D}_2 \neq \mathscr{D}_1\cdot\mathscr{D}_0^{-1}\cdot\mathscr{D}_1,$$

these Hamiltonian operators are not trivially related.[21]

(3) The three Hamiltonian operators $\mathscr{D}_0, \mathscr{D}_1, \mathscr{D}_2$, are compatible, meaning that any two of them form a Hamiltonian pair in the sense of Magri's theorem.[3,22] Therefore, Eq. (2.1) is a "tri-Hamiltonian system." The resulting recursion operators

$$\mathscr{R}_1 = \mathscr{D}_1\cdot\mathscr{D}_0^{-1}, \quad \mathscr{R}_2 = \mathscr{D}_2\cdot\mathscr{D}_0^{-1}, \quad \mathscr{R}_3 = \mathscr{D}_2\cdot\mathscr{D}_1^{-1}, \tag{2.4}$$

which map symmetries to symmetries, are trivially related by

$$\mathscr{R}_2 = \mathscr{R}_3\cdot\mathscr{R}_1,$$

but are otherwise independent "integrodifferential operators."

(4) There is an infinite sequence of commuting Hamiltonian flows corresponding to the evolutionary vector fields $\mathbf{v}_n = Q_n[u]\cdot\partial_u$, with characteristics $Q_n[u] = u^n\cdot u_x$. The nth flow can be written in Hamiltonian form in three distinct ways:

$$\begin{aligned} u_t = Q_n[u] &= u^n\cdot u_x \\ &= \mathscr{D}_0\cdot E([(n+2)(n+1)]^{-1}H_{n+2}) \\ &= \mathscr{D}_1\cdot E([(2n+1)(n+1)]^{-1}H_{n+1}) \\ &= \mathscr{D}_2\cdot E(n^{-2}H_n), \end{aligned} \tag{2.5}$$

whose Hamiltonians H_n are given by (2.2). These Hamiltonians are in involution with respect to any of the three Poisson brackets determined by the Hamiltonian operators:

$$\{H_n,H_m\}_j = \int E(H_n)\cdot\mathscr{D}_j E(H_m)\,dx = 0, \quad j = 0,1,2.$$

The recursion operators (2.4) reproduce the hierarchy of Hamiltonian flows,

$$\mathscr{R}_1(Q_n) = [(2n+3)/(n+1)]Q_{n+1},$$

$$\mathscr{R}_2(Q_n) = [(n+2)/(n+1)]Q_{n+2}.$$

Thus, starting with the initial flow (2.1), $\mathscr{R}_1$ produces every flow in the hierarchy (2.5), whereas $\mathscr{R}_2$ produces every oth-

er one. Interestingly, even though there are two independent recursion operators, the two hierarchies of zeroth-order conserved densities and corresponding first-order symmetries of (2.1) guaranteed by Magri's theorem happen to coincide. However, this remark is special to the polynomial hierarchy (2.2); on other hierarchies of Hamiltonian densities $H_n(u)$ defined by the recursion relation, starting with a nonpolynomial function $H_0(u)$, these recursion operators will act differently.

(5) The equation admits an additional third-order Hamiltonian operator

$$\mathscr{E} = D_x \cdot (1/u_x) \cdot D_x \cdot (1/u_x) \cdot D_x. \tag{2.6}$$

Each of the Hamiltonian systems (2.5) can be written in this new Hamiltonian form using the Hamiltonian operator $\mathscr{E}$,

$$u_t = u^n u_x = \mathscr{E} \cdot E([(n+1)(n+2)(n+3)(n+4)]^{-1} H_{n+4}).$$

(6) The Hamiltonian operators $\mathscr{D}_0$ and $\mathscr{E}$ are compatible; however, $\mathscr{E}$ is not compatible with either $\mathscr{D}_1$ or $\mathscr{D}_2$. Thus, the Riemann equation, and its higher degree generalizations (2.5) are in the rather anomalous position of being "quadri-Hamiltonian systems," meaning there are four distinct local Hamiltonian structures, but of the six different possible pairs of Hamiltonian operators, four are compatible, whereas two are not!

(7) The recursion operator

$$\hat{\mathscr{R}} = \mathscr{E} \cdot \mathscr{D}_0^{-1} = D_x \cdot (1/u_x) \cdot D_x \cdot (1/u_x),$$

is the square of a simpler first-order recursion operator

$$\mathscr{R} = D_x \cdot (1/u_x). \tag{2.7}$$

This latter operator acts on the first-order flows according to

$$\mathscr{R}(Q_n) = nQ_{n-1},$$

and so, up to multiple, "inverts" the first-order recursion operator $\mathscr{R}_1$. [Again, as in part (4), this is special to the polynomial hierarchy (2.5).]

(8) There is a rational second-order generalized symmetry $\hat{\mathbf{v}}_2 = \hat{Q}_2 \cdot \partial_u = u_x^{-2} \cdot u_{xx} \cdot \partial_u$ with corresponding flow governed by the evolution equation

$$u_t = \hat{Q}_2 = u_{xx}/u_x^2.$$

Thus the recursion operator $\mathscr{R}$ generates an additional hierarchy of higher order symmetries $\hat{\mathbf{v}}_n = \hat{Q}_n \cdot \partial_u$, where

$$\hat{Q}_{n+1} = \mathscr{R}(\hat{Q}_n), \quad n = 2,3,\ldots . \tag{2.8}$$

(9) The third-order generalized symmetry $\hat{\mathbf{v}}_3$ in the hierarchy (2.8) is Hamiltonian relative to the first-order Hamiltonian operator $\mathscr{D}_0$. Thus there is a rational first-order conserved density

$$\hat{H}_1 = 1/u_x$$

such that the third-order evolution equation

$$u_t = \hat{Q}_3 = (u_x \cdot u_{xxx} - 3u_{xx}^2)/u_x^4 \tag{2.9}$$

corresponding to $\hat{\mathbf{v}}_3$ is in fact bi-Hamiltonian

$$u_t = \mathscr{D}_0 \cdot E(\hat{H}_1) = \mathscr{E} \cdot E(\hat{H}_0).$$

Here

$$\hat{H}_0 = xu + \tfrac{1}{2}tu^2$$

is a (time-dependent) zeroth-order conserved density for (2.1). Consequently, every other member of the hierarchy (2.8), i.e., the vector fields $\hat{\mathbf{v}}_{2n+1}$, is bi-Hamiltonian, corresponding to a hierarchy of higher-order rational conserved densities $\hat{H}_n$, $n = 0,1,2,\ldots$.

(10) For any solution to the general first-order flow (2.5) corresponding to the vector field $\mathbf{v}_n$, each of the higher-order quantities $\int \hat{H}_m[u]dx$ is a linear function of t (provided that the integral converges). In fact, if $H(u)$ is any zeroth-order conserved density, with flow $u_t = D_x E(H)$, then

$$\hat{H}_m + t \frac{\partial^{2m+1} H}{\partial u^{2m+1}}$$

is a conserved density for such a flow. In particular, $\hat{H}_m$ is a conserved density for $\mathbf{v}_n$ whenever $2m \geqslant n-1$.

There are some additional facts concerning higher order symmetries and conservation laws for (2.1) which do *not* appear to have counterparts for the equations of gas dynamics. [This demonstrates that one must be careful when deciding which of the properties of (2.1) one wishes to generalize to two-dimensional systems in general, and gas dynamics in particular.] The proof of the following complete characterization of symmetries and conservation laws is not difficult.

Theorem 2.1: Define the rational differential functions

$$K_0 = u, \quad K_1 = x - tu, \quad K_2 = u/u_x - x, \quad K_3 = u_{xx}/u_x^3,$$
$$K_{n+1} = (1/u_x) D_x K_n, \quad n \geqslant 3.$$

(i) $T[u]$ is conserved density of (2.1) if and only if

$$T = u_x \cdot F(K_0, K_1, K_2, \ldots, K_n),$$

where F is an arbitrary smooth function of the differential functions $K_0,\ldots,K_n$.

(ii) $Q[u]$ is the characteristic of a generalized symmetry $\mathbf{v}_Q = Q \cdot \partial_u$ of (2.1) if and only if

$$Q = u_x \cdot G(K_0, K_1, K_2, \ldots, K_n),$$

where G is an arbitrary smooth function of the functions $K_0,\ldots,K_n$.

III. ZEROTH-ORDER CONSERVED DENSITIES

We now return to the general two-component Hamiltonian system (1.1), and investigate its symmetries, conserved densities, and Hamiltonian structures. We will follow the basic outline of the properties for the elementary one-dimensional wave equation (2.1) discussed in Sec. II. We therefore begin by discussing item (1), the existence of zeroth-order conserved densities. Much of this material is well-known.[5,13,23]

Associated with each hyperbolic system (1.1) is a second-order linear partial differential equation

$$A(u,v)F_{uu} = B(u,v)F_{vv}, \tag{3.1}$$

with $A = H_{vv}$, $B = H_{uu}$. The solutions of (3.1) are the zeroth-order conserved densities (entropies) for the hyperbolic system (1.1).[10,13]

Proposition 3.1: A functional $\mathscr{F}[u] = \int F(u,v)dx$ is conserved for the Hamiltonian system (1.1) if and only if F is a solution to Eq. (3.1).

Note that H itself is always a solution to (3.1). Also, if F

is a conserved density for the Hamiltonian system governed by H, then, reciprocally, H is a conserved density for the Hamiltonian system governed by F. In addition, we always have four elementary solutions given by the functions

$$1,\ u,\ v,\ u\cdot v. \tag{3.2}$$

These are conserved densities for *any* Hamiltonian system of the form (1.1).

The most important class of Hamiltonian systems (1.1) are those for which the corresponding partial differential equation (3.1) admits a separation of variables in the rectangular (u,v) coordinates.

Definition 3.2: A Hamiltonian density $H(u,v)$ is called *separable* if (a) the second-order partial derivatives H_{uu} and H_{vv} do not vanish identically, and (b) there exist functions $\lambda(u)$ and $\mu(v)$ such that

$$H_{uu}/H_{vv} = \lambda(u)/\mu(v). \tag{3.3}$$

If the Hamiltonian density is separable, then Eq. (3.1) determining the zeroth-order conserved densities takes the form

$$F_{uu}/\lambda(u) = F_{vv}/\mu(v). \tag{3.4}$$

In this paper, we will deal exclusively with separable Hamiltonian systems (1.1); an interesting open question is how many of these results can be generalized to the nonseparable cases. Gas dynamics, (1.6), is an example of a separable system, where

$$\lambda(u)\equiv 1 \text{ and } \mu(v) = \frac{F''(v)}{v} = \frac{f(v)}{v} = \frac{P'(v)}{v^2}.$$

In fact, the special case when $\lambda\equiv 1$ has added importance.

Definition 3.3: Let $H(u,v)$ be a Hamiltonian density. If the ratio $H_{uu}/H_{vv} = \mu(v)$ is a function of v alone, then H is said to be a *generalized gas dynamics Hamiltonian density.*

The elastic models (1.9) are of gas dynamics type, with $\mu(v) = \sigma'(v)$. The Born–Infeld Hamiltonian (1.8) is separable, with $\lambda(u) = u^{-4}$ and $\mu(v) = v^{-4}$, but not of gas dynamics type. [However, Verosky[24] noticed that the transformation

$$\tilde{u} = -(u^{-2}+v^{-2}),\quad \tilde{v} = \tfrac{1}{2}uv, \tag{3.5}$$

will change the Born–Infeld system into a polytropic gas dynamics system with adiabatic index $\gamma = -1$, i.e., $F(v) = 1/2v$. This remarkable transformation will be discussed in more detail in a forthcoming publication.[17]]

In the separable case, there are four fundamental hierarchies of solutions to the wave equation (3.4). Each of them takes the form

$$H_n(u,v) = \sum_{i=0}^{n} F_i(u)\cdot G_{n-i}(v),$$

where the functions F_i and G_i are generated by the recursion relations

$$\frac{\partial^2 F_i}{\partial u^2} = \lambda(u)F_{i-1},\quad \frac{\partial^2 G_i}{\partial v^2} = \mu(v)G_{i-1}, \tag{3.6}$$

and normalized so that

$$F_i(0) = F_i'(0) = 0,\quad G_i(0) = G_i'(0) = 0.$$

The hierarchies depend on the initial selection of $H_0 = F_0\cdot G_0$, and there are four obvious possibilities, given by the four elementary conserved densities (3.2):

$$H_0 = 1,\quad F_0 = G_0 = 1,$$
$$H_0 = u,\quad F_0 = u,\quad G_0 = 1,$$
$$H_0 = v,\quad F_0 = 1,\quad G_0 = v,$$
$$H_0 = uv,\quad F_0 = u,\quad G_0 = v.$$

In practice, it is expedient to combine the first and second, and the third and fourth hierarchies, leading to two fundamental hierarchies of conserved densities, which we denote by H_n and $\tilde{H}_n$, respectively, so that H_{2n} is the nth member of the first hierarchy, H_{2n+1} is the nth member of the second hierarchy, $\tilde{H}_{2n}$ is the nth member of the third hierarchy, and $\tilde{H}_{2n+1}$ is the nth member of the fourth hierarchy.

For reference, we list the first few members of each sequence in the gas dynamics case $\lambda\equiv 1$:

$$\begin{aligned}
&H_0 = 1, && \tilde{H}_0 = v,\\
&H_1 = u, && \tilde{H}_1 = uv,\\
&H_2 = \tfrac{1}{2}u^2 + G_1(v), && \tilde{H}_2 = \tfrac{1}{2}u^2v + \tilde{G}_1(v),\\
&H_3 = \tfrac{1}{6}u^3 + uG_1(v), && \tilde{H}_3 = \tfrac{1}{6}u^3v + u\tilde{G}_1(v),\\
&H_4 = \tfrac{1}{24}u^4 + \tfrac{1}{2}u^2G_1(v) + G_2(v), &&\\
&\tilde{H}_4 = \tfrac{1}{24}u^4v + u^2\tilde{G}_1(v) + \tilde{G}_2(v), &&
\end{aligned} \tag{3.7}$$

etc., where, according to (3.6),

$$G_1(v) = \int_0^v (v-w)\cdot\mu(w)dw,$$
$$G_2(v) = \int_0^v (v-w)\cdot\mu(w)\cdot G_1(w)dw,$$
$$\tilde{G}_1(v) = \int_0^v (v-w)\cdot w\cdot\mu(w)dw,$$
$$\tilde{G}_2(v) = \int_0^v (v-w)\cdot\mu(w)\cdot\tilde{G}_1(w)dw.$$

Note that the elastic Hamiltonian (1.10) appears in the first hierarchy as H_2, whereas the gas dynamics Hamiltonian (1.7) appears in the alternative hierarchy as $-\tilde{H}_2$.

Lemma 3.4: Let $H(u,v)$ be a generalized gas dynamics Hamiltonian. If $F(u,v)$ is a conserved density, then so is $\partial F/\partial u$.

In fact, the hierarchies (3.7) satisfy

$$\frac{\partial H_n}{\partial u} = H_{n-1}, \tag{3.8}$$

and similarly for $\tilde{H}_n$.

For a general separable system, each of the Hamiltonian functions generates a Hamiltonian flow, governed by the corresponding evolutionary system (1.1). We let

$$\mathbf{Q}_n = \mathscr{D}_0 E[H_n] = \mathbf{H}_n\cdot\mathbf{u}_x \tag{3.9}$$

[cf. (1.3) and (1.5)] denote the right-hand side of this equation, which is also the characteristic for the symmetry vector field $\mathbf{v}_n = \mathbf{Q}_n\cdot\partial_{\mathbf{u}}$. We define $\tilde{\mathbf{H}}_n$, $\tilde{\mathbf{Q}}_n$, and $\tilde{\mathbf{v}}_n$ from the alternative hierarchy $\tilde{H}_n$ similarly. In particular, $\mathbf{Q}_0 = \mathbf{Q}_1 = \tilde{\mathbf{Q}}_0 = 0$, meaning that the densities H_0, H_1, $\tilde{H}_0$ determine distinguished functionals (Casimirs) for the Hamiltonian structure (1.5), while $\tilde{\mathbf{Q}}_1 = \mathbf{u}_x$ is the common translational symmetry of the systems. All the Hamiltonians H_n and $\tilde{H}_n$

are in involution with respect to the Poisson bracket determined by the Hamiltonian operator (1.5), i.e.,

$$\{H_n, H_m\} = \{H_n, \tilde{H}_m\} = \{\tilde{H}_n, \tilde{H}_m\} = 0,$$

for all n, m. Consequently, the vector fields $\mathbf{v}_n$ and $\tilde{\mathbf{v}}_m$ all mutually commute.

IV. FIRST-ORDER HAMILTONIAN STRUCTURES: POLYTROPIC GAS DYNAMICS

We have already seen the first Hamiltonian structure of the gas dynamics system (1.6). We now turn to a discussion of the other first-order Hamiltonian structures in the case of polytropic gas dynamics. These results correspond to items (2)–(4) for the one-dimensional equation (2.1). Strangely, these results do not seem to generalize to the general gas dynamics system, much less the more general hyperbolic system (1.1).

Nutku[8] has shown that, in the case of polytropic gas dynamics, there are two additional first-order Hamiltonian structures. Using the Hamiltonian hierarchies (3.7), we find that we can write the polytropic gas dynamics equations (1.6), with $f(v) = v^{\gamma-2}$, in the alternative Hamiltonian form

$$u_t = \mathscr{D}_1 \cdot E(\gamma^{-1}\tilde{H}_1), \tag{4.1}$$

where $\mathscr{D}_1$ is the Hamiltonian operator

$$\mathscr{D}_1 = \begin{pmatrix} v^{\gamma-2}\cdot D_x + D_x \cdot v^{\gamma-2} & (\gamma-1)u\cdot D_x + u_x \\ (\gamma-1)u\cdot D_x + (\gamma-2)u_x & v\cdot D_x + D_x \cdot v \end{pmatrix}.$$

They have yet another Hamiltonian form,

$$u_t = \mathscr{D}_2 \cdot E(\tilde{H}_0), \tag{4.2}$$

where

$$\mathscr{D}_2 = \begin{pmatrix} uv^{\gamma-2}\cdot D_x + D_x \cdot uv^{\gamma-2} & \{\frac{1}{2}(\gamma-1)u^2 + 2[v^{\gamma-1}/(\gamma-1)]\}\cdot D_x \\ & \quad + uu_x + v^{\gamma-2}v_x \\ \{\frac{1}{2}(\gamma-1)u^2 + 2[v^{\gamma-1}/(\gamma-1)]\}\cdot D_x & uv\cdot D_x + D_x \cdot uv \\ \quad + (\gamma-2)uu_x + v^{\gamma-2}v_x & \end{pmatrix}.$$

The Hamiltonian operators $\mathscr{D}_0$, $\mathscr{D}_1$, $\mathscr{D}_2$ are mutually compatible, leading to three distinct Hamiltonian pairs. The corresponding recursion operators are just as in (2.4), i.e.,

$$\mathscr{R}_1 = \mathscr{D}_1 \cdot \mathscr{D}_0^{-1}, \quad \mathscr{R}_2 = \mathscr{D}_2 \cdot \mathscr{D}_0^{-1}, \quad \mathscr{R}_3 = \mathscr{D}_2 \cdot \mathscr{D}_1^{-1}, \tag{4.3}$$

and, as before, are trivially related by the identity $\mathscr{R}_2 = \mathscr{R}_3 \cdot \mathscr{R}_1$, but are otherwise distinct. Nevertheless, they both give rise to the same series of gas dynamics Hamiltonians, since

$$\mathscr{R}_1(\mathbf{Q}_n) = \mathbf{Q}_{n+1}, \quad \mathscr{R}_2(\mathbf{Q}_n) = \mathbf{Q}_{n+2},$$

and similarly for the alternative hierarchy $\tilde{\mathbf{v}}_n$. In other words, just as with the one-dimensional prototype (2.1), $\mathscr{R}_2$ leads to every other member of the hierarchies (3.9), and so for some reason $\mathscr{R}_3$ acts the same as $\mathscr{R}_1$ on these hierarchies. (As with the Riemann equation recursion operators, these recursion operators will act differently on other hierarchies.) Strangely, there does not appear to be a counterpart of these two recursion operators in the general nonpolytropic case, i.e., when the pressure is not proportional to a power of the density. We do not fully understand why this should be the case.

A similar situation exists for the "polytropic" versions of the elasticity models (1.9), i.e., when $F(v) = v^{\gamma-1}/(\gamma-2)\cdot(\gamma-1)$, $\gamma \neq 0,1,2$. Here we can write it in a second Hamiltonian form

$$u_t = \mathscr{D}_1 \cdot E((\gamma-2)^{-1}H_1). \tag{4.4}$$

However, even though this system cannot be written in Hamiltonian form using the operator $\mathscr{D}_2$, nevertheless it still admits all three recursion operators (4.3).

The only other case of (1.9) which appears to admit a second Hamiltonian structure is the curious choice $\sigma(v) = e^v$. Here we have the Hamiltonian operator

$$\mathscr{D}_1 = \begin{pmatrix} u\cdot e^v \cdot D_x + D_x \cdot u \cdot e^v & (e^v + \frac{1}{2}u^2)\cdot D_x + D_x \cdot e^v \\ e^v \cdot D_x + D_x \cdot (e^v + \frac{1}{2}u^2) & u\cdot D_x + D_x \cdot u \end{pmatrix},$$

and the equation can be written as

$$u_t = \mathscr{D}_1 \cdot E(\tilde{H}_0).$$

Interestingly, the resulting recursion operator $\mathscr{R}_1 = \mathscr{D}_1 \cdot \mathscr{D}_0^{-1}$ alternates between the two hierarchies of conserved densities (3.7), i.e.,

$$\mathscr{R}_1(\mathbf{Q}_n) = \tilde{\mathbf{Q}}_{n+2}, \quad \mathscr{R}_1(\tilde{\mathbf{Q}}_n) = \mathbf{Q}_{n+2}.$$

Note that this also implies that the gas dynamics equations (1.6), with $f(v) = e^v$, can also be written in the Hamiltonian form $u_t = \mathscr{D}_1 \cdot E(H_1)$. Apart from obvious rescalings and translations of v, this appears to be the only nontrivial case where this happens.

V. HIGHER-ORDER HAMILTONIAN STRUCTURE

In this section, we present the second Hamiltonian structure of a general separable Hamiltonian system (1.2). We begin by exhibiting the analog of the third-order Hamiltonian operator (2.6), thereby finding analogies to items (5) and (6) of Sec. II. We will then show how to write the system itself in bi-Hamiltonian form.

Following Sheftel',[13] define the functions

$$L(u) = \int_0^u \lambda(s)ds, \quad M(v) = \int_0^v \mu(s)ds,$$

and the matrix variables

$$U(u,v)=\begin{pmatrix} u & M(v) \\ v & L(u) \end{pmatrix}, \quad V(u,v)=\begin{pmatrix} L(u) & M(v) \\ v & u \end{pmatrix}. \tag{5.1}$$

Note that $\sigma_1 \cdot U = V^T \cdot \sigma_1$, where $\sigma_1 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Let

$$U_x=\begin{pmatrix} u_x & \mu(v)v_x \\ v_x & \lambda(u)u_x \end{pmatrix} \quad \text{and} \quad U_t=\begin{pmatrix} u_t & \mu(v)v_t \\ v_t & \lambda(u)u_t \end{pmatrix}$$

denote the total x and t derivatives of the matrix U; define V_x and V_t analogously. Also let

$$U_x^{-1}=\frac{1}{\delta}\begin{pmatrix} \lambda(u)u_x & -\mu(v)v_x \\ -v_x & u_x \end{pmatrix},$$

$$V_x^{-1}=\frac{1}{\delta}\begin{pmatrix} u_x & -\mu(v)v_x \\ -v_x & \lambda(u)u_x \end{pmatrix},$$

$$\delta=\lambda(u)\cdot u_x^2-\mu(v)\cdot v_x^2,$$

denote the matrix inverses of U_x and V_x.

The basic third-order Hamiltonian operator for a general separable Hamiltonian system can be written compactly as follows.

Theorem 5.1: The operator

$$\begin{aligned} \mathscr{E} &= D_x \cdot V_x^{-1} \cdot D_x \cdot U_x^{-1} \cdot \sigma_1 \cdot D_x \\ &= D_x \cdot V_x^{-1} \cdot D_x \cdot \sigma_1 \cdot V_x^{-T} \cdot D_x \end{aligned} \tag{5.2}$$

is Hamiltonian.

Proof: The fact that $\mathscr{E}$ is a skew-adjoint differential operator follows immediately from the second formula in (5.2). Rather than prove the Poisson form[22] of the Jacobi identity for $\mathscr{E}$, it is simpler to prove that the symplectic two-form

$$\Omega=\int \{d\mathbf{u}^T \wedge \mathscr{E}^{-1}\, d\mathbf{u}\}dx,$$

is closed[25]: $d\Omega = 0$. Let φ, ψ denote potential functions for u, v, respectively, so

$$\varphi_x = u, \quad \psi_x = v.$$

Thus, formally,

$$D_x^{-1}(d\varphi) = du, \quad D_x^{-1}(d\psi) = dv,$$

and hence

$$\begin{aligned} \Omega &= \int \{(v_x\, d\varphi + u_x\, d\psi) \\ &\qquad \wedge D_x^{-1}(\lambda(u)u_x\, d\varphi + \mu(v)v_x\, d\psi)\}dx \\ &= \int \{(\psi_{xx}\, d\varphi + \varphi_{xx}\, d\psi) \\ &\qquad \wedge D_x^{-1}(\lambda(\varphi_x)\varphi_{xx}\, d\varphi + \mu(\psi_x)\psi_{xx}\, d\psi)\}dx. \end{aligned}$$

Let L, M be as in (5.1), and let

$$\hat{L}(u)=\int_0^u s\cdot\lambda(s)ds,$$

$$\tilde{L}(u)=\int_0^u (u-s)\lambda(s)ds = uL(u)-\hat{L}(u),$$

$$\hat{M}(u)=\int_0^u s\cdot\mu(s)ds,$$

$$\tilde{M}(u)=\int_0^u (u-s)\mu(s)ds = uM(u)-\hat{M}(u).$$

Performing an inspired series of integration by parts, we find

$$\begin{aligned} \Omega &= -\int \{(\psi_x\, d\varphi + \varphi_x\, d\psi)\wedge(\lambda(\varphi_x)\varphi_{xx}\, d\varphi + \mu(\psi_x)\psi_{xx}\, d\psi) \\ &\qquad + d(\varphi_x\psi_x)\wedge D_x^{-1}(\lambda(\varphi_x)\varphi_{xx}\, d\varphi + \mu(\psi_x)\,\psi_{xx}\, d\psi)\}dx \\ &= \int \{(\varphi_x\lambda(\varphi_x)\varphi_{xx} - \psi_x\mu(\psi_x)\psi_{xx})d\varphi\wedge d\psi \\ &\qquad - d(\varphi_x\psi_x)\wedge(L(\varphi_x)d\varphi + M(\psi_x)d\psi) + d(\varphi_x\psi_x)\wedge D_x^{-1}d(\tilde{L}(\varphi_x)+\tilde{M}(\psi_x))\}dx, \\ &= \int \{(-\hat{L}(\varphi_x)+\hat{M}(\psi_x))(d\varphi_x\wedge d\psi + d\varphi\wedge d\psi_x) \\ &\qquad - d(\varphi_x\psi_x)\wedge(L(\varphi_x)d\varphi + M(\psi_x)d\psi) + d(\varphi_x\psi_x)\wedge D_x^{-1}d(\tilde{L}(\varphi_x)+\tilde{M}(\psi_x))\}dx. \end{aligned}$$

Therefore,

$$\begin{aligned} d\Omega = \int \{ &-\varphi_x\lambda(\varphi_x)d\varphi_x\wedge d\varphi\wedge d\psi_x \\ &+\psi_x\mu(\psi_x)d\psi_x\wedge d\varphi_x\wedge d\psi \\ &+ d(\varphi_x\psi_x)\wedge(\lambda(\varphi_x)d\varphi_x\wedge d\varphi \\ &+\mu(\psi_x)d\psi_x\wedge d\psi)\}dx, \end{aligned}$$

which clearly vanishes. This completes the proof that Ω is closed, and hence $\mathscr{E}$ is a Hamiltonian operator.

Proposition 5.2: The operators $\mathscr{E}$ and $\mathscr{D}_0$ form a Hamiltonian pair.

Proof: Here it is more convenient to use Poisson methods.[22] Let $\theta = \left(\begin{smallmatrix} \xi \\ \eta \end{smallmatrix}\right)$ be the basis univectors associated with the variables $\mathbf{u} = \left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$. Since we already know that $\mathscr{E}$ and $\mathscr{D}_0$ are Hamiltonian operators, we need only verify the additional compatibility condition[26]

$$\operatorname{pr} \mathbf{v}_{\mathscr{D}_0\theta}(\Theta_{\mathscr{E}}) + \operatorname{pr} \mathbf{v}_{\mathscr{E}\theta}(\Theta_{\mathscr{D}_0}) = 0, \tag{5.3}$$

where

$$\Theta_{\mathscr{D}_0}=\frac{1}{2}\int \theta^T\wedge\mathscr{D}_0(\theta)dx=\int \xi\wedge\eta_x\, dx$$

is the functional bivector associated with the Hamiltonian operator $\mathscr{D}_0$, and

$$\Theta_{\mathscr{E}} = \frac{1}{2}\int \theta^T \wedge \mathscr{E}(\theta)dx$$

$$= \frac{1}{2}\int [V_x^{-T}\theta]^T \wedge \sigma_1 \cdot D_x [V_x^{-T}\theta]dx$$

is the bivector associated with the Hamiltonian operator $\mathscr{E}$. Now $\Theta_{\mathscr{D}_0}$ is a constant coefficient bivector, and hence

$$\operatorname{pr} \mathbf{v}_{\mathscr{E}\theta}(\Theta_{\mathscr{D}_0}) = 0,$$

automatically. Thus, to verify (5.3), we need only compute

$$\operatorname{pr} \mathbf{v}_{\mathscr{D}_0\theta}(\Theta_{\mathscr{E}})$$

$$= \int [\operatorname{pr} \mathbf{v}_{\mathscr{D}_0\theta}(V_x^{-T})\cdot\theta]^T \wedge \sigma_1 \cdot D_x [V_x^{-T}\cdot\theta]dx,$$

$$= -\int [V_x^{-T}\cdot \operatorname{pr} \mathbf{v}_{\mathscr{D}_0\theta}(V_x^T)\cdot V_x^{-T}\cdot\theta]^T$$

$$\wedge \sigma_1 \cdot D_x [V_x^{-T}\cdot\theta]dx, \tag{5.4}$$

where

$$\operatorname{pr} \mathbf{v}_{\mathscr{D}_0\theta}(V_x^T)$$

$$= \operatorname{pr}\left\{\eta_x \frac{\partial}{\partial u} + \xi_x \frac{\partial}{\partial v}\right\}\begin{pmatrix}\lambda(u)\cdot u_x & v_x \\ \mu(v)\cdot v_x & u_x\end{pmatrix}$$

$$= \begin{pmatrix}\lambda(u)\cdot\eta_{xx} + \lambda'(u)\cdot u_x\cdot\eta_x & \eta_{xx} \\ \mu(v)\cdot\xi_{xx} + \mu'(v)\cdot v_x\cdot\xi_x & \xi_{xx}\end{pmatrix}.$$

It now appears that the only recourse is a rather long calculation to check that the functional trivector (5.4) vanishes. We will not reproduce this calculation here, but just remark that, after several integrations by parts, the final result only depends on the four basis trivectors

$$\begin{aligned}&\xi_x \wedge \eta_x \wedge \xi_{xx}, \quad \xi_x \wedge \eta_x \wedge \eta_{xx}, \\ &\xi_x \wedge \xi_{xx} \wedge \eta_{xx}, \quad \eta_x \wedge \xi_{xx} \wedge \eta_{xx}.\end{aligned} \tag{5.5}$$

Thus one need only check that the coefficient of each of these trivectors is zero, an exercise we leave to the reader.

We now show how to combine the two Hamiltonian operators $\mathscr{E}$ and $\mathscr{D}_0$ to make a separable Hamiltonian system bi-Hamiltonian.

Theorem 5.3: Let $H(u,v)$ be a separable Hamiltonian density. Then there exists a second zeroth-order conserved density $H^*(u,v)$ such that the corresponding Hamiltonian system (1.1) can be written in bi-Hamiltonian form

$$\mathbf{u}_t = \mathscr{D}_0 E[H] = \mathscr{E} E[H^*]. \tag{5.6}$$

Proof: Here, all the calculations are local, i.e., over a suitably small domain in (u,v) space. Let $H^*(u,v)$ be a separable density satisfying the same equation (3.3) as H. A straightforward calculation using (3.3) shows that

$$\mathscr{E} E[H^*] = D_x \begin{pmatrix} H^*_{uuv}/\lambda \\ H^*_{uvv}/\mu \end{pmatrix}.$$

This will coincide with (1.1) provided H^* satisfies

$$H_v = H^*_{uuv}/\lambda, \quad H_u = H^*_{uvv}/\mu. \tag{5.7}$$

Note first that, given H^*, we can always determine a corresponding Hamiltonian density H, satisfying (3.3), since the compatibility conditions for (5.7), i.e.,

$$\frac{\partial^2}{\partial u\,\partial v}\left(\frac{H^*_{uu}}{\lambda}\right) = \frac{\partial}{\partial u}\left(\frac{H^*_{uuv}}{\lambda}\right)$$

$$= \frac{\partial}{\partial v}\left(\frac{H^*_{uvv}}{\mu}\right) = \frac{\partial^2}{\partial u\,\partial v}\left(\frac{H^*_{vv}}{\mu}\right), \tag{5.8}$$

are clearly satisfied since H^* solves (3.3). Conversely, given $H(u,v)$, which satisfies (3.3), we need to show that there exists a function $H^*(u,v)$, which also satisfies (3.3), and satisfies (5.7). To accomplish this, we first determine a function $G(u,v)$ which satisfies

$$G_u = \lambda H_v, \quad G_v = \mu H_u.$$

This is possible (locally) by virtue of (3.3). Further, let $F(u,v)$ be any function such that

$$F_{uv} = G.$$

Now according to (5.7), we will have $H^*_{uv} = G$ also, hence

$$H^*(u,v) = F(u,v) + \rho(u) + \sigma(v) \tag{5.9}$$

for certain functions $\rho(u)$ and $\sigma(v)$ of one variable. We only need to determine ρ and σ so that H^* satisfies (3.3). Now, according to (5.8)

$$\frac{\partial^2}{\partial u\,\partial v}\left(\frac{F_{uu}}{\lambda} - \frac{F_{vv}}{\mu}\right) = 0,$$

hence

$$F_{uu}/\lambda - F_{vv}/\mu = \varphi(u) + \psi(v). \tag{5.10}$$

If we choose ρ and σ to satisfy

$$\rho'' = -\lambda\cdot\varphi, \quad \sigma'' = -\mu\cdot\psi,$$

then it is not hard to see that H^*, as determined by (5.9), satisfies (3.3), since

$$\frac{H^*_{uu}}{\lambda} - \frac{H^*_{vv}}{\mu} = \frac{F_{uu}}{\lambda} + \frac{\rho''}{\lambda} - \frac{F_{vv}}{\mu} - \frac{\sigma''}{\mu} = 0$$

by virtue of (5.10). This completes the construction of the appropriate density H^*, and hence proves the theorem.

If the Hamiltonian density H in Theorem 5.3 is one of the densities H_n in the hierarchy (3.7), then it is not hard to see that the corresponding density $H^*(u,v)$ can be taken to be the density H_{n+2}; similarly, if $H = \tilde{H}_n$, then $H^* = \tilde{H}_{n+2}$. (See also Lemma 6.2 below.)

Example 5.4: Consider Eqs. (1.6) of polytropic gas dynamics, so that $f(v) = v^{\gamma-2}$. Here $\lambda(u) = 1$, and $\mu(v) = v^{\gamma-3}$. Therefore the matrix variables (5.1) coincide

$$U = V = \begin{pmatrix} u & v^{\gamma-2}/(\gamma-2) \\ v & u \end{pmatrix}.$$

According to the proof of Theorem 5.3, we can write the gas dynamics equations in the new Hamiltonian form

$$\mathbf{u}_t = \mathscr{E} E[H^*].$$

The Hamiltonian operator is

$$\mathscr{E} = D_x \cdot U_x^{-1} \cdot D_x \cdot U_x^{-1} \cdot \sigma_1 \cdot D_x,$$

and the new Hamiltonian is

$$H^*(u,v) = \tilde{H}_4(u,v) = \tfrac{1}{24}u^4 v + u^2 \cdot v^\gamma/\gamma(\gamma-1)$$

$$+ v^{2\gamma-1}/2\gamma(\gamma-1)^2(2\gamma-1),$$

provided $\gamma \neq \frac{1}{2}$ (otherwise the last summand is logarithmic).

Similar formulas hold for nonpolytropic gas dynamics.

As discussed in Sec. IV, polytropic gas dynamics has two additional Hamiltonian structures. A similar, but even lengthier, calculation along the lines of that in Proposition 5.2 shows that for polytropic gas dynamics, the Hamiltonian operators $\mathscr{E}$ and $\mathscr{D}_1$ are *not* compatible, nor are the Hamiltonian operators $\mathscr{E}$ and $\mathscr{D}_2$. [In this case, there are other basis trivectors besides the ones listed in (5.5) which show up in the compatibility conditions (5.3), e.g., $\xi \wedge \eta_x \wedge \xi_{xx}$, and one checks that the coefficient of at least one of these trivectors does not vanish.] Thus, as with the one-dimensional Riemann equation (2.1), the polytropic gas dynamic equations form a quadri-Hamiltonian system with four compatible and two incompatible Hamiltonian pairs.

VI. RECURSION OPERATORS AND HIGHER-ORDER SYMMETRIES

We now turn to items (7) and (8) of Sec. II, and discuss the higher-order symmetries for hyperbolic systems. We begin by invoking Magri's theorem[3] to construct recursion operators from the Hamiltonian pairs of differential operators. The first consequence of the developments in the preceding section is the following result of Sheftel'.[13]

Proposition 6.1: The operator

$$\hat{\mathscr{R}} = \mathscr{E} \cdot \mathscr{D}_0^{-1} = D_x \cdot V_x^{-1} \cdot D_x \cdot U_x^{-1} \tag{6.1}$$

is a recursion operator for the separable Hamiltonian system (1.1).

Thus applying the recursion operator $\hat{\mathscr{R}}$ to any symmetry of the gas dynamics system leads to another symmetry. A straightforward calculation shows that for the zeroth-order conserved densities, this recursion operator does not lead to anything new.

Lemma 6.2: Let $\mathbf{v}_n$ and $\tilde{\mathbf{v}}_n$ be the nth-order Hamiltonian symmetries determined by the two hierarchies (3.7). Then the recursion operator acts on their characteristics according to

$$\hat{\mathscr{R}}(\mathbf{Q}_n) = \mathbf{Q}_{n-2}, \quad \hat{\mathscr{R}}(\tilde{\mathbf{Q}}_n) = \tilde{\mathbf{Q}}_{n-2}. \tag{6.2}$$

In the case of a generalized gas dynamics Hamiltonian, so $\lambda(u) \equiv 1$, the matrix variables (5.1) coincide:

$$U = V = \begin{pmatrix} u & M(v) \\ v & u \end{pmatrix}.$$

Thus, just as in the one-dimensional case, the recursion operator (6.1) is the square of a simpler recursion operator[13]

$$\mathscr{R} = D_x \cdot U_x^{-1}, \tag{6.3}$$

On the zeroth-order symmetries,

$$\mathscr{R}(\mathbf{Q}_n) = \mathbf{Q}_{n-1}, \quad \mathscr{R}(\tilde{\mathbf{Q}}_n) = \tilde{\mathbf{Q}}_{n-1}. \tag{6.4}$$

In the polytropic case, $\mathscr{R}$ is the "inverse" to the recursion operator $\mathscr{R}_1$ on the hierarchies (3.7), although as always, this is special to these particular hierarchies.

Since we cannot obtain any higher-order symmetries by applying Sheftel's recursion operator to the first-order symmetries, we need to look elsewhere for the analogies to the higher order rational symmetries of the Riemann equation. We begin by stating the basic condition for a generalized vector field to be a symmetry of (1.1).

Proposition 6.3: The function $\mathbf{Q}[\mathbf{u}] = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$ is the characteristic of a symmetry vector field $\mathbf{v}_{\mathbf{Q}} = \mathbf{Q} \cdot \partial_{\mathbf{u}}$ of (1.2) if and only if $\mathbf{Q}[\mathbf{u}]$ satisfies the identity

$$D_t \mathbf{Q} = D_x(\mathbf{H} \cdot \mathbf{Q}) \tag{6.5}$$

on solutions to the system (1.2).

This result is standard[27] using the fact that the operator $D_t - D_x \cdot \mathbf{H}$ is the Fréchet derivative operator for the system of differential equations (1.2). As a consequence of Proposition 6.3 and the usual condition for a recursion operator,[28] we also deduce the following.

Corollary 6.4: The Fréchet derivative and recursion operators commute:

$$(D_t - D_x \cdot \mathbf{H}) \cdot \hat{\mathscr{R}} = \hat{\mathscr{R}} \cdot (D_t - D_x \cdot \mathbf{H}). \tag{6.6}$$

This can also be proved directly.[13] The commutation identity (6.6) also clearly holds for the first-order recursion operator (6.3) in the generalized gas dynamics case.

Theorem 6.5: Let $H = H_n$ be a Hamiltonian in the first hierarchy constructed in Sec. III, and let $\mathbf{v}_n$ be the corresponding Hamiltonian vector field. Let $\hat{\mathbf{v}}_{2m}$ denote the generalized vector field of order $2m$ with characteristic

$$\hat{\mathbf{Q}}_{2m} = \hat{\mathscr{R}}^m(x\mathbf{u}_x). \tag{6.7}$$

Then $\hat{\mathbf{v}}_{2m}$ is a symmetry for the flow generated by $\mathbf{v}_n$ provided $2m \geqslant n-1$. Similarly, $\hat{\mathbf{v}}_{2m}$ is a symmetry for the flow generated by $\tilde{\mathbf{v}}_n$ corresponding to the Hamiltonian $\tilde{H}_n$ provided $2m \geqslant n$.

Proof: We need to verify the symmetry criterion (6.5) of Proposition 6.3. Using the recursion condition (6.6), (6.2), and (6.7), we see that, on solutions to (1.2) with $\mathbf{H} = \mathbf{H}_n$,

$$\begin{aligned}(D_t - D_x \cdot \mathbf{H}_n)\hat{\mathbf{Q}}_{2m} &= (D_t - D_x \cdot \mathbf{H}_n) \cdot \hat{\mathscr{R}}^m(x\mathbf{u}_x) \\ &= \hat{\mathscr{R}}^m \cdot (D_t - D_x \cdot \mathbf{H}_n)(x\mathbf{u}_x) \\ &= \hat{\mathscr{R}}^m \cdot \{x \cdot D_x(\mathbf{u}_t - \mathbf{H}_n \cdot \mathbf{u}_x) - \mathbf{H}_n \cdot \mathbf{u}_x\} \\ &= -\hat{\mathscr{R}}^m(\mathbf{Q}_n) = -\mathbf{Q}_{n-2m},\end{aligned}$$

which vanishes if $n - 2m \leqslant 1$. The proof for $\tilde{H}_n$ is the same, but now $\tilde{\mathbf{Q}}_{n-2m}$ vanishes if $n - 2m \leqslant 0$.

In the case of generalized gas dynamics, there is a more extensive hierarchy of symmetries because the recursion operator is a first-order differential operator. The same calculation yields the following theorem.

Theorem 6.6: Suppose $H = H_n$ be one of the nth-order generalized gas dynamics Hamiltonians, as in (3.7), and let $\mathbf{v}_n$ be the corresponding first-order Hamiltonian flow. Let $\hat{\mathbf{v}}_m$ denote the generalized vector field of order m with characteristic

$$\hat{\mathbf{Q}}_m = \mathscr{R}^m(x\mathbf{u}_x) = \mathscr{R}^{m-1} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{6.8}$$

Then $\hat{\mathbf{v}}_m$ is a symmetry for the flow generated by $\mathbf{v}_n$ provided $m \geqslant n-1$. Similarly, $\hat{\mathbf{v}}_m$ is a symmetry for the flow generated by $\tilde{\mathbf{v}}_n$ corresponding to the Hamiltonian $H = \tilde{H}_n$ provided $m \geqslant n$.

Finally, we note that in polytropic gas dynamics, we can still form additional recursion operators by combining the Hamiltonian operator $\mathscr{E}$ with the operators $\mathscr{D}_1$, $\mathscr{D}_2$, even though they are not compatible.[29] However, the resulting higher-order symmetries appear to always be nonlocal since we cannot explicitly invert $\mathscr{D}_1$ or $\mathscr{D}_2$.

VII. HIGHER-ORDER CONSERVATION LAWS FOR GENERALIZED GAS DYNAMICS

Proposition 3.1 gives a complete description of the zeroth-order conservation laws for the general Hamiltonian system (1.1). For the classical gas dynamics Hamiltonian (1.6), Verosky[14] found an additional first-order conserved density,

$$\hat{H}_1[u,v] = \frac{v_x}{\delta} = \frac{v_x}{u_x^2 - \mu(v)\cdot v_x^2}. \tag{7.1}$$

We now show how Verosky's conservation law fits into our general framework, and derive analogous laws for generalized gas dynamics Hamiltonians. This will complete our extension of the results of Sec. II to two-dimensional Hamiltonian systems. We first note that a generalized gas dynamics Hamiltonian system can be cast into a suggestive matrix form, which is very reminiscent of the scalar Riemann equation (2.1).

Lemma 7.1: Suppose $H(u,v)$ is a generalized gas dynamics Hamiltonian. Then the corresponding Hamiltonian system (1.2) is equivalent to the matrix equation

$$U_t = \mathbf{H}\cdot U_x. \tag{7.2}$$

This follows from an elementary direct computation. The key to Verosky's conservation law and its higher order generalizations is the following matrix divergence identity.

Lemma 7.2: Let $H(u,v)$ be a generalized gas dynamics Hamiltonian. Then

$$D_t\cdot(U_x^{-1}) - D_x(\mathbf{H}\cdot U_x^{-1}) = -(\mathbf{H}_x\cdot U_x^{-1} + U_x^{-1}\cdot\mathbf{H}_x) \tag{7.3}$$

holds on solutions to the system (1.2).

Proof: We first note that the matrices $\mathbf{H}$ and U_x commute:

$$\mathbf{H}\cdot U_x = U_x\cdot\mathbf{H}.$$

Also, differentiating (7.2) we find

$$U_{xt} = \mathbf{H}\cdot U_{xx} + \mathbf{H}_x\cdot U_x.$$

Therefore we immediately deduce (7.3):

$$\begin{aligned} &D_t\cdot(U_x^{-1}) - D_x(\mathbf{H}\cdot U_x^{-1}) \\ &\quad = -U_x^{-1}\cdot U_{xt}\cdot U_x^{-1} - \mathbf{H}_x\cdot U_x^{-1} \\ &\qquad + \mathbf{H}\cdot U_x^{-1}\cdot U_{xx}\cdot U_x^{-1} \\ &\quad = -\mathbf{H}_x\cdot U_x^{-1} - U_x^{-1}\cdot\mathbf{H}_x. \end{aligned}$$

In particular, the (2,1) entry of the matrix identity (7.3) reads

$$D_t\left(\frac{v_x}{\delta}\right) + D_x\left(\frac{H_{uu}\cdot v_x - H_{uv}\cdot u_x}{\delta}\right) = -2H_{uuu}. \tag{7.4}$$

For the classical gas dynamics Hamiltonian (1.7), $H_{uuu}\equiv 0$, and we recover Verosky's conservation law, with density (7.1). For more general gas dynamics Hamiltonians, H_{uuu} will no longer be 0, and v_x/δ will no longer be a conserved density; however, we can simply modify it to get a time-dependent conservation law.

Theorem 7.4: Let $H(u,v)$ be a gas dynamics Hamiltonian. Then the function

$$\hat{H}_1^* = v_x/\delta + 2tH_{uuu} \tag{7.5}$$

is a conserved density for the corresponding flow (1.1).

Proof: Let $T(u,v) = H_{uuu}$. Note first that according to Lemma 3.4, $T(u,v)$ is a conserved density for the flow. According to (7.4), to prove that (7.5) is also a conserved density, it suffices to show that there exists a function $X(u,v)$ such that the divergence identity

$$D_t(t\cdot T(u,v)) - D_x(t\cdot X(u,v)) = T(u,v) \tag{7.6}$$

holds. Evaluating the left-hand side of (7.6), and using Eq. (1.1), we find that X must satisfy the pair of first-order partial differential equations

$$X_u = H_{uv}T_u + H_{uu}T_v, \quad X_v = H_{vv}T_u + H_{uv}T_v.$$

The integrability condition $X_{uv} = X_{vu}$ simplifies to the condition

$$H_{vv}T_{uu} = H_{uu}T_{vv}. \tag{7.7}$$

But this follows immediately from Lemma 3.1, using the fact that T is a conserved density. This completes the proof.

Corollary 7.5: The integral

$$\hat{\mathscr{H}}_1 = \int \frac{v_x}{\delta}\,dx,$$

when it converges, is a linear function of t for any gas dynamics Hamiltonian.

The first-order conserved density $\hat{H}_1 = v_x/\delta$ leads to a Hamiltonian flow using the basic Hamiltonian operator $\mathscr{D}_0$. We now connect this flow with the symmetries generated in Theorem 6.6. This will allow us to apply the theorem of Magri to the Hamiltonian pair $\mathscr{E}$ and $\mathscr{D}_0$, and thereby generate a new hierarchy of higher order conservation laws in gas dynamics. The starting point is the following straightforward result.

Proposition 7.6: The symmetry $\hat{\mathbf{v}}_3$ is Hamiltonian with respect to the Hamiltonian structure determined by $\mathscr{D}_0$, and the corresponding conserved density is -2 times Verosky's density (7.1).

Proof: According to (6.8), the characteristic for the symmetry $\hat{\mathbf{v}}_3$ is given by

$$\hat{\mathbf{Q}}_3 = \mathscr{R}^2\cdot\begin{pmatrix}1\\0\end{pmatrix} = -D_x\cdot U_x^{-2}\cdot U_{xx}\cdot U_x^{-1}\cdot\begin{pmatrix}1\\0\end{pmatrix}.$$

On the other hand, the Hamiltonian flow corresponding to Verosky's density has characteristic

$$D_x\cdot\begin{pmatrix}E_v(\hat{H}_1)\\E_u(\hat{H}_1)\end{pmatrix}.$$

Therefore, it suffices to verify that

$$\begin{pmatrix}E_v(\hat{H}_1)\\E_u(\hat{H}_1)\end{pmatrix} = 2U_x^{-2}\cdot U_{xx}\cdot U_x^{-1}\cdot\begin{pmatrix}1\\0\end{pmatrix}.$$

We have thus reduced the proposition to a straightforward computation, which we leave to the reader.

Theorem 7.7: For a generalized gas dynamics Hamiltonian, there is a hierarchy of higher order Hamiltonian densities $\hat{H}_m$, $m = 1,2,\ldots$, with m indicating the order of derivatives on which they depend, and corresponding commuting bi-Hamiltonian systems

$$\mathbf{u}_t = \hat{\mathbf{Q}}_{2m+1} = \mathscr{D}_0 E[\hat{H}_m] = \mathscr{E} E[\hat{H}_{m-1}], \quad m\geqslant 1. \tag{7.8}$$

The Hamiltonians are in involution with respect to both the $\mathscr{D}_0$ and $\mathscr{E}$ Poisson brackets.

This follows directly from Magri's theorem.[3,22] There is, however, a technical point to be resolved: whether we can always invert the differential operator $\mathscr{D}_0$ to continue the recursive construction of the densities from (7.8). However this follows from the theorem in Ref. 30 that shows that this always is the case when one of the members of a Hamiltonian pair is a constant coefficient differential operator.

Note that if one of the higher-order densities $\hat{H}_m$ is conserved, then so is every subsequent density $\hat{H}_k$, $k \geqslant m$. Thus we need only know which is the first conserved density for our Hamiltonian system (1.1).

Theorem 7.8: If $H = H_n$ is a Hamiltonian density in the first generalized gas dynamics hierarchy, then the higher-order density $\hat{H}_m$ is conserved for the Hamiltonian system (1.2) provided $n \leqslant 2m + 1$. If $H = \tilde{H}_n$ is in the second generalized gas dynamics hierarchy, then $\hat{H}_m$ is conserved provided $n \leqslant 2m + 2$.

Proof: We apply the recursion operator $\hat{\mathscr{R}} = \mathscr{R}^2$ to the (time-dependent) flow corresponding to the Hamiltonian density $\hat{H}_1^*$ [cf. (7.5)]. Using (3.8) and (6.4), we conclude that, as in the one-dimensional case, the function

$$\hat{H}_m + t \frac{\partial^{2m+1} H}{\partial u^{2m+1}}$$

is a conserved density for (1.1) with $H = H_n$ or $\tilde{H}_n$. In particular, if $2m + 1 > n$, the second summand vanishes; more particularly, if $H = H_n$, and $n = 2m + 1$, then the second summand is just a constant multiple of t, and can be ignored. This completes the proof.

For more general separable Hamiltonian systems not of gas dynamics type, higher order conservation laws do not appear to exist in general. The principal reason for this is that the recursion operator (6.1) is second order, and the hierarchy of higher-order symmetries (6.7) omits the obvious Hamiltonian candidates. However, we do not have a complete proof that these systems do not have higher-order conservation laws, and, indeed, Verosky's transformation (3.5) shows that at least one special example—the Born–Infeld equation—they do exist. This is an area that requires further investigation.

In a later publication, we hope to return to the application of our results to physically interesting initial value problems and discontinuous solutions/shock waves. Another interesting direction for further research is to try to extend these results to three-dimensional systems of conservation laws, especially those of nonisentropic gas dynamics for which Verosky has discovered additional higher order conservation laws.[31]

ACKNOWLEDGMENTS

We would like to thank John Verosky for many helpful remarks, and for showing us his remarkable transformation (3.5).

This work was supported in part by NATO Collaborative Research Grant No. RG 86/0055. One of the authors (P.O.) was supported in part by NSF Grant DMS 86-02004.

[1] C. S. Gardner, "Korteweg–de Vries equation and generalizations. IV. The Korteweg–de Vries equation as a Hamiltonian system," J. Math. Phys. **12**, 1548 (1971).
[2] V. I. Arnol'd, "Sur la géometrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits," Ann. Inst. Fourier Grenoble **16**, 319 (1966).
[3] F. Magri, "A simple model of the integrable Hamiltonian equation," J. Math. Phys. **19**, 1156 (1978).
[4] D. D. Holm, J. E. Marsden, T. Ratiu, and A. Weinstein, "Nonlinear stability of fluid and plasma equilibria," Phys. Rep. **123**, 1 (1985).
[5] Yu. I. Manin, "Algebraic aspects of nonlinear differential equations," J. Sov. Math. **11**, 1 (1979).
[6] J. Cavalcante and H. P. McKean, "The classical shallow water equations: Symplectic geometry," Physica D **4**, 253 (1982).
[7] B. A. Dubrovin and S. A. Novikov, "Hamiltonian formalism of one-dimensional systems of hydrodynamic type and the Bogolyubov–Whitham averaging method," Sov. Math. Dokl. **27**, 665 (1983).
[8] Y. Nutku, "On a new class of completely integrable systems. II. Multi-Hamiltonian structure," J. Math. Phys. **28**, 2579 (1987).
[9] M. Slemrod, "Dynamics of first order phase transitions," in *Phase Transformations and Material Instabilities in Solids*, edited by M. E. Gurtin (Academic, New York, 1984), pp. 163–203.
[10] C. M. Dafermos, "Estimates for conservation laws with little viscosity," SIAM J. Math. Anal. **18**, 409 (1987).
[11] T. B. Benjamin and S. Bowman, "Discontinuous solutions of one-dimensional Hamiltonian systems," Proc. R. Soc. London Ser. B **413**, 263 (1987).
[12] M. B. Sheftel', "On the infinite-dimensional noncommutative Lie–Bäcklund algebra associated with the equations of one-dimensional gas dynamics," Theor. Math. Phys. **56**, 878 (1983).
[13] M. B. Sheftel', "Integration of Hamiltonian systems of hydrodynamic type with two dependent variables with the aid of the Lie–Bäcklund group," Func. Anal. Appl. **20**, 227 (1986).
[14] J. M. Verosky, "First order conserved densities for gas dynamics," J. Math. Phys. **27**, 3061 (1986).
[15] P. J. Olver, *Applications of Lie Groups to Differential Equations*, Graduate Texts in Mathematics, Vol. 107 (Springer, New York, 1986).
[16] G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[17] Reference 16, p. 579.
[18] M. Arik, F. Neyzi, Y. Nutku, P. J. Olver, and J. M. Verosky, "Hamiltonian structures for the Born–Infeld equation," in preparation.
[19] Reference 16, p. 123.
[20] Y. Nutku, "On a new class of completely integrable systems. I. Infinitely many conservation laws," J. Math. Phys. **26**, 1237 (1985).
[21] B. Fuchssteiner and A. S. Fokas, "Symplectic structures, their Bäcklund transformations and hereditary symmetries," Physica D **4**, 47 (1981).
[22] Reference 15, Chap. 7.
[23] P. Morrison (personal communication).
[24] J. M. Verosky (personal communication).
[25] P. J. Olver, "On the Hamiltonian structure of evolution equations," Math. Proc. Camb. Phil. Soc. **88**, 71 (1980).
[26] Reference 15, corollary 7.21.
[27] Reference 15, Chap. 5.
[28] Reference 15, Theorem 5.30.
[29] Reference 15, remarks at bottom of p. 450.
[30] P. J. Olver, "BiHamiltonian systems," in *Ordinary and Partial Differential Equations*, edited by B. D. Sleeman and R. J. Jarvis, Pitman Research Notes in Mathematics Series, No. 157 (Longman Scientific and Technical, New York, 1987), pp. 176–193.
[31] J. M. Verosky, "Higher order symmetries of the compressible one-dimensional isentropic fluid equations," J. Math. Phys. **25**, 884 (1984).

The scalar product in two-particle relativistic quantum mechanics

H. Sazdjian
Division de Physique Théorique,[a] *Institut de Physique Nucléaire, Université Paris XI, 91406, Orsay Cedex, France*

(Received 16 July 1987; accepted for publication 24 February 1988)

In the framework of two-particle relativistic quantum mechanics, a Poincaré-invariant scalar product and the corresponding physical Hilbert space of states are constructed. This is achieved by finding a tensor current of rank 2, $j_{\mu\nu}(x_1,x_2)$, satisfying two independent conservation laws, relative to particles 1 and 2, respectively. Then the scalar product is obtained by integrating the current $j_{\mu\nu}$ over two three-dimensional spacelike hypersurfaces. The Hermiticity of the Poincaré group generators is ensured by the fact that the kernel of the current $j_{\mu\nu}$ is translation invariant and covariant. A simple expression of the scalar product is obtained when one chooses for the two spacelike hypersurfaces two constant parallel hyperplanes. The positivity of the norm is, in general, ensured if the spectrum of the eigenvalues of the total mass squared operator comes out to be positive.

I. INTRODUCTION

In order to have a general framework of two-particle relativistic quantum mechanics,[1-3] it is necessary to define, in addition to the wave equations, the physical Hilbert space in which the states are considered, and more precisely, to construct the scalar product of the theory. It is this last problem that we intend to solve throughout this work.

The main property that we demand from the scalar product is its Poincaré invariance, which then guarantees, with the covariance of its kernel, the Hermiticity property of the Poincaré group generators and permits a unitary realization of the Poincaré group. Of course, one also has to ensure the positivity of the norm of physical states.

In free, one-particle relativistic quantum mechanics a Poincaré-invariant scalar product is constructed by means of a conserved four-vector current, $j_\mu(x)$, built up from two wave functions that are solutions of the corresponding wave equation. The conservation of the current guarantees that its three-dimensional integral over a spacelike surface, which defines the scalar product, is independent of that surface, and this in turn ensures the Poincaré invariance of the scalar product itself.[4]

In the two-particle case one naturally should extend this procedure. One should construct, from two wave functions, a tensor current of rank 2, $j_{\mu\nu}(x_1,x_2)$, so that it satisfies the two conservation laws

$$\partial_1^\mu j_{\mu\nu}(x_1,x_2)=0,\qquad \partial_2^\nu j_{\mu\nu}(x_1,x_2)=0. \tag{1.1}$$

The scalar product should then be constructed as a double three-dimensional integral of this current over two spacelike hypersurfaces Σ_1 and Σ_2:

$$(\Psi,\Phi)=\int_{\Sigma_1,\Sigma_2} j_{\mu\nu}(x_1,x_2)\,d\sigma_1^\mu(x_1)\,d\sigma_2^\nu(x_2). \tag{1.2}$$

It can easily be seen that the two conservation laws (1.1) guarantee the independence of the integral (1.2) from the types of the surfaces Σ_1 and Σ_2, and, more particularly, ensure the Poincaré invariance of the scalar product defined by (1.2).

It is this procedure that we shall adopt to construct the scalar product of the theory. The difficulty of the problem consists in finding the tensor current $j_{\mu\nu}$ satisfying the conservation laws (1.1). Contrary to one-particle mechanics, in the two-particle case the interaction potentials depend in general on the total four-momentum p of the system,[3] that is, in particular, on the energy, except perhaps in the c.m. (center of momentum) frame. This is because the potentials depend on the relative coordinate $x\equiv x_1-x_2$ only through its transverse components with respect to p, that is, through the variables

$$x_\mu^T=x_\mu-[(p\cdot x)/p^2]p_\mu,\quad x^{T2}=x^2-[(p\cdot x)^2/p^2]. \tag{1.3}$$

In the c.m. frame, x^T reduces to the variable $(0,\vec{x})$ and hence becomes independent of p. However, for realistic potentials, the latter may also exhibit—through the coupling constants, for instance—an explicit dependence on the total mass squared and display a form like

$$V=V(x^{T2},p^2,\ldots), \tag{1.4}$$

where the dots stand for other possible variables such as the relative momentum or Dirac matrices.

This means that for two wave functions that are eigenfunctions of two different eigenvalues of p^2, the interaction potential will not be the same for each of them. It is this feature that makes it difficult to build up the tensor current $j_{\mu\nu}$. The problem can be solved, however, as we shall show in this paper. The expression for the resulting current $j_{\mu\nu}$ turns out to be a complicated nonlocal function of the corresponding wave functions and potentials.

We construct the scalar product (1.2) by choosing for the surfaces Σ_1 and Σ_2 two parallel hyperplanes perpendicular to a constant timelike vector n. With this choice, the expression for the scalar product is considerably simplified and takes a local form. It differs from the expression for the scalar product of two free particles by terms taking into account the total momentum dependence of the potential.

[a] Laboratory associated with Centre National de la Recherche Scientifique.

 0022-2488/88/071620-14$02.50

In order to relate the quantum-mechanical wave functions to field-theoretic quantities (such as matrix elements of currents), it is necessary to fix the normalization constants of the former by comparing their norm to that obtained from the Bethe–Salpeter equation with approximations leading to local potentials.[5] This comparison fixes the overall normalization constant of the norm. It is not equal to unity, as long as the scale of the norm is fixed by the normalization condition of the field theory state $|p\rangle$, and this merely reflects the fact that quantum-mechanical two-particle wave functions are only projections of field theory states $|p\rangle$ onto two-particle states. This normalization constant depends on the characteristics of the quantum-mechanical wave function, such as its total mass squared and the mean value of its relative momentum squared.

Expressions for scalar products in the case of p^2-dependent potentials have also been obtained, in the three-dimensional formalism, by Lepage[6] and Caswell and Lepage,[7] who used a three-dimensional reduction method on the Bethe–Salpeter equation. Their formulas cannot, however, be applied in our case, since the three-dimensional wave equations they obtain differ from ours. In Lepage's work[6] there is an asymmetric treatment of the two constituents of the system, while in our wave equations the two particles are treated in a symmetric way. However, this feature does not prevent them from satisfying the requirement that if the mass of one of the particles is taken to be infinite, the other particle satisfies a Klein–Gordon or a Dirac equation in an external potential. In the work of Caswell and Lepage[7] the asymmetry phenomenon is absent, but still their wave equations are different. These features do not mean that there is a contradiction between these works and ours. It is well-known that by wave-function transformations one can always modify the structure of a wave equation and, correspondingly, the kernel of the scalar product. Furthermore, there are many ways of reducing the Bethe–Salpeter equation to a three-dimensional one. The result depends on the trial three-dimensional kernel one chooses as the zeroth-order kernal to iterate in order to reach the Bethe–Salpeter kernel.

We would also like to emphasize the difference in our approach in constructing the scalar product. Our main objective has been the formulation of two-particle relativistic quantum mechanics from the general principles of the Hamiltonian formalism (more precisely, the manifestly covariant formalism with constraints) without any reference to quantum field theory, parallel to the way nonrelativistic quantum mechanics is formulated. This is why the wave equations and the corresponding scalar products are constructed by self-consistent methods and not by using known properties of field-theoretic quantities. Furthermore, we also provide the expressions for conserved currents, which therefore contain, at least explicitly, more information about the local properties of wave functions than in previous work. On the other hand, Ref. 5 establishes the connection between the two types of formalism: the one obtained from field theory, the other from relativistic Hamiltonian quantum mechanics.

In order to introduce the method we use to construct the tensor current $j_{\mu\nu}$, we first consider, in Sec. II, the case of one-particle relativistic quantum mechanics with energy-dependent potentials. This problem is not generally well known, although a particular case of it already occurs in the problem of a spinless particle placed in an external electromagnetic potential.[8] General aspects of this question have been treated by several authors,[9–12] and in Ref. 13 a summary of related properties was presented. One main feature of energy-dependent potentials is the possible occurrenceof the so-called associated vectors.[13–15] These states usually appear when zero-norm states occur in the spectrum, and thus they modify with their special properties the completeness relation of ordinary states. Associated vectors occur, however, for particular domains of the coupling constants or for particular shapes of the potentials. We shall assume throughout this work that the potentials we are considering are so chosen as to be free of such pathologies, and no zero-norm states occur in the spectrum.

The plan of the paper is as follows. In Sec. II we consider one-particle relativistic quantum mechanics with energy-dependent potentials and construct the corresponding scalar product. In Sec. III we consider two spin-0 particle systems. (A summary of this section was already presented in Ref. 16.) Section IV is devoted to the study of the particular properties of the scalar product, as constructed on two constant, parallel, spacelike hyperplanes. In Sec. V we consider spin-$\frac{1}{2}$ fermion–antifermion systems, and Sec. VI concerns systems composed of one spin-$\frac{1}{2}$ fermion and one spin-0 boson. The Conclusion follows in Sec. VII.

II. ONE-PARTICLE RELATIVISTIC QUANTUM MECHANICS WITH ENERGY-DEPENDENT POTENTIALS

One-particle relativistic equations with external potentials lose, in general, part of their Poincaré invariance. For instance, with static potentials one loses at least spatial translation invariance and Lorentz invariance. Therefore, in this case, one is mostly interested in the Hermiticity property of the Hamiltonian of the system, which guarantees the orthogonality of energy eigenfunctions with different eigenvalues. If the potential is rotation invariant, then one would also check the Hermiticity property of the total angular momentum operator, which is usually an easier task than for the Hamiltonian case. In general, the expressions for the conserved currents that are found in the free case remain true also in the presence of such external potentials, and continue to serve to define the scalar product of the system and to ensure the Hermiticity property of the Hamiltonian.

In this section we shall consider the more complicated case of static potentials depending on the Hamiltonian itself or, more particularly, on the energy of the system when acting on energy eigenfucntions. A well-known example of such a case is provided by the Klein–Gordon equation for a charged spinless particle in the presence of a static electric potential.[8,9] Here the effective potential, resulting from the minimal coupling, displays a linear dependence on the energy. More complicated dependence may arise when one considers effective potentials obtained from approximations of field theoretic quantities, as in the quasipotential approach to the Bethe–Salpeter equation[17–19] for the two-body relative motion.

We shall therefore consider general static potentials of the type

$$V = V(\vec{x}, p_0), \tag{2.1}$$

where p_0 is the energy operator acting as the time-derivative operator $i\partial/\partial x^0$. The potential V may also depend on the three-momentum $\vec{p}$. But this kind of dependence, provided it is appropriately symmetrized, will not play any particular role in the subsequent discussions and therefore will be omitted from the notation.

Strictly speaking, with arbitrary energy-dependent potentials (2.1), one partly loses the Hamiltonian description of the system at the level of the initial-value problem. In general, with energy-independent (or, at most, linearly or quadratically dependent) potentials, one can predict for an arbitrary state $|\psi(t_0)\rangle$, given at time t_0 (with its first-order derivative, if we deal with the Klein–Gordon equation), its evolution at later times. This is obtained by spanning the state $|\psi\rangle$ by means of the completeness relation along the eigenstates of the energy operator, the time evolution of each of them being well known. It turns out, however, that for potentials of the type (2.1), the kernel of the scalar product will be also energy dependent, which, in turn, reappears in the completeness relation. Therefore an arbitrary state $|\psi\rangle$, given at time t_0, can no longer be decomposed along the eigenstates of the energy operator, since the completeness relation will involve, usually in a complicated form, the time-derivative operator $i\partial/\partial t$. The procedure to be adopted in this case is the following. One must first solve the energy eigenvalue problem. Then one has to consider, in general, only those states that are known superpositions of energy eigenstates. For such states the time evolution problem will be completely solved. This is the price to be paid for the Hamiltonian treatment of wave equations with energy-dependent potentials.

We now turn to the question of construction of the scalar product of the theory. As we described in Sec. I, we search for a conserved current $j_\mu(x)$ built up from two energy eigenfunctions Ψ and Φ. We write j_μ in the form

$$j_\mu = j_\mu^{(0)} + j_\mu^{(1)}, \tag{2.2}$$

where $j_\mu^{(0)}$ represents the current obtained in the free theory (or in theories with energy-independent potentials), and $j_\mu^{(1)}$ is a remainder. In general, $j_\mu^{(0)}$ does not alone ensure current conservation. We therefore seek an expression for $j_\mu^{(1)}$ that will ensure the current conservation of j_μ.

We shall consider the two cases of spin-0 and spin-$\frac{1}{2}$ particles.

A. The Klein–Gordon equation

The wave equation with potential (2.1) is

$$(\partial^2 + m^2 + V(\vec{x}, i\partial_0))\Psi(x) = 0. \tag{2.3}$$

The potential $V(\vec{x}, p_0)$ is assumed to be superficially Hermitian in the sense that when p_0 is replaced by a real eigenvalue, then V is Hermitian in the usual L_2 norm. Furthermore, we assume that the shape of V, its coupling constant, and its p_0 dependence are such that all the eigenvalues of p_0 come out to be real.

We first consider the case of two different eigenvalues p_0 and p_0' of the energy operator $i\partial_0$, with corresponding eigenfunctions Φ and Ψ, respectively; they can also be decomposed as

$$\Phi(x) = e^{-ip_0 x^0}\phi(\vec{x}), \quad \Psi(x) = e^{-ip_0' x^0}\psi(\vec{x}). \tag{2.4}$$

The expression for the norm of a wave function will be obtained as a limit, with $p_0' \to p_0$, of the expression for the scalar product defined for $p_0' \neq p_0$. With this aim we shall introduce from the start a small negative imaginary part, $-i\varepsilon$, into the eigenvalues p_0 and p_0' and define the corresponding wave functions (2.4) as limits, with $\varepsilon \to 0$:

$$\Phi(x) = \lim_{\varepsilon \to 0} \Phi_\varepsilon(x) \equiv e^{-i(p_0 - i\varepsilon)x^0}\phi(\vec{x}) \quad (\varepsilon > 0),$$

$$\Psi(x) = \lim_{\varepsilon \to 0} \Psi_\varepsilon(x) \equiv e^{-i(p_0' - i\varepsilon)x^0}\psi(\vec{x}) \tag{2.5}$$

(p_0 and p_0' are real). The wave functions Φ_ε and Ψ_ε satisfy Eq. (2.3) in the limit $\varepsilon = 0$.

The "free" current $j_\mu^{(0)}$ built up from Ψ and Φ is

$$j_\mu^{(0)}(x) = i\,\Psi_\varepsilon^*(x)\overleftrightarrow{\partial}_\mu \Phi_\varepsilon(x). \tag{2.6}$$

For simplicity we shall use these notations in the following discussion:

$$V \equiv V(\vec{x}, p_0 - i\varepsilon), \quad V' \equiv V(\vec{x}, p_0' - i\varepsilon), \tag{2.7}$$

where p_0 and p_0' are the eigenvalues of Φ and Ψ, respectively.

By using the equations of motion, we get

$$\partial^\mu j_\mu^{(0)} = i\,\Psi_\varepsilon^*(V'^* - V)\Phi_\varepsilon. \tag{2.8}$$

Therefore the current $j_\mu^{(1)}$, (2.2), must satisfy the equation

$$\partial^\mu j_\mu^{(1)} = -i\,\Psi_\varepsilon^*(V'^* - V)\Phi_\varepsilon. \tag{2.9}$$

The solution of this equation is easily obtained. We introduce the advanced Green's function $G_A(x)$,

$$G_A(x) = -i\theta(-x^0)\Delta(x, m=0), \tag{2.10}$$

which satisfies the equation

$$\partial^2 G_A(x) = -i\delta^4(x). \tag{2.11}$$

It can be considered the limit as $\varepsilon \to 0$, $\varepsilon > 0$, of the ε-dependent function $G_A(x, \varepsilon)$:

$$G_A(x) = \lim_{\varepsilon \to 0} G_A(x, \varepsilon) \equiv G_A(x) e^{\varepsilon x^0}, \tag{2.12}$$

the Fourier transform of which is

$$G_A(k, \varepsilon) = \int e^{ik\cdot x} G_A(x, \varepsilon) d^4x = \frac{i}{k^2 - 2ik_0\varepsilon}. \tag{2.13}$$

Then $j_\mu^{(1)}$ is given by the expression

$$j_\mu^{(1)}(x) = \partial_\mu \int G_A(x - x')\Psi_\varepsilon^*(x') \times [V(\vec{x}', p_0' + i\varepsilon) - V(\vec{x}', p_0 - i\varepsilon)]\Phi_\varepsilon(x') d^4x'. \tag{2.14}$$

The scalar product of the two wave functions Ψ and Φ is now

$$(\Psi,\Phi)=\int d^3\vec{x}\, j_0(x)$$
$$=\int d^3\vec{x}(j_0^{(0)}(x)+j_0^{(1)}(x))$$
$$=\int d^3\vec{x}\,\Psi_\varepsilon^*(x)(p_0'+p_0)\Phi_\varepsilon(x)$$
$$+\int d^3\vec{x}\,d^4x'\,\partial_0 G_A(x-x')$$
$$\times\Psi_\varepsilon^*(x')(V'^*-V)\Phi_\varepsilon(x'). \quad (2.15)$$

We shall concentrate below on the second term $\int d^3x j_0^{(1)}(x)$. We use the expressions (2.5) of Ψ_ε and Φ_ε:

$$\int d^3\vec{x}\, j_0^{(1)}(x)=\int d^3\vec{x}\,d^4x'\,\partial_0 G_A(x-x')$$
$$\times e^{i(p_0'-p_0+2i\varepsilon)x'^0}$$
$$\times\{\psi^*(\vec{x}')(V'^*-V)\phi(\vec{x}')\}. \quad (2.16)$$

Notice that the term in the curly brackets is now independent of x'^0. We make the following change of variables:

$$\vec{y}=\vec{x}',\quad \vec{y}'=\vec{x}-\vec{x}',\quad y'^0=-(x'^0-x^0). \quad (2.17)$$

Expression (2.16) then becomes

$$\int d^3\vec{x}\, j_0^{(1)}(x)=\partial_0\int d^3\vec{y}\,d^4y'\,G_A(y',2\varepsilon)e^{-i(p_0'-p_0)y'^0}$$
$$\times e^{i(p_0'-p_0+2i\varepsilon)x^0}\{\psi^*(\vec{y})(V'^*-V)\phi(\vec{y})\}, \quad (2.18)$$

where we have used definition (2.12) of $G_A(y,\varepsilon)$. The integration with respect to y' gives the Fourier transform (2.13) of $G_A(y',2\varepsilon)$:

$$\int d^3\vec{x}\, j_0^{(1)}(x)$$
$$=i\partial_0\int d^3\vec{x}\, e^{i(p_0'-p_0+2i\varepsilon)x^0}$$
$$\times\frac{\psi^*(\vec{x})[V(\vec{x},p_0'+i\varepsilon)-V(\vec{x},p_0-i\varepsilon)]\phi(\vec{x})}{(p_0'-p_0)^2+4i\varepsilon(p_0'-p_0)}$$
$$=-\int d^3\vec{x}\,\Psi_\varepsilon^*(x)$$
$$\times\frac{[V(\vec{x},p_0'+i\varepsilon)-V(\vec{x},p_0-i\varepsilon)]}{p_0'-p_0+2i\varepsilon}\Phi_\varepsilon(x), \quad (2.19)$$

where we incorporated the exponentials in the wave functions, according to (2.5).

Finally, we get for the scalar product

$$(\Psi,\Phi)=\int d^3\vec{x}\,\Psi^*(x)\left\{p_0'+p_0\right.$$
$$\left.-\frac{[V(\vec{x},p_0'+i\varepsilon)-V(\vec{x},p_0-i\varepsilon)]}{p_0'-p_0+2i\varepsilon}\right\}\Phi(x). \quad (2.20)$$

It can be checked, by using the equations of motion, that this scalar product is independent of x^0, and that for $p_0\neq p_0'$ it is zero, which is a manifestation of the orthogonality property of two energy eigenfunctions with different eigenvalues. The Hermiticity property of the operator p_0 is also easily checked.

For the evaluation of the norm of states we take $p_0'=p_0$ and then the limit $\varepsilon\to 0$. This gives

$$(\Psi_a,\Psi_b)=\int d^3\vec{x}\,\Psi_a^*(x)\left[2p_0-\frac{\partial V(\vec{x},p_0)}{\partial p_0}\right]\Psi_b(x)$$
$$=2p_0\delta_{ab}, \quad (2.21)$$

where the labels a,b distinguish different eigenfunctions with the same energy.

If the potential was independent of p_0 we would get the expression for the free norm of states for the Klein–Gordon equation. It is a well-known fact in this case that the norm is positive for positive-energy solutions and negative for negative-energy solutions.[8,13] The physical Hilbert space is then chosen to correspond to the subspace of solutions with positive energies.

In the general case (2.21), the question of the positivity of the norm for positive-energy solutions is less straightforward. The potential V must satisfy appropriate conditions concerning its shape and its coupling constant, as well as its energy dependence, in order to maintain the positivity of the norm of states with positive energies. We then have to assume that the physically acceptable potentials are those that satisfy the above conditions and permit again the definition of the physical Hilbert space as the one corresponding to the subspace of positive-energy solutions of Eq. (2.3). (See also the discussion at the end of Sec. III.)

The scalar product (2.20) can be straightforwardly generalized to wave functions Ψ and Φ, which are *known* superpositions of energy eigenfunctions (see the discussion at the beginning of this section). The kernel of the scalar product (2.20) is written now in an operator form, with a self-evident notation:

$$(\Psi,\Phi)=\lim_{\varepsilon\to 0}\int d^3\vec{x}\,\Psi_\varepsilon(x)$$
$$\times\left\{i\overrightarrow{\partial}_0-\frac{[V(\vec{x},-i\overleftarrow{\partial}_0)-V(\vec{x},i\overrightarrow{\partial}_0)]}{(-i\overleftarrow{\partial}_0-i\overrightarrow{\partial}_0)}\right\}\Phi_\varepsilon(x). \quad (2.22)$$

B. The Dirac equation

The wave equation is

$$[i\gamma\cdot\partial-m-V(\vec{x},i\partial_0,\gamma)]\Psi(x)=0, \quad (2.23)$$

with V superficially Hermitian (when $i\partial_0$ is replaced by a real eigenvalue p_0):

$$\gamma_0 V^\dagger\gamma_0=V. \quad (2.24)$$

As in the spin-0 case, we assume that the choice of V has been such that the spectrum of p_0 is real.

For two different eigenfunctions of the energy operator, Ψ and Φ, with different eigenvalues p_0' and p_0, respectively, the free current is

$$j_\mu^{(0)}=\bar{\Psi}_\varepsilon\gamma_\mu\Phi_\varepsilon, \quad (2.25)$$

where we have used notation (2.5), and it satisfies the equation

$$\partial^\mu j_\mu^{(0)}(x) = i\bar{\Psi}_\varepsilon(x)[V(\vec{x},p_0'+i\varepsilon,\gamma) - V(\vec{x},p_0-i\varepsilon,\gamma)]\Phi_\varepsilon(x). \tag{2.26}$$

The current $j_\mu^{(1)}$, (2.2), can be constructed as in the spin-0 case. Its expression is

$$j_\mu^{(1)}(x) = \partial_\mu \int G_A(x-x')\bar{\Psi}_\varepsilon(x') \times [V(\vec{x}',p_0'+i\varepsilon,\gamma) - V(\vec{x}',p_0-i\varepsilon,\gamma)] \times \Phi_\varepsilon(x')d^4x'. \tag{2.27}$$

Following similar calculations as in the spin-0 case we end up with the scalar product

$$(\Psi,\Phi) = \int d^3\vec{x}\,\bar{\Psi}(x) \times \left\{\gamma_0 - \frac{[V(\vec{x},p_0'+i\varepsilon,\gamma) - V(\vec{x},p_0-i\varepsilon,\gamma)]}{p_0'-p_0+2i\varepsilon}\right\} \times \Phi(x), \tag{2.28}$$

and the norm

$$(\Psi_a,\Psi_b) = \int d^3\vec{x}\,\bar{\Psi}_a(x)\left[\gamma_0 - \frac{\partial V(\vec{x},p_0,\gamma)}{\partial p_0}\right]\Psi_b(x) = 2|p_0|\delta_{ab}. \tag{2.29}$$

The remarks made in the spin-0 case as to the physically acceptable potentials, which must guarantee the positivity of the norm of positive-energy states, hold also here.

III. TWO SPIN-0 PARTICLE SYSTEMS

In the manifestly covariant relativistic quantum mechanics the wave function of the two spin-0 particle system satisfies two independent wave equations,[3] each of them being a generalization of the Klein–Gordon equation for particle 1 or 2:

$$H_a\Psi(x_1,x_2) \equiv (p_a^2 - m_a^2 - V)\Psi(x_1,x_2) = 0 \quad (a=1,2), \tag{3.1}$$

where V is a Poincaré-invariant interaction potential.

The compatibility condition of these two equations reads

$$[H_1,H_2]\Psi = 0, \tag{3.2}$$

which, considered in its strong sense, gives the constraint on V:

$$[p_1^2 - p_2^2, V] = 0. \tag{3.3}$$

The Poincaré-invariant general (local) solution of this equation is

$$V = V(x^{T2},x^T\cdot v,p^2,v^2,p\cdot v), \tag{3.4}$$

where we use the following notation:

$$\begin{aligned} &p = p_1 + p_2, \quad v = \tfrac{1}{2}(p_1 - p_2), \quad X = \tfrac{1}{2}(x_1 + x_2), \\ &x = x_1 - x_2, \quad x_\mu^T = x_\mu - (\hat{p}\cdot x)\hat{p}_\mu, \\ &\hat{p}_\mu = p_\mu/(p^2)^{1/2}, \quad p^2 > 0, \quad x^{T2} = x^2 - (\hat{p}\cdot x)^2, \\ &\nabla_\mu = \partial_{1\mu} + \partial_{2\mu}, \quad \partial_\mu = \tfrac{1}{2}(\partial_{1\mu} - \partial_{2\mu}), \end{aligned} \tag{3.5}$$

and for any vector Y we shall define its "transverse" and "longitudinal" parts with respect to the total momentum p by

$$\begin{aligned} &Y_\mu^T = Y_\mu - (\hat{p}\cdot Y)\hat{p}_\mu, \\ &Y_\mu^L = (\hat{p}\cdot Y)\hat{p}_\mu, \quad Y_L = (\hat{p}\cdot Y). \end{aligned} \tag{3.6}$$

By subtracting the two equations (3.1) from each other we get an equation that determines the covariant relative time evolution of the system:

$$(p_1^2 - p_2^2)\Psi = 2p\cdot v\Psi = (m_1^2 - m_2^2)\Psi, \tag{3.7}$$

the solution of which, for eigenfunctions of p_μ, is

$$\Psi(x_1,x_2) = e^{-ip\cdot X}\psi(x) = e^{-ip\cdot X}e^{-i(m_1^2-m_2^2)p\cdot x/(2p^2)}\psi(x^T), \tag{3.8}$$

and $\psi(x^T)$ defines a three-dimensional "internal" wave function.[20] Taking the sum of the two equations (3.1), we get the "eigenvalue" equation

$$[\tfrac{1}{4}p^2 - \tfrac{1}{2}(m_1^2 + m_2^2) + (1/4p^2)(m_1^2 - m_2^2)^2 + v^{T2} - V]\psi(x^T) = 0, \tag{3.9}$$

which is a three-dimensional Schrödinger-type equation.

The potential V, (3.4), can also be generalized so as to include nonlocal functions in x^T, but in this work we shall only consider potentials of the local form in x^T, (3.4). We assume V to be superficially Hermitian, in the sense that when the p_μ are replaced by real eigenvalues, then V is Hermitian in the usual L_2 norm.

We notice that the longitudinal operators $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$ commute with Eqs. (3.1), or with (3.7) and (3.9), which means that they have simultaneously Lorentz-invariant eigenvalues.[3] We assume that the physically acceptable potentials are those that lead to real eigenvalues of both $(\hat{p}\cdot p_1)^2$ and $(\hat{p}\cdot p_2)^2$, and that these in turn lead for each $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$ to two distinct eigenvalues with opposite signs. These different eigenvalues split the general Hilbert space $\mathscr{H}$ of normalizable solutions (also extended to Dirac's distributional sense) of Eqs. (3.1) into four subspaces, according to the signs of the eigenvalues of $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$. The physical Hilbert space $\mathscr{H}^*$ is defined as the subspace of $\mathscr{H}$ correspondng to the positive signs of both $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$:

$$\begin{aligned} &\hat{p}\cdot p_1 = \tfrac{1}{2}(p^2)^{1/2}(1 + (m_1^2 - m_2^2)/p^2), \quad p^2 > |m_1^2 - m_2^2|, \\ &\hat{p}\cdot p_2 = \tfrac{1}{2}(p^2)^{1/2}(1 - (m_1^2 - m_2^2)/p^2). \end{aligned} \tag{3.10}$$

This ensures, among other things, the positivity of $(p^2)^{1/2}$, and is a direct generalization of a similar condition in one-particle relativistic quantum mechanics.[8,13] One also requires that the norm of states in $\mathscr{H}^*$ be positive.

To construct the conserved tensor current of rank 2, $j_{\mu\nu}$, we shall follow the method already utilized in the one-particle case. We define $j_{\mu\nu}$ as the sum of two quantities:

$$j_{\mu\nu}(x_1,x_2) = j_{\mu\nu}^{(0)}(x_1,x_2) + j_{\mu\nu}^{(1)}(x_1,x_2), \tag{3.11}$$

where $j_{\mu\nu}^{(0)}$ is the "free" current,[13,21]

$$j_{\mu\nu}^{(0)}(x_1,x_2) = i^2\Psi_\varepsilon^*(x_1,x_2)\overleftrightarrow{\partial}_{1\mu}\overleftrightarrow{\partial}_{2\nu}\Phi_\varepsilon(x_1,x_2); \tag{3.12}$$

where we have considered the case of two eigenfunctions of

$i\nabla_\mu$, Ψ and Φ, with different eigenvalues p'_μ and p_μ, respectively; and where we have introduced, as in (2.5), a small negative imaginary part in the energies

$$\begin{aligned} &p_\mu \to p_\mu - i\varepsilon n_\mu, \quad n=(1,\vec{0}), \quad \varepsilon>0, \\ &p'_\mu \to p'_\mu - i\varepsilon n_\mu. \end{aligned} \tag{3.13}$$

A notation similar to (2.7) will also be used in the following:

$$V = V(x, p - i\varepsilon n, \ldots), \quad V' = V(x, p' - i\varepsilon n, \ldots), \tag{3.14}$$

where the transverse variables x^T, (3.4), are calculated with respect to $p_\mu - i\varepsilon n_\mu$ and $p'_\mu - i\varepsilon n_\mu$, respectively.

In view of the two conservation laws (1.1), $j^{(0)}_\mu$ satisfies the two equations

$$\begin{aligned} \partial^\mu_1 j^{(0)}_{\mu\nu} \equiv F_{1\nu} &= \Psi^*_\varepsilon [-(V'^* - V)\partial_{2\nu} + \overleftarrow{\partial}_{2\nu}(V'^* - V) \\ &\quad + (\partial_{2\nu}(V'^* + V))]\Phi_\varepsilon \\ &= \Psi^*_\varepsilon [(i/2)(V'^* - V)(p'_\nu + p_\nu) \\ &\quad + (V'^* - V)\partial_\nu - \overleftarrow{\partial}_\nu (V'^* - V) \\ &\quad - (\partial_\nu (V'^* + V))]\Phi_\varepsilon, \end{aligned} \tag{3.15a}$$

$$\begin{aligned} \partial^\nu_2 j^{(0)}_{\mu\nu} \equiv F_{2\mu} &= \Psi^*_\varepsilon [-(V'^* - V)\partial_{1\mu} + \overleftarrow{\partial}_{1\mu}(V'^* - V) \\ &\quad + (\partial_{1\mu}(V'^* + V))]\Phi_\varepsilon \\ &= \Psi^*_\varepsilon [(i/2)(V'^* - V)(p'_\mu + p_\mu) \\ &\quad - (V'^* - V)\partial_\mu + \overleftarrow{\partial}_\mu (V'^* - V) \\ &\quad + (\partial_\mu (V'^* + V))]\Phi_\varepsilon. \end{aligned} \tag{3.15b}$$

This means that $j^{(1)}_{\mu\nu}$, (3.1), must be a solution of the equations

$$\partial^\mu_1 j^{(1)}_{\mu\nu}(x_1,x_2) = -F_{1\nu}(x_1,x_2), \tag{3.16a}$$

$$\partial^\nu_2 j^{(1)}_{\mu\nu}(x_1,x_2) = -F_{2\mu}(x_1,x_2). \tag{3.16b}$$

These two equations should be compatible with each other, that is, we should find

$$\partial^\nu_2 F_{1\nu} = \partial^\mu_1 F_{2\mu}. \tag{3.17}$$

It can be checked that this equation is actually satisfied as a consequence of the compabitility of the two wave equations (3.1). One finds

$$\begin{aligned} \partial^\nu_2 F_{1\nu} = \partial^\mu_1 F_{1\mu} \equiv F(x_1,x_2) &= \Psi^*_\varepsilon \{ 2\overleftarrow{\partial}_\alpha (\partial^\alpha V'^*) + 2(\partial_\alpha V)\partial^\alpha \\ &\quad + (V'^* - V)[\tfrac{1}{4}(p - i\varepsilon n)^2 - \partial^2] \\ &\quad - [\tfrac{1}{4}(p' + i\varepsilon n)^2 - \overleftarrow{\partial}^2](V'^* - V) \\ &\quad + (\partial^2 (V'^* + V))\}\Phi_\varepsilon. \end{aligned} \tag{3.18}$$

The solution of Eqs. (3.16), which vanishes when the interaction is switched off, is

$$\begin{aligned} j^{(1)}_{\mu\nu}(x_1,x_2) = &-i\partial_{1\mu} \int G_A(x_1 - x'_1) F_{1\nu}(x'_1,x_2) d^4x'_1 \\ &- i\partial_{2\nu} \int G_A(x_2 - x'_2) F_{2\mu}(x_1,x'_2) d^4x'_2 \\ &- \partial_{1\mu}\partial_{2\nu} \int G_A(x_1 - x'_1) G_A(x_2 - x'_2) \\ &\times F(x'_1,x'_2) d^4x'_1\, d^4x'_2, \end{aligned} \tag{3.19}$$

where G_A has been defined in Eqs. (2.10) and (2.11).

If we factorize in $F_{1\nu}$, (3.15a), $F_{2\mu}$, (3.15b), and F, (3.18), the wave functions Ψ and Φ, we can easily check that the kernel of the current $j^{(1)}_{\mu\nu}$, and hence that of $j_{\mu\nu}$, (3.1), is translation invariant and is a pure Lorentz tensor of rank 2, as a consequence of the Poincaré invariance property of the potential V. These two features, together with the current conservations (1.1), ensure the Hermiticity properties of the Poincaré group generators and therefore the unitary realization of the whole group.

For the subsequent calculations we can use the factorization property of the F's in terms of the wave functions and make appropriate changes of variables in the integrals of the expression for $j^{(1)}_{\mu\nu}$, (3.19). We can write, using the total coordinate dependences of the wave functions (3.8),

$$\begin{aligned} &F_{a\nu}(x_1,x_2) = F_{a\nu}(X,x) = e^{i(p' - p + 2i\varepsilon n)\cdot X} \widetilde{F}_{a\nu}(x) \\ &\qquad\qquad (a = 1,2), \\ &F(x_1,x_2) = F(X,x) = e^{i(p' - p + 2i\varepsilon n)\cdot X} \widetilde{F}(x), \end{aligned} \tag{3.20}$$

where the expressions for $\widetilde{F}_{a\nu}(x)$ and $\widetilde{F}(x)$ can be obtained from (3.15), (3.18), and (3.8).

By using these expressions for $F_{a\nu}$ and F in (3.19), one can also write $j^{(1)}_{\mu\nu}$ in the form

$$\begin{aligned} j^{(1)}_{\mu\nu}(X,x) = &-i \int d^4x' (\partial_\mu G_A(x - x')) e^{-(i/2)(p' - p + 2i\varepsilon n)\cdot(x - x')} F_{1\nu}(X,x') \\ &+ i \int d^4x' (\partial_\nu G_A(x' - x)) e^{(i/2)(p' - p + 2i\varepsilon n)\cdot(x - x')} F_{2\mu}(X,x') \\ &+ 4 \int d^4X'\, d^4x' (\partial_\mu G_A(\tfrac{1}{2}x - \tfrac{1}{2}x' - X'))(\partial_\nu G_A(\tfrac{1}{2}x' - \tfrac{1}{2}x - X')) e^{i(p' - p + 2i\varepsilon n)\cdot X'} F(X,x'). \end{aligned} \tag{3.21}$$

In order to construct the scalar product relative to the wave functions Ψ and Φ, we shall choose for the surfaces Σ_1 and Σ_2 of formula (1.2) two parallel spacelike hyperplanes, perpendicular to a unit timelike four-vector n, which we choose, as usual, parallel to the x^0 axis (3.13):

$$n\cdot x_1 = t_1, \quad n = (1,\vec{0}), \quad n\cdot x_2 = t_2. \tag{3.22}$$

Then the scalar product is

$$(\Psi,\Phi) = \int j_{00}(X,x) d^3\vec{X}\, d^3\vec{x} = \int [\, j^{(0)}_{00}(X,x) + j^{(1)}_{00}(X,x)\,] d^3\vec{X}\, d^3\vec{x}. \tag{3.23}$$

We shall now concentrate on the second term of (3.23). Using expression (3.21), we have

$$\int j_{00}^{(1)}(X,x)d^3\vec{X}\,d^3\vec{x} = -i\int d^3\vec{X}\,d^3\vec{x}\,d^4x'(\partial_0 G_A(x-x'))e^{-(i/2)(p'-p+2i\varepsilon n)\cdot(x-x')}F_{10}(X,x')$$

$$+i\int d^3\vec{X}\,d^3\vec{x}\,d^4x'(\partial_0 G_A(x'-x))e^{(i/2)(p'-p+2i\varepsilon n)\cdot(x-x')}F_{20}(X,x')$$

$$+4\int d^3\vec{X}\,d^3\vec{x}\,d^4X'd^4x'(\partial_0 G_A(\tfrac{1}{2}x-\tfrac{1}{2}x'-X'))(\partial_0 G_A(\tfrac{1}{2}x'-\tfrac{1}{2}x-X'))e^{i(p'-p+2i\varepsilon n)\cdot X'}F(X,x'). \tag{3.24}$$

By using the X dependence of the F's (3.20), the integration with respect to $d^3\vec{X}$ yields the factor $(2\pi)^3\delta^3(\vec{p}-\vec{p}')$. In the first two integrals of (3.24) one can also perform the integration easily with respect to $d^3\vec{x}$, using the formula

$$\int \partial_0 G_A(x)d^3\vec{x} = i\theta(-x^0). \tag{3.25}$$

In the third integral the changes of variables

$$X'\to -X',\quad \vec{Y}=\vec{X}'-\tfrac{1}{2}\vec{x}+\tfrac{1}{2}\vec{x}',\quad \vec{y}=\tfrac{1}{2}\vec{x}-\tfrac{1}{2}\vec{x}'+\vec{X}'$$

are needed between $\vec{X}'$ and $\vec{x}$; then one uses formula (3.25) again. Finally one is left with an integral over X'^0, which is carried out in a straightforward way. One obtains

$$\int j_{00}^{(1)}(X,x)d^3\vec{X}\,d^3\vec{x} = (2\pi)^3\delta^3(\vec{p}'-\vec{p})e^{i(p_0'-p_0+2i\varepsilon)X^0}\left\{\int d^4x'\,\theta(x'^0-x^0)e^{(i/2)(p_0'-p_0+2i\varepsilon)(x^0-x'^0)}\tilde{F}_{10}(x')\right.$$

$$+\int d^4x'\,\theta(x^0-x'^0)e^{(i/2)(p_0'-p_0+2i\varepsilon)(x^0-x'^0)}\tilde{F}_{20}(x')$$

$$\left.+\frac{i}{(p_0'-p_0+2i\varepsilon)}\int d^4x'\,e^{(i/2)(p_0'-p_0+2i\varepsilon)|x^0-x'^0|}\tilde{F}(x')\right\}, \tag{3.26}$$

where the expressions for the $\tilde{F}$'s are defined in (3.20), (3.15), (3.18), and (3.8).

The remaining calculations are carried out by using the explicit expressions for the $\tilde{F}$'s. One mainly transfers, by partial integrations, the derivative operators acting on the potentials on the other terms of the integral, and one also uses the wave equations (3.1). One arrives at the expression

$$\int j_{00}^{(1)}(X,x)d^3\vec{X}\,d^3\vec{x} = (2\pi)^3\delta^3(\vec{p}'-\vec{p})e^{i(p_0'-p_0+2i\varepsilon)X^0}$$

$$\times\left\{\int d^4x'e^{(i/2)(p_0'-p_0+2i\varepsilon)|x^0-x'^0|}\frac{i}{4}(p_0'+p_0)\psi^*(x')(V'^*-V)\phi(x')\right.$$

$$+\frac{1}{2}\int d^4x'\,\varepsilon(x^0-x'^0)e^{(i/2)(p_0'-p_0+2i\varepsilon)|x^0-x'^0|}\psi^*(x')$$

$$\left.\times[(\partial_0'(V'^*+V))-(V'^*-V)\partial_0'+\overleftarrow{\partial}_0'(V'^*-V)]\phi(x')\right\}. \tag{3.27}$$

At this stage one uses the following properties and relations:

$$(i/2)(p_0'-p_0+2i\varepsilon)e^{(i/2)(p_0'-p_0+2i\varepsilon)|x^0-x'^0|} = -2\delta(x^0-x'^0)-\partial_0'\left[\varepsilon(x^0-x'^0)e^{(i/2)(p_0'-p_0+2i\varepsilon)|x^0-x'^0|}\right], \tag{3.28}$$

$$(p_0-i\varepsilon)\partial_0\phi = \vec{p}\cdot\vec{\partial}\phi-(i/2)(m_1^2-m_2^2)\phi, \tag{3.29}$$

$$(p_0-i\varepsilon)(\partial_0 V) = (\vec{p}\cdot\vec{\partial}V), \tag{3.30}$$

the latter two relations being simple transcriptions of relations (3.7) and (3.3), respectively. These relations yield surface terms that may be dropped. One finally obtains the following expression for the scalar product of two eigenfunctions Φ and Ψ of $i\nabla_\mu$ with different eigenvalues p_μ and p'_μ, respectively:

$$(\Psi_{p'},\Phi_p) = \int j_{00}(X,x)d^3\vec{X}\,d^3\vec{x}$$

$$= \int d^3\vec{X}\,d^3\vec{x}\,\Psi^*(X,x)\left\{i^2\overleftrightarrow{\partial}_{10}\overleftrightarrow{\partial}_{20}-(p_0'+p_0)\frac{[V(x,p'+i\varepsilon n,\ldots)-V(x,p-i\varepsilon n,\ldots)]}{(p_0'-p_0+2i\varepsilon)}\right\}\Phi(X,x)$$

$$= (2\pi)^3\delta^3(\vec{p}'-\vec{p})e^{i(p_0'-p_0+2i\varepsilon)X^0}\int d^3\vec{x}\,\psi^*(x)$$

$$\times\left\{\frac{1}{4}(p_0'+p_0)^2+(\overleftarrow{\partial}_0^2-2\overleftarrow{\partial}_0\partial_0+\partial_0^2)-(p_0'+p_0)\frac{[V(x,p'+i\varepsilon n,\ldots)-V(x,p-i\varepsilon n,\ldots)]}{(p_0'-p_0+2i\varepsilon)}\right\}\phi(x), \tag{3.31}$$

where the limit $\varepsilon = 0$ is understood.

By using the equations of motion and the property (3.29) and by dropping surface terms, one checks that this scalar product is independent of X^0 as it should be. This also shows that for $p'_0 \neq p_0$ (i.e., $p'^2 \neq p^2$) it vanishes, and hence the two wave functions are orthogonal.

To calculate the norm, we take $p'_0 = p_0$ and then the limit $\varepsilon = 0$. We obtain

$$
\begin{aligned}
&(\Psi_{p',a}, \Psi_{p,b})_{p'^2 = p^2} \\
&\quad = \int d^3\vec{X}\, d^3\vec{x}\, e^{ip'\cdot X} \psi_a^*(x) \left[i^2 \overleftrightarrow{\partial}_{10} \overleftrightarrow{\partial}_{20} - 4p_0^2 \frac{\partial V}{\partial p^2} \right] \\
&\qquad \times \psi_b(x) e^{-ip\cdot X} \\
&\quad = (2\pi)^3 2p_0 \delta^3(\vec{p}' - \vec{p}) \delta_{ab} f_a(p^2), \qquad (3.32)
\end{aligned}
$$

where the labels a,b distinguish different eigenfunctions with the same mass squared p^2. The origin and the meaning of the additional normalization factor $f_a(p^2)$ on the right-hand side of (3.32) will be explained below.

To check the independence of the scalar product on the relative time x^0, it is sufficient to consider the case of the norm (3.32), since for two different eigenvalues p'^2 and p^2 it is already zero for any value of $\vec{p}$. On deriving Eq. (3.32) with respect to x^0 one uses the properties (3.29) and (3.30) and converts the whole quantity into a surface term, which can be dropped, thus displaying the independence of the scalar product on the two times X^0 and x^0.

The expressions for the scalar product and the norm are simplified in the c.m. frame. In this frame the operators $i\partial_{10}$ and $i\partial_{20}$ become identical to the operators $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$, which have well-defined eigenvalues (3.10).

Furthermore, in this frame the transverse vector x^T reduces to $(0,\vec{x})$, which is independent of p^2. Therefore the only p^2 dependence of the potential V arises now from its explicit dependence on this variable in the c.m. frame, and no longer also from its implicit kinematic dependence through x^T.

The expressions for the internal parts of the scalar product (3.31) and the norm (3.32) then become, respectively,

$$
\begin{aligned}
&(\psi_{p'^2}, \phi_{p^2})_{\text{c.m.}} \\
&\quad = \int d^3\vec{x}\, \psi^*(x) \left\{ (p'_{10} + p_{10})(p'_{20} + p_{20}) - (p'_0 + p_0) \right. \\
&\qquad \left. \times \frac{[V(x,p'_0 + i\varepsilon,\ldots) - V(x,p'_0 - i\varepsilon,\ldots)]}{(p'_0 - p_0 + 2i\varepsilon)} \right\}_{\text{c.m.}} \phi(x), \qquad (3.33)
\end{aligned}
$$

$$
\begin{aligned}
(\psi_{p^2,a}, \psi_{p^2,b}) &= \int d^3\vec{x}\, \psi_a^*(x^T) \\
&\quad \times \left[4\hat{p}\cdot p_1 \hat{p}\cdot p_2 - 4p^2 \frac{\partial V}{\partial p^2} \right]_{\text{c.m.}} \psi_b(x^T) \\
&= 2p_0 f_a(p^2) \delta_{ab}. \qquad (3.34)
\end{aligned}
$$

We now turn to the interpretation of the normalization factor f in (3.32) and (3.34). If the Hilbert space of physical states is identified with the physical Hilbert space of two-particle relativistic quantum mechanics, then the completeness relation implies that $f = 1$. However, if the physical states $|p\rangle$ are supposed to belong to a larger space than that of two-particle quantum mechanics, such as in quantum field theory, then the two-particle Hilbert space is only a subspace of the total physical Hilbert space. A quantum field theory bound state $|p\rangle$ is not made up of only two constituents, but also contains contributions from multiconstituent subspaces. For this reason it is the global normalization condition of the quantum field theory state $|p\rangle$ that should fix the normalization factor of the two-particle quantum-mechanical wave function. This factor cannot be determined by quantum mechanics itself, but by its relationship with quantum field theory. This problem was solved by the author in a previous work,[5] where the connection of two-particle relativistic quantum mechanics with the Bethe–Salpeter equation was exhibited. In particular, the normalization condition of the quantum-mechanical wave function was obtained from that of the Bethe–Salpeter wave function in the c.m. frame and for the case of local potentials. It is this relation that fixes the normalization factor f of the norm (3.34).

Upon comparing formula (3.34) and formula (6.18) of Ref. 5 we reach two conclusions. First, the kernel of the norm in (6.18) of Ref. 5 is the same as in (3.34). This is not a trivial result. Formula (6.18) of Ref. 5 was obtained from a general field-theoretic normalization condition, mainly with the use of the "relativistic instantaneous approximation." It outlines the consistency of the latter approximation as a relativistic approximation for physical quantities. Second, it yields the expression for the normalization factor f,

$$
f = \left[\frac{\hat{p}\cdot p_1}{2(m_1^2 - \langle v^{T2} \rangle)^{1/2}} + \frac{\hat{p}\cdot p_2}{2(m_2^2 - \langle v^{T2} \rangle)^{1/2}} \right]^{-1}, \qquad (3.35)
$$

where $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$ are given by (3.10), and the mean value $\langle v^{T2} \rangle$ is calculated in the c.m. frame—in the L_2 norm, for instance. (It is an approximate value.)

Knowledge of the factor f, (3.35), is crucial when relating the quantum-mechanical wave function to physical quantities typical of field theory, such as decay coupling constants,[22] mainly in the case of nonperturbative interactions.

We end this section by examining the question of the positivity of the norm. Since the norm is Poincaré invariant, it is sufficient to examine this question in the c.m. frame. If the potential V is independent of p^2 there, then the kernels of the scalar product and of the norm, (3.33) and (3.34), become identical to those of the free expressions of two-particle relativistic quantum mechanics. In particular, if the eigenvalues of $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$ have been chosen positive, as in (3.10), then the norm is positive. This condition was actually imposed[3] as one of the defining conditions of the two-particle physical Hilbert space $\mathscr{H}^*$.

If V depends on p^2 in the c.m. frame, then the norm is no longer manifestly positive. However, one can still show, by a general argument, that it is actually positive.

This argument is based on results mentioned in Ref. 13, Sec. 1C concerning the properties of the so-called associated vectors, which may occur in the case of energy-dependent potentials. The appearance of such states in the theory has the following three consequences: (a) the occurrence of zero norm states; (b) the disappearance of the mass gap between the spectrum of the physical Hilbert space $\mathscr{H}^*$ and that of

the space of unphysical solutions; and (c) the occurrence of multiple poles in the Green's functions.

If the potential is characterized by a set of coupling constants $\{\alpha\}\equiv\{\alpha_1,\ldots,\alpha_n\}$, then the particular values $\{\overline{\alpha}\}$ that produce associated vectors delineate the frontier of the domain $\mathscr{D}$ of $\{\alpha\}$ inside which the physical Hilbert space $\mathscr{H}^*$ can be unambiguously defined. Behind this frontier, some eigenvalues of p^2 become negative or complex, and also complex norms appear.

Therefore the domain $\mathscr{D}$ of the values of the coupling constants $\{\alpha\}$ that permits the definition of a physical Hilbert space $\mathscr{H}^*$ (characterized by $\hat{p}\cdot p_1>0$, $\hat{p}\cdot p_2>0$ [Eqs. (3.10)] and a mass gap with respect to the space of unphysical solutions) ensures at the same time the strict positivity of the norms of physical states. The occurrence of nonpositive norms would contradict the fact that for $\{\alpha\}$ belonging to $\mathscr{D}$, all eigenvalues of p^2 are real and positive.

As a matter of example, we consider the case of the attractive Coulomb-like scalar potential

$$V=-\alpha m_1 m_2/2(-p^2x^{T2})^{1/2}\quad(\alpha>0). \tag{3.36}$$

The corresponding bound-state spectrum, obtained from Eq. (3.9), is given by the formulas

$$p^2_{\pm}=m_1^2+m_2^2\pm 2m_1m_2(1-\alpha^2/16(l+n+1)^2)^{1/2}, \tag{3.37}$$

where l and n are the usual orbital and radial quantum numbers. In order for p^2 to be real and positive, with a mass gap between p^2_+ and p^2_-, it is necessary that $\alpha<4$. The physical Hilbert space $\mathscr{H}^*$ corresponds to the positive square root of p^2_+, that is, to $\hat{p}\cdot p_1>0$ and $\hat{p}\cdot p_2>0$ (for any positive values of m_1 and m_2).

The norm of the internal part of the wave functions of physical states takes this form in the c.m. frame:

$$(\psi_{nlm},\psi_{nlm})=4m_1m_2(1-\alpha^2/16(l+n+1)^2)^{1/2}\times\int d^3\vec{x}\,\psi^*_{nlm}(x)\psi_{nlm}(x). \tag{3.38}$$

It is also positive for $\alpha<4$.

For $\alpha=4$, the norm of the ground state of $\mathscr{H}^*$ ($l=0$, $n=0$) becomes zero, thus indicating the presence of an associated vector. For this value of α, the eigenvalues of p^2 are still real and positive. But in this case, the mass gap between p^2_+ and p^2_- disappears.

For $\alpha>4$, some eigenvalues of p^2 become complex, and similarly for the norms of the corresponding states.

IV. PROPERTIES OF THE SCALAR PRODUCT CONSTRUCTED ON PARALLEL HYPERPLANES

The relative simplicity of the expression of the scalar product [(3.31) and (3.32)] as compared to the complicated form of the current $j_{\mu\nu}$ [(3.11), (3.12), and (3.19)] is not fortuitous. It is a consequence of the fact that we chose for the spacelike hypersurfaces Σ_1 and Σ_2, (1.2), two parallel hyperplanes perpendicular to the same constant timelike vector n, (3.22), and not, for instance, two unrelated hyperplanes defined by two different vectors n_1 and n_2. This feature actually simplifies the construction of the scalar product in that particular case, and it can be shown that here the exact knowledge of the current $j_{\mu\nu}$, satisfying the two conservation laws (1.1), is not necessary. It is sufficient to construct another current $\tilde{j}_{\mu\nu}$ satisfying a single conservation law,

$$\partial^\mu_1\tilde{j}_{\mu\nu}+\partial^\mu_2\tilde{j}_{\nu\mu}=0, \tag{4.1}$$

corresponding to the global translation invariance of the scalar product, in order to find expressions (3.31) and (3.32). The reason is that the difference between the current $\tilde{j}_{\mu\nu}$, (4.1), and the true current $j_{\mu\nu}$, (1.1), yields, for the choice of parallel hyperplanes (3.22), surface terms that then do not contribute in the expression of the scalar product. In this section we intend to show the validity of the above property.

Let us choose for the surfaces Σ_1 and Σ_2 of the scalar product (1.2) two parallel hyperplanes perpendicular to a constant timelike vector n, defined by Eqs. (3.22). The scalar product (1.2) can then be written as

$$(\Psi,\Phi)=\int d^4x_1\,d^4x_2\,\delta(n\cdot x_1-t_1)\times\delta(n\cdot x_2-t_2)n^\mu n^\nu j_{\mu\nu}(x_1,x_2). \tag{4.2}$$

We impose on the scalar product (global) translation invariance, that is, invariance under the changes $t_1\to t_1+n\cdot a$, $t_2\to t_2+n\cdot a$ (a_μ: a constant vector). This imposes on the covariant current $j_{\mu\nu}$, after dropping surface terms, the condition

$$\partial^\mu_1 j_{\mu\nu}+\partial^\mu_2 j_{\nu\mu}=0. \tag{4.3}$$

Let us then construct a current $\tilde{j}_{\mu\nu}$ that satisfies this equation [or (4.1)]. Because the current $j_{\mu\nu}$ satisfies two independent equations (1.1), it is clear that the single equation (4.3) has no unique solution (once the boundary conditions are fixed). We assume that one chooses among the solutions of Eq. (4.3) or (4.1) the one that has the simplest form. We construct the current $\tilde{j}_{\mu\nu}$ by the same methods as before, that is, by decomposing it into two terms,

$$\tilde{j}_{\mu\nu}=j^{(0)}_{\mu\nu}+\tilde{j}^{(1)}_{\mu\nu}, \tag{4.4}$$

the first term $j^{(0)}_{\mu\nu}$ corresponding to the "free" current (3.12), and the second term $\tilde{j}^{(1)}_{\mu\nu}$ to a remainder.

In general, $j^{(0)}_{\mu\nu}$ alone does not satisfy, except in the free case, Eq. (4.1). One then constructs $\tilde{j}^{(1)}_{\mu\nu}$ in such a way that it cancels the nonvanishing part of the quantity $\partial^\mu_1 j^{(0)}_{\mu\nu}+\partial^\mu_2 j^{(0)}_{\nu\mu}$. Because of the Poincaré invariance of the wave equations and their compatibility, $\tilde{j}^{(1)}_{\mu\nu}$ will have the following structure:

$$\tilde{j}^{(1)}_{\mu\nu}=\Psi^{**}K_{\mu\nu}{}^*\Phi, \tag{4.5}$$

where the kernel $K_{\mu\nu}$ is translation invariant and is a pure Lorentz tensor of rank 2; the stars on both sides of $K_{\mu\nu}$ mean that this kernel may involve integral operators. The translation invariance of $K_{\mu\nu}$ implies that the only X dependence of $\tilde{j}^{(1)}_{\mu\nu}$ is contained in the wave functions Ψ and Φ, and therefore, using their structure (3.8), $\tilde{j}^{(1)}_{\mu\nu}$ can also be written as

$$\tilde{j}^{(1)}_{\mu\nu}(X,x)=e^{i(p'-p+2i\varepsilon n)\cdot X}\psi^{**}K_{\mu\nu}{}^*\phi, \tag{4.6}$$

where, as usual, we have considered Ψ and Φ as eigenfunctions of $i\nabla_\mu$ with eigenvalues p'_μ and p_μ, respectively, and the limit $\varepsilon=0$ is to be taken at the end.

By specializing to the vector $n=(1,\vec{0})$, we consider the

scalar product built up from the current $\tilde{j}_{\mu\nu}$, (4.4)–(4.6) above:

$$(\Psi_{p'},\Phi_p)=\int d^3\vec{X}\,d^3\vec{x}\,\tilde{j}_{00}. \tag{4.7}$$

Because $\tilde{j}_{\mu\nu}$ satisfies Eq. (4.1), this scalar product is translation invariant and hence independent of X^0. Using the structure (4.6) of $\tilde{j}^{(1)}_{\mu\nu}$ and a similar structure of $j^{(0)}_{\mu\nu}$, one immediately deduces that for $p'^2\neq p^2$, the scalar product (4.7) is zero, which expresses the orthogonality of eigenfunctions of the total energy operator with different eigenvalues. This result is true for any value of the three-momentum $\vec{p}=\vec{p}'$.

The Lorentz invariance of the scalar product (4.7) requires also its independence on the relative time variable x^0. However, since the scalar product (4.7) is zero for different eigenvalues p'_0 and p_0 for any $\vec{p}$, it is then sufficient to verify this property on the norm alone. This involves the same function on both sides of its kernel and the same potential V through the kernel $K_{\mu\nu}$ (4.5). By using the properties (3.29) and (3.30) of Φ and V, one can then transform the derivative operator ∂_0, applying on the norm, into surface terms, which can then be dropped. Therefore the scalar product (4.7) is Poincaré invariant. Furthermore, the translation invariance and covariance property of the kernel of the current $\tilde{j}_{\mu\nu}$ ensures the Hermiticity property of the Poincaré group generators. The scalar product (4.7) is therefore identical to the one constructed from the current $j_{\mu\nu}$ satisfying the two conservation laws (1.1).

In the rest of this section we shall construct explicitly a current $\tilde{j}_{\mu\nu}$ and the resulting scalar product.

From Eqs. (3.15) we deduce

$$\begin{aligned}\partial^\mu_1 j^{(0)}_{\mu\nu}+\partial^\mu_2 j^{(0)}_{\nu\mu}&=F_{1\nu}+F_{2\nu}\\&=\Psi^*_\varepsilon i(V'^*-V)(p'_\nu+p_\nu)\Phi_\varepsilon.\end{aligned} \tag{4.8}$$

Then $\tilde{j}^{(1)}_{\mu\nu}$ must satisfy the equation

$$\partial^\mu_1 \tilde{j}^{(1)}_{\mu\nu}+\partial^\mu_2 \tilde{j}^{(1)}_{\nu\mu}=-\Psi^*_\varepsilon i(V'^*-V)(p'_\nu+p_\nu)\Phi_\varepsilon. \tag{4.9}$$

Among the solutions of this equation, one can choose those which are symmetric tensors in μ,ν:

$$\tilde{j}^{(1)}_{\mu\nu}(x_1,x_2)=\tilde{j}^{(1)}_{\nu\mu}(x_1,x_2). \tag{4.10}$$

Then the left-hand side of Eq. (4.9) becomes

$$\partial^\mu_1 \tilde{j}^{(1)}_{\mu\nu}+\partial^\mu_2 \tilde{j}^{(1)}_{\mu\nu}=(\partial^\mu_1+\partial^\mu_2)\tilde{j}^{(1)}_{\mu\nu}=\nabla^\mu\tilde{j}^{(1)}_{\mu\nu}. \tag{4.11}$$

Taking into account the fact that the total coordinate X is contained, on the right-hand side of Eq. (4.9), in the wave functions alone, one easily obtains the following solution, which vanishes when the interaction is switched off:

$$\tilde{j}^{(1)}_{\mu\nu}=-\Psi^*_\varepsilon(p'+p)_\mu(p'+p)_\nu \times\frac{(V'^*-V)}{(p'^2-p^2+2i\varepsilon(p'_0+p_0))}\Phi_\varepsilon. \tag{4.12}$$

The scalar product constructed from $\tilde{j}_{\mu\nu}$, (4.4), is then

$$\begin{aligned}(\Psi_{p'},\Phi_p)&=\int d^3\vec{X}\,d^3\vec{x}\,\tilde{j}_{00}(X,x)\\&=\int d^3\vec{X}\,d^3\vec{x}\,\Psi^*(X,x)\left\{i^2\overleftrightarrow{\partial}_{10}\overleftrightarrow{\partial}_{20}-(p'_0+p_0)\right.\\&\left.\times\frac{[V(x,p'+i\varepsilon n,\ldots)-V(x,p-i\varepsilon n,\ldots)]}{(p'_0-p_0+2i\varepsilon)}\right\}\Phi(X,x),\end{aligned} \tag{4.13}$$

which coincides with the scalar product (3.31), and a similar conclusion is reached for the norm (3.32). (We have used the fact that because of the $d^3\vec{X}$ integration, $\vec{p}=\vec{p}'$ and hence $p'^2_0-p^2_0=p'^2-p^2$.)

In conclusion, if one builds up the two-particle scalar product on parallel hyperplanes defined by a common constant timelike vector n, then the knowledge of a current satisfying one conservation law, that of translation invariance (4.1), is sufficient to find the correct result. This property considerably simplifies the corresponding calculations and yields the scalar product in an almost straightforward way.

In the two following sections, where we deal with systems involving fermions, we shall continue, for the sake of generality, to present both methods of calculation.

V. SPIN-½ FERMION–ANTIFERMION SYSTEMS

The fermion–antifermion wave function satisfies two independent wave equations[3] which are generalizations of the Dirac equations relative to the fermion (particle 1) and to the antifermion (particle 2), and are the analogs of Eqs. (3.1) of the spin-0 case:

$$H_1\Psi\equiv[\gamma\cdot p_1-m_1-(-\eta\cdot p_2+m_2)V]\Psi=0, \tag{5.1a}$$

$$H_2\Psi\equiv[\eta\cdot p_2+m_2+(\gamma\cdot p_1+m_1)V]\Psi=0, \tag{5.1b}$$

where the wave function Ψ is a 16-component spinor of rank 2:

$$\Psi=\Psi_{\alpha_1\alpha_2}(x_1,x_2)\quad(\alpha_1,\alpha_2=1,\ldots,4). \tag{5.2}$$

The matrices γ and η are the Dirac matrices acting on the fermion and antifermion spinor indices, respectively (labeled by subindices 1 and 2):

$$\begin{aligned}\gamma_\mu\Psi&\equiv\gamma_{1\mu}\Psi=(\gamma_\mu)_{\alpha_1\beta_1}\Psi_{\beta_1\alpha_2},\\ \eta_\mu\Psi&\equiv\Psi\gamma_{2\mu}=\Psi_{\alpha_1\beta_2}(\gamma_\mu)_{\beta_2\alpha_2};\end{aligned} \tag{5.3}$$

we shall also define

$$\sigma_{\mu\nu}=(1/2i)[\gamma_\mu,\gamma_\nu],\quad \xi_{\mu\nu}=(1/2i)[\eta_\mu,\eta_\nu]. \tag{5.4}$$

(The γ and η matrices commute.)

The potential V is a Poincaré-invariant function of the coordinates, momenta, and Dirac matrices. The compatibility condition (3.2) of the wave equations requires that V depend on the relative coordinates x through the transverse components x^T alone[3] (3.5):

$$V=V(x^T,p_1,p_2,\gamma,\eta) \tag{5.5}$$

[V satisfying Eq. (3.3)].

Equations (5.1) completely determine the longitudinal, relative coordinate x_L dependence of the wave function

through Eq. (3.7), which here is a consequence of Eqs. (5.1) and whose solution is, for eigenfunctions of the total momentum p, given by the decomposition[20] (3.8).

In the following we assume that V is superficially Hermitian, so that

$$\overline{V} \equiv \gamma_0\eta_0 V^\dagger \gamma_0\eta_0 = V \tag{5.6}$$

in the usual L_2 norm and when the p_μ in V are replaced by real eigenvalues.

In order to construct the current $j_{\mu\nu}$ satisfying the two conservation laws (1.1), we again decompose it into two parts (3.11), where now the trial current $j^{(0)}_{\mu\nu}$ is not chosen as the "free" current but as a modified one that proves more convenient in the following:

$$j^{(0)}_{\mu\nu} = \mathrm{Tr}[\overline{\Psi}_\varepsilon \gamma_\mu \eta_\nu \Phi_\varepsilon - \overline{\Psi}_\varepsilon \overline{V'^*} \gamma_\mu \eta_\nu V \Phi_\varepsilon], \tag{5.7}$$

where the trace bears on the spinor indices, $\overline{\Psi} = [\gamma_0\eta_0\Psi]^\dagger$, and the notation V' is as in (3.14); the complex conjugation of V' comes from the presence of the ε term in the energy eigenvalues (to be taken to the limit zero at the end). Because of (5.6), $\overline{V^{*\prime}} = V^{*\prime}$. For the sake of simplicity in notation we shall omit, until the final formulas, the function "Tr" from the various expressions, but it must be understood that the trace is always taken.

The current $j^{(0)}_{\mu\nu}$ satisfies the two equations

$$\begin{aligned}\partial^\mu_1 j^{(0)}_{\mu\nu} \equiv F_{1\nu} &= \overline{\Psi}_\varepsilon [(V'^* - V)\partial_{2\nu} - \overleftarrow{\partial}_{2\nu}(V'^* - V) - (\partial_{2\nu}(V'^* + V))]\Phi_\varepsilon - i\partial^\alpha_2 [\overline{\Psi}_\varepsilon (\xi_{\nu\alpha} V - V'^* \xi_{\nu\alpha})\Phi_\varepsilon] \\ &= \overline{\Psi}_\varepsilon [-(i/2)(V'^* - V)(p'_\nu + p_\nu) - (V'^* - V)\partial_\nu + \overleftarrow{\partial}_\nu (V'^* - V) + (\partial_\nu (V'^* + V))]\Phi_\varepsilon \\ &\quad - i\partial^\alpha_2 [\overline{\Psi}_\varepsilon (\xi_{\nu\alpha} V - V'^* \xi_{\nu\alpha})\Phi_\varepsilon], \end{aligned} \tag{5.8a}$$

$$\begin{aligned}\partial^\nu_2 j^{(0)}_{\mu\nu} \equiv F_{2\mu} &= \overline{\Psi}_\varepsilon [(V'^* - V)\partial_{1\mu} - \overleftarrow{\partial}_{1\mu}(V'^* - V) - (\partial_{1\mu}(V'^* + V))]\Phi_\varepsilon - i\partial^\alpha_1 [\overline{\Psi}_\varepsilon (\sigma_{\mu\alpha} V - V'^* \sigma_{\mu\alpha})\Phi_\varepsilon] \\ &= \overline{\Psi}_\varepsilon [-(i/2)(V'^* - V)(p'_\mu + p_\mu) + (V'^* - V)\partial_\mu - \overleftarrow{\partial}_\mu (V'^* - V) - (\partial_\mu (V'^* + V))]\Phi_\varepsilon \\ &\quad - i\partial^\alpha_1 [\overline{\Psi}_\varepsilon (\sigma_{\mu\alpha} V - V'^* \sigma_{\mu\alpha})\Phi_\varepsilon]. \end{aligned} \tag{5.8b}$$

The remaining current $j^{(1)}_{\mu\nu}$, (3.11), must then satisfy the two equations

$$\partial^\mu_1 j^{(1)}_{\mu\nu} = -F_{1\nu}, \tag{5.9a}$$

$$\partial^\nu_2 j^{(1)}_{\mu\nu} = -F_{2\mu}, \tag{5.9b}$$

and be a solution of them.

The two equations (5.9) are compatible among themselves (i.e., integrable). One finds

$$\begin{aligned}\partial^\nu_2 F_{1\nu} = \partial^\mu_1 F_{2\mu} \equiv F(x_1, x_2) &= -\overline{\Psi}_\varepsilon \{2\overleftarrow{\partial}_\alpha (\partial^\alpha V'^*) + 2(\partial_\alpha V)\partial^\alpha + (V'^* - V)[\tfrac{1}{4}(p - i\varepsilon n)^2 - \partial^2] \\ &\quad - [\tfrac{1}{4}(p' + i\varepsilon n)^2 - \overleftarrow{\partial}^2](V'^* - V) + (\partial^2 (V'^* + V))\}\Phi_\varepsilon. \end{aligned} \tag{5.10}$$

We notice that, except for a global minus sign, this expression is formally the same as in the spin-0 case (3.18). Furthermore, the expressions of $F_{1\nu}$, (5.8a), and $F_{2\mu}$, (5.8b), are also formally the same as in the spin-0 case, Eqs. (3.15), except for the last total derivative terms involving the spin matrices, and a global minus sign.

The solution of Eqs. (5.9), which vanishes when the interaction is switched off, is given by formula (3.19). The construction of the scalar product follows similar lines as in the spin-0 case. We choose for the surfaces Σ_1 and Σ_2 of Eq. (1.2) the two parallel hyperplanes defined by Eqs. (3.22). Because of the antisymmetry of the spin operators $\sigma_{\nu\alpha}$ and $\xi_{\nu\alpha}$, the total derivative terms [present in $F_{1\nu}$ and $F_{2\mu}$, (5.8a) and (5.8b)] will lead to surface terms in the scalar product, and therefore their contributions may be dropped. This means that the contribution of $j^{(1)}_{\mu\nu}$ to the scalar product will be formally the same as in the spin-0 case, up to a global minus sign.

One finally gets for the scalar product

$$\begin{aligned}(\Psi_{p'}, \Phi_p) &= \int j_{00}(X,x)\, d^3\vec{X}\, d^3\vec{x} \\ &= \int d^3\vec{X}\, d^3\vec{x}\, \mathrm{Tr}\left\{\overline{\Psi}(X,x)\left[\gamma_0\eta_0 - V'^*\gamma_0\eta_0 V + (p'_0 + p_0)\frac{[V(x,p' + i\varepsilon n,\ldots) - V(x,p - i\varepsilon n,\ldots)]}{(p'_0 - p_0 + 2i\varepsilon)}\right]\Phi(X,x)\right\} \\ &= (2\pi)^3 \delta^3(\vec{p}' - \vec{p}) e^{i(p'_0 - p_0 + 2i\varepsilon)X^0} \int d^3\vec{x}\, \mathrm{Tr}\left\{\overline{\psi}(x)\left[\gamma_0\eta_0 - V'^*\gamma_0\eta_0 V \right.\right. \\ &\quad \left.\left. + (p'_0 + p_0)\frac{[V(x,p' + i\varepsilon n,\ldots) - V(x,p - i\varepsilon n,\ldots)]}{(p'_0 - p_0 + 2i\varepsilon)}\right]\phi(x)\right\}, \end{aligned} \tag{5.11}$$

the limit $\varepsilon = 0$ being understood.

For the norm one obtains

$$(\Psi_{p',a},\Psi_{p,b})_{p'^2=p^2}$$
$$= (2\pi)^3\delta^3(\vec{p}'-\vec{p})\int d^3\vec{x}\,\mathrm{Tr}$$
$$\times\left\{\bar{\psi}_a(x)\left[\gamma_0\eta_0 - V\gamma_0\eta_0 V + 4p_0^2\frac{\partial V}{\partial p^2}\right]\psi_b(x)\right\}$$
$$= (2\pi)^3 2p_0\delta^3(\vec{p}'-\vec{p})\delta_{ab}f_a(p^2). \tag{5.12}$$

The X^0 independence of the scalar product (5.11) can be checked by using the wave equations (5.1). In particular, these can be combined to yield the following equation:

$$(1+\gamma_0\eta_0 V)p_0\Psi = (m_1\gamma_0 - m_2\eta_0 - \gamma_0\vec{\gamma}\cdot\vec{v} + \eta_0\vec{\eta}\cdot\vec{v})$$
$$\times(1-\gamma_0\eta_0 V)\Psi, \tag{5.13}$$

which is useful for the above demonstration.

The x^0 independence concerns mainly the norm (5.12), since the scalar product is zero for different eigenvalues p'^2 and p^2, for any $\vec{p}$. One then uses the properties (3.29) and (3.30) of the wave function and of the potential to show the x^0 independence of the norm.

The normalization factor f of the norm (5.12), which has a field-theoretical origin, can be obtained by comparing formula (5.12) to that of the normalization condition of the Bethe–Salpeter equation derived within the framework leading to local potentials. The corresponding expression was given in Ref. 5, formula (6.29), in the c.m. frame. One deduces that f has the same expression, (3.35), as in the spin-0 case.

If the potential V is explicitly independent of p^2 in the c.m. frame, the expression of the norm (5.12) shows that its kernel still depends on V. This implies that V must satisfy some inequality conditions to guarantee the positivity of the norm. This question was examined in more detail in Ref. 3, Sec. VII A. In this case one can also apply the transformation

$$\Psi = [1-(\gamma\cdot\hat{p}\eta\cdot\hat{p}V)^2]^{-1/2}\Psi'$$

on the wave function in order to transform the kernel of the norm to that of the "free" norm[3] in the c.m. frame.

We end this section by constructing the scalar product (5.11) by the method described in Sec. IV, that is, by constructing a current $\tilde{j}_{\mu\nu}$, (4.4), satisfying the conservation law (4.1). We take for $j^{(0)}_{\mu\nu}$ the same current as in (5.7). One then obtains

$$\partial_1^\mu j^{(0)}_{\mu\nu} + \partial_2^\mu j^{(0)}_{\nu\mu}$$
$$= -i\bar{\Psi}_\varepsilon(V'^* - V)(p'_\nu + p_\nu)\Phi_\varepsilon$$
$$- i\partial_2^\alpha(\bar{\Psi}_\varepsilon A_{\nu\alpha}\Phi_\varepsilon) - i\partial_1^\alpha(\bar{\Psi}_\varepsilon B_{\nu\alpha}\Phi_\varepsilon), \tag{5.14}$$

where

$$A_{\nu\alpha} = (\xi_{\nu\alpha}V - V'^*\xi_{\nu\alpha}),$$
$$B_{\nu\alpha} = (\sigma_{\nu\alpha}V - V'^*\sigma_{\nu\alpha}). \tag{5.15}$$

The first term on the right-hand side of (5.14) is the same, up to a minus sign, as the one appearing in the spin-0 case (4.8). Therefore its contribution to the scalar product will be analogous.

The remaining current $\tilde{j}^{(1)}_{\mu\nu}$, (4.4), must satisfy the equation

$$\partial_1^\mu \tilde{j}^{(1)}_{\mu\nu} + \partial_2^\mu \tilde{j}^{(1)}_{\nu\mu} = i\bar{\Psi}_\varepsilon(V'^* - V)(p'_\nu + p_\nu)\Phi_\varepsilon$$
$$+ i\partial_2^\alpha f_{1\nu\alpha} + i\partial_1^\alpha f_{2\nu\alpha}, \tag{5.16}$$

where

$$f_{1\nu\alpha} = \bar{\Psi}_\varepsilon A_{\nu\alpha}\Phi_\varepsilon, \quad f_{2\nu\alpha} = \bar{\Psi}_\varepsilon B_{\nu\alpha}\Phi_\varepsilon, \tag{5.17}$$

A simple solution of Eq. (5.16) is

$$\tilde{j}^{(1)}_{\mu\nu} = \bar{\Psi}_\varepsilon(p'_\mu + p_\mu)(p'_\nu + p_\nu)$$
$$\times\frac{(V'^* - V)}{(p'^2 - p^2 + 2i\varepsilon(p'_0 + p_0))}\Phi_\varepsilon$$
$$- \partial_{1\mu}\int G_A(x_1 - x'_1)\partial_2^\alpha f_{1\nu\alpha}(x'_1, x_2)d^4x'_1$$
$$- \partial_{2\nu}\int G_A(x_2 - x'_2)\partial_1^\alpha f_{2\mu\alpha}(x_1, x'_2)d^4x'_2, \tag{5.18}$$

where G_A is defined in (2.10) and (2.11).

Because of the antisymmetry of the tensors $f_{1\nu\alpha}$ and $f_{2\nu\alpha}$ and the presence of the total derivatives ∂_2^α and ∂_1^α in the last two terms of $j^{(1)}_{\mu\nu}$, (5.18), these will yield only surface terms in the corresponding scalar product and therefore will not contribute in it. Consequently one finds the same expression (5.11) for the scalar product.

VI. SPIN-½–SPIN-0 PARTICLE SYSTEMS

In this section we consider systems composed of one spin-½ fermion (particle 1) and one spin-0 boson (particle 2). The wave function is a four-component spinor,

$$\Psi = \Psi_\alpha(x_1, x_2) \quad (\alpha = 1,\ldots,4), \tag{6.1}$$

and satisfies two independent wave equations[3] that are generalizations of the Dirac and Klein–Gordon equations, respectively:

$$H_1\Psi \equiv (\gamma\cdot p_1 - m_1 - V)\Psi = 0, \tag{6.2a}$$
$$H_2\Psi \equiv [p_2^2 - m_2^2 - (\gamma\cdot p_1 + m_1)V]\Psi = 0. \tag{6.2b}$$

The potential V is a Poincaré-invariant function of the coordinates, momenta, and Dirac matrices. The compatibility condition (3.2) of the wave equations requires that V depend on the relative coordinates x through the transverse components x^T alone[3] (3.5):

$$V = V(x^T, p_1, p_2, \gamma) \tag{6.3}$$

[V satisfying Eq. (3.3)].

Equations (6.2) completely determine the longitudinal, relative coordinate x_L dependence of the wave function through Eq. (3.7), which is a consequence of Eqs. (6.2) here, and whose solution is, for eigenfunctions of the total momentum p, given by the decomposition (3.8).

In the following we assume that V is superficially Hermitian, so that

$$\bar{V} \equiv \gamma_0 V^\dagger \gamma_0 = V \tag{6.4}$$

in the usual L_2 norm and when the p_μ in V are replaced by real eigenvalues.

In order to construct the current $j_{\mu\nu}$ satisfying the two conservation laws (1.1), we decompose it into two parts, (3.11), where $j^{(0)}_{\mu\nu}$ is chosen as being the "free" current:

$$j^{(0)}_{\mu\nu} = i\bar{\Psi}_\varepsilon \gamma_\mu \overleftrightarrow{\partial}_{2\nu} \Phi_\varepsilon, \tag{6.5}$$

where $\bar{\Psi} = \Psi^\dagger \gamma_0$.

The current $j^{(0)}_{\mu\nu}$ satisfies the two equations

$$\begin{aligned}\partial^\mu_1 j^{(0)}_{\mu\nu} \equiv F_{1\nu} &= \bar{\Psi}_\varepsilon [-(V'^* - V)\partial_{2\nu} + \overleftarrow{\partial}_{2\nu}(V'^* - V) + (\partial_{2\nu}(V'^* + V))]\Phi_\varepsilon \\ &= \bar{\Psi}_\varepsilon [(i/2)(V'^* - V)(p'_\nu + p_\nu) + (V'^* - V)\partial_\nu - \overleftarrow{\partial}_\nu (V'^* - V) - (\partial_\nu (V'^* + V))]\Phi_\varepsilon, \end{aligned} \tag{6.6a}$$

$$\begin{aligned}\partial^\nu_2 j^{(0)}_{\mu\nu} \equiv F_{2\mu} &= \bar{\Psi}_\varepsilon [-(V'^* - V)\partial_{1\mu} + \overleftarrow{\partial}_{1\mu}(V'^* - V) + (\partial_{1\mu}(V'^* + V))]\Phi_\varepsilon - i\partial^\alpha_1 [\bar{\Psi}_\varepsilon (V'^* \sigma_{\mu\alpha} - \sigma_{\mu\alpha} V)\Phi_\varepsilon] \\ &= \bar{\Psi}_\varepsilon [(i/2)(V'^* - V)(p'_\mu + p_\mu) - (V'^* - V)\partial_\mu + \overleftarrow{\partial}_\mu (V'^* - V) + (\partial_\mu (V'^* + V))]\Phi_\varepsilon \\ &\quad - i\partial^\alpha_1 [\bar{\Psi}_\varepsilon (V'^* \sigma_{\mu\alpha} - \sigma_{\mu\alpha} V)\Phi_\varepsilon]. \end{aligned} \tag{6.6b}$$

Compared to Eqs. (3.15) we notice that $F_{1\nu}$ has the same formal expression as in the spin-0 case, while $F_{2\mu}$ differs only by the last total derivative term.

Now $F_{1\nu}$ and $F_{2\mu}$ also satisfy the compatibility equation

$$\begin{aligned}\partial^\nu_2 F_{1\nu} = \partial^\mu_1 F_{2\mu} \equiv F(x_1,x_2) = \bar{\Psi}_\varepsilon \{ &2\overleftarrow{\partial}_\alpha (\partial^\alpha V'^*) + 2(\partial_\alpha V)\partial^\alpha + (V'^* - V)[\tfrac{1}{4}(p - i\varepsilon n)^2 - \partial^2] \\ &- [\tfrac{1}{4}(p' + i\varepsilon n)^2 - \overleftarrow{\partial}^2](V'^* - V) + (\partial^2 (V'^* + V))\}\Phi_\varepsilon, \end{aligned} \tag{6.7}$$

where formally F has the same expression as in the spin-0 case (3.18).

The remaining current $j^{(1)}_{\mu\nu}$ has the expression (3.19) in terms of the F's. The construction of the scalar product follows similar lines as in the spin-0 case. We choose for the surfaces Σ_1 and Σ_2 of Eq. (1.2) the two parallel hyperplanes defined by Eqs. (3.22). Because of the antisymmetry of the spin operator $\sigma_{\mu\alpha}$, the total derivative term present in $F_{2\mu}$, (6.6b), leads to surface terms in the scalar product, and therefore their contributions may be dropped. This means that the contribution of $j^{(1)}_{\mu\nu}$ to the scalar product will be formally the same as in the spin-0 case.

One finally gets for the scalar product

$$\begin{aligned}(\Psi_{p'}, \Phi_p) &= \int j_{00}(X,x) d^3\vec{X}\, d^3\vec{x} \\ &= \int d^3\vec{X}\, d^3\vec{x}\, \bar{\Psi}(X,x) \left\{ i\gamma_0 \overleftrightarrow{\partial}_{20} - (p'_0 + p_0) \frac{[V(x,p' + i\varepsilon n,\ldots) - V(x,p - i\varepsilon n,\ldots)]}{(p'_0 - p_0 + 2i\varepsilon)} \right\} \Phi(X,x) \\ &= (2\pi)^3 \delta^3(\vec{p}' - \vec{p}) e^{i(p'_0 - p_0 + 2i\varepsilon)X^0} \int d^3\vec{x}\, \bar{\psi}(x) \\ &\quad \times \left\{ \frac{1}{2}(p'_0 + p_0)\gamma_0 - i\overleftrightarrow{\partial}_0 \gamma_0 - (p'_0 + p_0) \frac{[V(x,p' + i\varepsilon n,\ldots) - V(x,p - i\varepsilon n,\ldots)]}{(p'_0 - p_0 + 2i\varepsilon)} \right\} \phi(x), \end{aligned} \tag{6.8}$$

where the limit $\varepsilon = 0$ is understood.

For the norm one obtains

$$\begin{aligned}&(\Psi_{p',a}, \Psi_{p,b})_{p'^2 = p^2} \\ &= (2\pi)^3 \delta^3(\vec{p}' - \vec{p}) \int d^3\vec{x}\, \bar{\psi}_a(x) \\ &\quad \times \left[\gamma_0 p_0 - i\gamma_0 \overleftrightarrow{\partial}_0 - 4p_0^2 \frac{\partial V}{\partial p^2} \right] \psi_b(x) \\ &= (2\pi)^3 2p_0 \delta^3(\vec{p}' - \vec{p}) \delta_{ab} f_a(p^2). \end{aligned} \tag{6.9}$$

The X^0 independence of the scalar product (6.8) can be checked by using the wave equation (6.2a) and the property (3.29), and by dropping surface terms. The x^0 independence of the norm is checked by using the properties (3.29) and (3.30).

The normalization factor f is obtained by comparing formula (6.9) with the expression of the norm obtained from the Bethe–Salpeter equation in the c.m. frame. It is given in Ref. 5, Eq. (6.41). One finds that f has the same expression as in the spin-0 case, Eq. (3.35).

The construction of the scalar product by the method of Sec. IV follows similar lines as in the fermion–antifermion case. One obtains

$$\begin{aligned}\partial^\mu_1 j^{(0)}_{\mu\nu} + \partial^\mu_2 j^{(0)}_{\nu\mu} &= i\bar{\Psi}_\varepsilon (V'^* - V)(p'_\nu + p_\nu)\Phi_\varepsilon \\ &\quad + i\partial^\alpha_1 (\bar{\Psi}_\varepsilon B_{\nu\alpha} \Phi_\varepsilon), \end{aligned} \tag{6.10}$$

$B_{\nu\alpha}$ being given by (5.15).

The term $\tilde{j}^{(1)}_{\mu\nu}$ is then given by

$$\begin{aligned}\tilde{j}^{(1)}_{\mu\nu} = &-\bar{\Psi}_\varepsilon (p'_\mu + p_\mu)(p'_\nu + p_\nu) \\ &\times \frac{(V'^* - V)}{(p'^2 - p^2 + 2i\varepsilon(p'_0 + p_0))} \Phi_\varepsilon \\ &+ \partial_{2\nu} \int G_A(x_2 - x'_2) \partial^\alpha_1 f_{2\mu\alpha}(x_1, x'_2) d^4x'_2, \end{aligned} \tag{6.11}$$

$f_{2\mu\alpha}$ being defined by (5.17). One finally finds the same expression (6.8) for the scalar product.

In the c.m. frame the expression for the kernel of the norm (6.9) becomes simplified, since the operator $i\partial_{20}$ be-

comes identical to the operator $\hat{p}\cdot p_2$, which has a well-defined eigenvalue given by (3.10).

The comments made at the end of Sec. III about the positivity of the norm in the spin-0 case remain valid in the present case as well.

VII. CONCLUSION

We constructed in this paper the scalar product corresponding to the Hilbert space of states in two-particle relativistic quantum mechanics. The method adopted here was that of finding a tensor current of rank 2, $j_{\mu\nu}(x_1,x_2)$, satisfying two independent conservation laws relative to particles 1 and 2, respectively. Then the scalar product was obtained by integrating $j_{\mu\nu}$ over two three-dimensional spacelike hypersurfaces defined by the elementry areas $d\sigma_1^{\mu}(x_1)$ and $d\sigma_2^{\nu}(x_2)$.

The two conservation laws of $j_{\mu\nu}$ ensure the Poincaré invariance of the scalar product. The fact that the kernel of the current $j_{\mu\nu}$ is translation invariant and covariant then ensures the Hermiticity property of the Poincaré group generators and the unitary realization of the group.

A simple expression for the scalar product is obtained when one chooses for the two integration hypersurfaces two constant parallel hyperplanes. What is new in the kernel of the scalar product is essentially the presence of terms representing the total momentum dependences of the interaction potential V. One also finds, by a limiting procedure, the expression for the norm of states. In order to be in contact with field-theoretic quantities, the normalization constant of the norm is fixed by comparison with the normalization condition of the Bethe–Salpeter wave function through the link between two-particle relativistic quantum mechanics and field theory.[5] It was also shown that the positivity of the norm can, in general, be ensured if the spectrum of the eigenvalues of the total mass squared operator comes out to be positive. Furthermore, the choice of positive eigenvalues for each of the longitudinal momenta $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$ enters into the definition of the physical Hilbert space $\mathscr{H}^*$.

The construction of a Poincaré-invariant scalar product and the corresponding physical Hilbert space, together with the construction of Poincaré-invariant wave equations and the unitary realization of the Poincaré group, provides the basis for a consistent formulation and treatment of two-particle relativistic quantum mechanics in the interacting case. The connection established between this framework and the Bethe–Salpeter equation[5] permits, in addition, the evaluation of physical quantities having a field-theoretic interpretation.

[1] *Relativistic action at a distance: classical and quantum aspects, Proceedings of the Barcelona Workshop 1981, Lecture Notes in Physics,* Vol. 162, edited by J. Llosa (Springer, Berlin, 1982).

[2] *Constraint's Theory and Relativistic Dynamics, Proceedings of the Firenze Workshop 1986,* edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987).

[3] H. Sazdjian, Phys. Rev. D **33**, 3401 (1986).

[4] The Poincaré invariance of the scalar product is not enough to ensure the Poincaré invariance of the theory. One must also have Poincaré-invariant wave equations. In the one-particle case, this happens only in free theories.

[5] H. Sazdjian, J. Math. Phys. **28**, 2618 (1987).

[6] G. P. Lepage, Phys. Rev. A **16**, 863 (1977).

[7] W. E. Caswell and G. P. Lepage, Phys. Rev. A **18**, 810 (1978).

[8] H. Feshbach and F. Villars, Rev. Mod. Phys. **30**, 24 (1958).

[9] H. Snyder and J. Weinberg, Phys. Rev. **57**, 307 (1940); L. I. Schiff, H. Snyder, and J. Weinberg, *ibid.* **57**, 315 (1940).

[10] B. Schroer and J. A. Swieca, Phys. Rev. D **2**, 2938 (1970).

[11] L. E. Lundberg, Commun. Math. Phys. **31**, 243, 295 (1973); M. Schechter, Ann. Phys. (NY) **101**, 601 (1976); R. Weder, Helv. Phys. Acta **50**, 105 (1977); J. Func. Anal. **27**, 100 (1978).

[12] S. A. Fulling, Phys. Rev. D **14**, 1939 (1976); J. Rafelski, L. P. Fulcher, and A. Klein, Phys. Rep. C **38**, 227 (1978); J. Ambjørn and S. Wolfram, Ann. Phys. (NY) **147**, 33 (1983).

[13] V. A. Rizov, H. Sazdjian, and I. T. Todorov, Ann. Phys. (NY) **165**, 59 (1985).

[14] M. V. Keldysh, Usp. Mat. Nauk. **26**(4), 15 (1971) [Russian Math. Surveys **26**(4), 15 (1971)].

[15] G. V. Radzievski, Usp. Mat. Nauk. **37**(2), 81 (1982) [Russian Math. Surveys **37**(2), 81 (1982)].

[16] H. Sazdjian, Phys. Lett. B **180**, 146 (1986).

[17] A. A. Logunov and A. N. Tavkhelidze, Nuovo Cimento **29**, 380 (1963); A. A. Logunov, A. N. Tavkhelidze, I. T. Todorov, and O. A. Khrustalev, *ibid.* **30**, 134 (1963).

[18] M. H. Partovi and E. L. Lomon, Phys. Rev. D **2**, 1999 (1970).

[19] I. T. Todorov, Phys. Rev. D **3**, 2351 (1971); "Quasipotential approach to the two-body problem in quantum field theory," in *Properties of Fundamental Interactions,* Vol. 9, edited by A. Zichichi (Editrice, Bologna, 1973), Part C, p. 951.

[20] For simplicity we use the same notations for the four-dimensional internal wave function $\psi(x)$ of the first line of Eq. (3.8) and the three-dimensional internal wave function $\psi(x^T)$ of the second line. They will be distinguished from each other by the presence of their arguments: x for the former, x^T for the latter.

[21] L. Bel, "Quantum mechanics of predictive Poincaré invariant systems," in *Differential Geometry and Relativity,* edited by M. Cahen and M. Flato (Reidel, Dordrecht, 1976), p. 197.

[22] H. Sazdjian, Phys. Rev. D **33**, 3425 (1986).

Coherent states for the harmonic oscillator representations of the orthosymplectic supergroup Osp(1/2N,$\Re$)

A. B. Balantekin
Department of Physics, University of Wisconsin, Madison, Wisconsin 53706

H. A. Schmitt
Department of Physics, University of Wisconsin, Madison, Wisconsin 53706 and Department of Physics, University of Arizona, Tucson, Arizona 85721

B. R. Barrett
Physics Division, National Science Foundation, Washington, D.C. 20550 and Department of Physics, University of Arizona, Tucson, Arizona 85721[a)]

(Received 24 September 1987; accepted for publication 16 March 1988)

Coherent states for the harmonic oscillator representations of the noncompact supergroup Osp(1/2N,$\Re$) are introduced and the invariant integration measure is calculated by studying transformation properties of supercoset variables. The generalized Bogoliubov transformation for mixed systems of bosons and fermions is also obtained. An example for the simple harmonic oscillator is given.

I. INTRODUCTION

Starting with their first application in quantum optics,[1] coherent states played an important role in the study of quantum mechanical systems,[2] especially for establishing the correspondence between classical and quantum mechanics.[3] Coherent states for the group SU(2), the so-called Bloch states, were introduced by Radcliffe[4] and by Arecchi, Courtens, Gilmore, and Thomas[5] to study two-level atoms in quantum optics. After an earlier attempt,[6] the concept of coherent states was generalized to arbitrary Lie groups independently by Perelomov[7] and Gilmore and co-workers.[8] Generalized coherent states for a group are defined by the action of some representation of this group acting on a fixed vector in a given space. This fixed vector should be chosen in such a way that the coherent states obtained are "closest to classical states." Various authors elaborated on the precise meaning of this statement and gave specific criteria,[9,10] which we discuss in the forthcoming sections.

In a parallel development, supersymmetry concepts were successfully used to provide a unified description of mixed systems of bosons and fermions.[11] The generators of supersymmetry transformations form a superalgebra whose even generators belong to an ordinary Lie algebra. Although the representation theory of superalgebras has recently been studied in some detail,[12] little attention has been paid to the coherent states for supergroups. Such states can be very useful, for example, to exploit the geometry associated with the supercoherent spaces in the nuclear physics applications of supersymmetry.[13] Bars and Günaydin examined the unitary irreducible representations of the noncompact SU(M,P/$N+Q$)-type supergroups in terms of supercoherent states constructed out of particle states in a super-Fock space.[14] These supercoherent states form holomorphic functions of a graded matrix in a way analogous to the analytic representations of ordinary noncompact groups. More recently, Bowick and Gürsey have shown that[15] a system with first-class bosonic constraints obeying a Lie algebra can be BRST-quantized[16] using the nonlinear representations of an associated superalgebra, obtained by considering the fractional linear transformations of the supercoset variables.

In this paper we wish to extend previous results for coherent states to the harmonic oscillator representations of the orthosymplectic supergroups. We introduce coherent states associated with the noncompact superalgebra Osp(1/2,$\Re$) in Sec. II. In Sec. III, we show how the one dimensional simple harmonic oscillator can be represented with these coherent states. In Sec. IV, we introduce the Osp(1/2N,$\Re$) coherent states, obtain the generalized Bogoliubov transformations for mixed systems of bosons and fermions, and calculate the invariant integration measure by studying the transformation properties of the supercoset variables. Finally, in Sec. V we summarize our work and discuss future extensions and applications of our results.

II. Osp(1/2,$\Re$) COHERENT STATES

In this section we will consider the noncompact Osp(1/2,$\Re$) superalgebra. In our notation the noncompact superalgebra Osp(N/2M,$\Re$) has the compact subalgebra O(N) that acts on the fermionic subspace and the noncompact subalgebra Sp(2M,$\Re$) that acts on the bosonic subspace. The commutation–anticommutation relations of the Osp(1/2,$\Re$) superalgebra are given by

$$[K_0, K_\pm] = \pm K_\pm , \tag{2.1a}$$
$$[K_+, K_-] = -2K_0 , \tag{2.1b}$$
$$[K_0, F_\pm] = \pm \tfrac{1}{2} F_\pm , \tag{2.1c}$$
$$[K_+, F_+] = 0 = [K_-, F_-] , \tag{2.1d}$$
$$[K_\pm, F_\mp] = \mp F_\pm , \tag{2.1e}$$
$$\{F_\pm, F_\pm\} = K_\pm , \tag{2.1f}$$
$$\{F_+, F_-\} = K_0 , \tag{2.1g}$$

where K_0, $K_\pm$, and $F_\pm$ are the five generators of the superalgebra. Osp(1/2,$\Re$) has the subalgebra Sp(2,$\Re$) ~ SU(1,1)

[a)] Permanent address after 15 August 1987.

 0022-2488/88/071634-06$02.50

spanned by $K_{\pm}$ and K_0 [cf. Eqs. (2.1a) and (2.1b)]. The representations of Osp(1/2,$\Re$) are studied in detail in Ref. 17 and further in Refs. 14, 18, and 19. Useful results from the representation theory of Osp(1/2,$\Re$) and Sp(2,$\Re$) are briefly summarized in Appendix A. In the following discussion we restrict ourselves to the positive, discrete series representations for both Osp(1/2,$\Re$) and Sp(2,$\Re$).

The coherent state for the discrete series representation of Sp(2,$\Re$), characterized by the label k, is given by[7]

$$|\alpha\rangle = N\exp(\alpha K_+)|k,m=k\rangle\,, \tag{2.2}$$

where α is a complex number. If the coherent state is normalized to unity the normalization constant in Eq. (2.2) is given by $N=(1-|\alpha|^2)^k$. Since the set of coherent states given by Eq. (2.2) is overcomplete, the decomposition of identity is given by

$$\int d\mu(\alpha^*,\alpha)|\alpha\rangle\langle\alpha| = \sum_m |k,m\rangle\langle k,m| = \mathbf{1}\,, \tag{2.3a}$$

where the measure is given by

$$d\mu(\alpha^*,\alpha) = \frac{d^2\alpha}{\pi}\frac{(2k-1)}{(1-|\alpha|^2)^2}\,, \quad \text{for } k\neq\frac{1}{2}\,, \tag{2.3b}$$

and

$$d\mu(\alpha^*,\alpha) = \frac{d^2\alpha}{\pi}\frac{1}{(1-|\alpha|^2)}\,, \quad \text{for } k=\frac{1}{2}\,. \tag{2.3c}$$

These states satisfy the criterion[9,10] for being "closest to classical," i.e., they minimize the dispersion

$$\Delta C_2 = \langle C_2\rangle - \langle K_0\rangle^2 + \langle K_+\rangle\langle K_-\rangle\,, \tag{2.4a}$$

where the Casimir operator of Sp(2,$\Re$) is given by[20]

$$C_2 = K_0^2 - \tfrac{1}{2}(K_+K_- + K_-K_+)\,. \tag{2.4b}$$

Vector coherent-state representations for noncompact symplectic groups were studied in Ref. 21. In this paper we restrict ourselves only to the so-called harmonic oscillator representations of the orthosymplectic supergroup Osp(1/2N,$\Re$). A study of the coherent states based on generic lowest-weight representations for Osp(1/2N,$\Re$) is beyond the scope of the present article.

In analogy with Eq. (2.2) we define the coherent states for Osp(1/2,$\Re$) as

$$|\alpha,\theta\rangle = N\exp(\alpha K_+ + \theta F_+)|\tau,k=\tau,m=\tau\rangle\,, \tag{2.5}$$

where α is a complex number and θ is a Grassmann number. The appropriate representations of the Osp(1/2,$\Re$) are given in Appendix A and we refer the reader to Ref. 22 for the description of the properties of Grassmann numbers. The expansion of the state $|\alpha,\theta\rangle$ into the Osp(1/2,$\Re$) basis states is given in Appendix B, Eq. (B1). If we normalize the coherent states $|\alpha,\theta\rangle$ to unity, the normalization constant in Eq. (2.5) can easily be evaluated using Eq. (B1) and the orthonormality of the Osp(1/2,$\Re$) states. We get

$$\langle\alpha,\theta|\alpha',\theta'\rangle = N^2(1-\alpha^*\alpha')^{-2\tau}[1+\bar\theta\theta'\tau/(1-\alpha^*\alpha')] = 1\,. \tag{2.6a}$$

where $\bar\theta$ is the complex conjugate of θ. We find it convenient to write this normalization in the form

$$N = [\mathrm{Sdet}\,\mathscr{M}(\alpha^*,\alpha;\bar\theta,\theta)]^{-\tau}\,, \tag{2.6b}$$

where the matrix $\mathscr{M}$ is given by

$$\mathscr{M}(\alpha^*,\alpha';\bar\theta,\theta') = \begin{pmatrix} 1 & \theta'/\sqrt{2} \\ \bar\theta/\sqrt{2} & (1-\alpha^*\alpha') \end{pmatrix}. \tag{2.6c}$$

In writing Eqs. (2.6a) and (2.6b) we used the fact that, since $\theta\theta=0$, a power series expansion of a function of θ is finite. The superdeterminant of the graded matrix

$$\mathscr{M} = \begin{pmatrix} \mathscr{A} & \mathscr{B} \\ \mathscr{C} & \mathscr{D} \end{pmatrix} \tag{2.7a}$$

is defined as

$$\mathrm{Sdet}\,\mathscr{M} = \det(\mathscr{A}-\mathscr{B}\mathscr{D}^{-1}\mathscr{C})/\det\mathscr{D}\,. \tag{2.7b}$$

In Eq. (2.7a), the elements of the matrices $\mathscr{A}$ and $\mathscr{D}$ are commuting variables and those of the matrices $\mathscr{B}$ and $\mathscr{C}$ are anticommuting variables.

One can again calculate the decomposition of identity by introducing a suitable measure. In Sec. IV we calculate the measure for Osp(1/2N,$\Re$) coherent states. Here we only quote the result:

$$\int d\mu(\alpha^*,\alpha;\bar\theta,\theta)|\alpha,\theta\rangle\langle\alpha,\theta| = \sum_m |\tau,\tau,m\rangle\langle\tau,\tau,m| + \sum_{m'} |\tau,\tau+\tfrac{1}{2},m'\rangle\langle\tau,\tau+\tfrac{1}{2},m'| = \mathbf{1}\,, \tag{2.8a}$$

where the measure is given by

$$d\mu(\alpha^*,\alpha;\bar\theta,\theta) = (2/\pi)d\bar\theta\,d\theta\,d^2\alpha\,\mathrm{Sdet}\,\mathscr{M}(\alpha^*,\alpha;\bar\theta,\theta)\,. \tag{2.8b}$$

In the above integrations over the Grassmann variables, we fix our normalizations as

$$\int d\bar\theta\,d\theta(1,\theta,\bar\theta) = 0, \tag{2.9a}$$

$$\int d\bar\theta\,d\theta\,\theta\bar\theta = 1\,. \tag{2.9b}$$

In Ref. 9, it was shown that coherent states, invariably constructed from the highest-weight state of a Lie algebra, are closest to the classical states. Namely, they minimize

$$\Delta C_2 = \langle C_2\rangle - g^{ij}\langle X_i\rangle\langle X_j\rangle\,, \tag{2.10}$$

where the X_i are the generators of the algebra, g^{ij} is the Cartan–Killing metric tensor, and $C_2 = g^{ij}X_iX_j$ is the quadratic Casimir operator.[20] Using the expectation values of the Osp(1/2,$\Re$) generators between our coherent states, given in Appendix B, one can easily show that the above result generalizes to the Osp(1/2,$\Re$) superalgebra. The states given in Eq. (2.5) minimize the dispersion

$$\Delta C_2 = \langle C_2\rangle - \langle K_0\rangle^2 + \tfrac{1}{2}(\langle K_+\rangle\langle K_-\rangle + \langle K_-\rangle\langle K_+\rangle) - \tfrac{1}{2}(\langle F_+\rangle\langle F_-\rangle - \langle F_-\rangle\langle F_+\rangle)\,, \tag{2.11}$$

where the quadratic Casimir operator is given by Eq. (A1).

III. EXAMPLE: APPLICATION TO THE HARMONIC OSCILLATOR

Let b and $b^\dagger$ be the annihilation and creation operators for a one-dimensional harmonic oscillator. The three operators

$$K_+ = \tfrac{1}{2} b^\dagger b^\dagger , \tag{3.1a}$$

$$K_- = \tfrac{1}{2} bb , \tag{3.1b}$$

$$K_0 = \tfrac{1}{2}(b^\dagger b + \tfrac{1}{2}) \tag{3.1c}$$

form an Sp(2,$\Re$) algebra, the quadratic Casimir operator of which is $-\frac{3}{16}$ (cf. Ref. 20). Consequently for the realization given in Eq. (3.1), k is either $\frac{1}{4}$ or $\frac{3}{4}$. Since the operators $K_\pm$ change the harmonic oscillator quantum number, n, by 2, the even-n and odd-n states correspond to separate representations of Sp(2,$\Re$). We choose to include the even-n states (including the ground state) in the $k = \frac{1}{4}$ representation, and the odd-n states in the $k = \frac{3}{4}$ representation. By formally introducing a grading between even-n and odd-n states, one can embed both representations into a single irreducible representation of Osp(1/2,$\Re$) with $\tau = \frac{1}{4}$. The odd generators of this superalgebra are

$$F_+ = \tfrac{1}{2} b^\dagger , \tag{3.2a}$$

$$F_- = \tfrac{1}{2} b . \tag{3.2b}$$

The superalgebra Osp(1/2,$\Re$) generated by $K_\pm$, K_0, and $F_\pm$ given in Eqs. (3.1) and (3.2) appears to be the largest symmetry algebra of the one-dimensional harmonic oscillator. To be precise, the right-hand side of the above equations should be multiplied by a term of the form $(1/\sqrt{2})(\alpha^\dagger + \alpha)$, where α and $\alpha^\dagger$ are annihilation and creation operators for a single fermion. However, the commutation–anticommutation relations of Eqs. (2.1) would be satisfied if this term commutes with the harmonic oscillator operators and if its square is unity. Consequently, the identity operator can be substituted for this term. Aspects of symmetries and supersymmetries of the quantum harmonic oscillator are further discussed in Refs. 23 and 24. In Ref. 25, the eigenstates of the supersymmetric annihilation operator of the supersymmetric harmonic oscillator are named as supercoherent states. Those supercoherent states are the supersymmetric generalization of Glauber coherent states. The group structure and properties of our coherent states are rather different.

We can now write the coherent state for the above realization of Osp(1/2,$\Re$) in terms of harmonic oscillator states. One can easily make the identification

$$|\tau = \tfrac{1}{4}, k = \tfrac{1}{4}, m = \tfrac{1}{4} + p\rangle = |2p\rangle , \tag{3.3a}$$

$$|\tau = \tfrac{1}{4}, k = \tfrac{3}{4}, m = \tfrac{3}{4} + p\rangle = |2p-1\rangle , \tag{3.3b}$$

where p is an integer, between the Osp(1/2,$\Re$) (left-hand side) and harmonic oscillator (right-hand side) states. The coherent state is

$$|\theta,\alpha\rangle = N\left[\sum_{p=0}^{\infty} [(2p)!]^{1/2} \frac{\alpha^p}{2^p p!} |2p\rangle + \frac{\theta}{2} \sum_{p=1}^{\infty} \sqrt{p}\,[(2p)!]^{1/2} \frac{\alpha^{p-1}}{2^p p!} |2p-1\rangle\right], \tag{3.4a}$$

with

$$N = (1-|\alpha|^2)^{1/4}\left(1 + \tfrac{1}{8}\theta\bar{\theta}/(1-|\alpha|^2)\right) . \tag{3.4b}$$

The coherent-state wave functions, i.e., the overlap between the coherent state and the position eigenstates, are

$$\langle x|\alpha,\theta\rangle = N(mw/\pi\hbar)^{1/4} e^{-\xi^2/2}(1+\alpha)^{-1/2} \times \exp\left(\frac{\xi^2\alpha}{1+\alpha}\right)\left[1 + \frac{1}{2}\frac{\theta\xi}{(1+\alpha)}\right], \tag{3.5}$$

where m and w are the mass and the frequency of the oscillator and $\xi = \sqrt{mw/\hbar}\, x$.

IV. CALCULATION OF THE INTEGRATION MEASURE FOR THE HARMONIC OSCILLATOR REPRESENTATIONS OF Osp(1/2N,$\Re$)

In this section we introduce the Osp(1/2N,$\Re$) coherent states and calculate the invariant integration measure. Osp(1/2N,$\Re$) superalgebras are of particular interest, since they are the only superalgebras for which the direct products of all finite-dimensional representations are completely reducible.[26] By introducing the boson creation (annihilation) operators $b_i^\dagger (b_j)$, $i,j = 1,\ldots,N$, we can write the generators of the noncompact Sp(2N,$\Re$) [the even generators of Osp(1/2N,$\Re$)] as bilinear products[20]:

$$b_i^\dagger b_j^\dagger ,\ b_i b_j ,\ b_i^\dagger b_j + b_j^\dagger b_i . \tag{4.1}$$

By further introducing the creation and annihilation operators for one single fermion, $a^\dagger$ and a, we can write the odd generators of Osp(1/2N,$\Re$) as[14]

$$(1/\sqrt{2})(a + a^\dagger) b_i ,\ (1/\sqrt{2})(a + a^\dagger) b_i^\dagger . \tag{4.2}$$

Note that the fermion operator appears only in the combination $(1/\sqrt{2})(a + a^\dagger)$, the square of which is unity.

We now consider the most general linear transformation of the boson–fermion creation and annihilation operators:

$$b_i \to b_i' = u_{ij} b_j + v_{ik} b_k^\dagger + \theta_i (1/\sqrt{2})(a + a^\dagger) , \tag{4.3a}$$

$$b_i^\dagger \to b_i'^\dagger = v_{ij}^* b_j + u_{ij}^* b_j^\dagger + (1/\sqrt{2})(a + a^\dagger)\bar{\theta}_i , \tag{4.3b}$$

$$(1/\sqrt{2})(a + a^\dagger) \to (1/\sqrt{2})(a' + a'^\dagger) = \lambda (1/\sqrt{2})(a + a^\dagger) + \chi_i b_i + \bar{\chi}_i b_i^\dagger , \tag{4.3c}$$

where u_{ij}, v_{ij}, and λ are complex (commuting) numbers and $\theta_i, \bar{\theta}_i, \chi_i$, and $\bar{\chi}_i$ are Grassmann (anticommuting) numbers. The transformation given in Eq. (4.3) is manifestly supersymmetric, since it mixes bosons and fermions. We want this transformation to be canonical, i.e., the transformed operators should also satisfy the boson commutation and the fermion anticommutation relations. This condition imposes certain relations between the variables $u, v, \theta, \bar{\theta}, \chi$, and $\bar{\chi}$. We find it convenient to express these relations in a matrix form. Introducing the $(2N+1)\times(2N+1)$ supermatrix

$$\mathcal{N} = \begin{pmatrix} u & v & \theta \\ v^* & u^* & -\bar{\theta} \\ \chi & \bar{\chi} & \lambda \end{pmatrix}, \tag{4.4}$$

we find that the transformation (4.3) is canonical if

$$\mathcal{N} H \mathcal{N}^{\mathrm{ST}} = H , \tag{4.5}$$

where $\mathcal{N}^{\mathrm{ST}}$ is the supertranspose of the matrix $\mathcal{N}$, and

$$H = \begin{pmatrix} G & 0 \\ 0 & 1 \end{pmatrix}, \tag{4.6a}$$

with the $2N \times 2N$ matrix G being

$$G=\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{4.6b}$$

In the above equations I is an $N\times N$ unit matrix. The supertranspose of a matrix

$$\mathcal{N}_1=\begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix}, \tag{4.7a}$$

where the matrices $\mathcal{A}$ and $\mathcal{D}$ have complex elements and the matrices $\mathcal{B}$ and $\mathcal{C}$ have Grassmann elements, is defined as

$$\mathcal{N}_1^{\mathrm{ST}}=\begin{pmatrix} \mathcal{A}^T & \mathcal{C}^T \\ -\mathcal{B}^T & \mathcal{D}^T \end{pmatrix}. \tag{4.7b}$$

Equation (4.5) indicates that the matrix $\mathcal{N}$ is an element of the Osp$(1/2N,\Re)$ *supergroup*.[27] To summarize, a transformation of the form

$$\begin{pmatrix} b' \\ b'^{\dagger} \\ (1/\sqrt{2})(a'+a'^{\dagger}) \end{pmatrix}=\mathcal{N}\begin{pmatrix} b \\ b^{\dagger} \\ (1/\sqrt{2})(a+a^{\dagger}) \end{pmatrix} \tag{4.8}$$

is canonical if $\mathcal{N}$ is an element of the Osp$(1/2N,\Re)$ supergroup. The transformation given in Eq. (4.8) can be considered a generalization of the Bogoliubov transformation[28] to the mixed systems of bosons and fermions.

In order to define the coherent states, it is convenient to introduce the operator

$$T=\exp[\tfrac{1}{2}Z_{ij}b_i^{\dagger}b_j^{\dagger}]\exp[\tfrac{1}{2}\psi_i(a+a^{\dagger})b_i^{\dagger}], \tag{4.9}$$

where Z is an $N\times N$ complex, symmetric matrix and ψ is an $N\times 1$ Grassmann column vector. One can easily show that

$$Tb_iT^{-1}=b_i-Z_{ij}b_j^{\dagger}-\tfrac{1}{2}\psi_i(a+a^{\dagger}), \tag{4.10}$$

from which it follows that the transformation (4.3a) can be written in the form

$$b_i'=u_{ij}Tb_jT^{-1} \tag{4.11a}$$

if

$$Z_{jk}=-u_{jl}^{-1}v_{lk},\quad \psi_j=-\sqrt{2}u_{jl}^{-1}\theta_l. \tag{4.11b}$$

Consequently if we define the Osp$(1/2N,\Re)$ coherent state as

$$|Z,\psi\rangle=\exp[\tfrac{1}{2}Z_{ij}b_i^{\dagger}b_j^{\dagger}]\exp[\tfrac{1}{2}\psi_i(a+a^{\dagger})b_i^{\dagger}]|0\rangle, \tag{4.12}$$

where $|0\rangle$ is the common vacuum for the boson and fermion operators, one can identify quasiparticle operators annihilating the coherent state:

$$[b_j-Z_{jk}b_k^{\dagger}-\tfrac{1}{2}\psi_j(a+a^{\dagger})]|Z,\psi\rangle=0. \tag{4.13}$$

We postulate that the coherent state defined in Eq. (4.12) would minimize the dispersion of the Casimir operator, Eq. (2.10). However, since we have not constructed the basis states of the representations of Osp$(1/2N,\Re)$ for $N>2$, we cannot explicitly prove this statement and verify the analog of Eq. (2.11).

We now apply the transformation (4.3) to Eq. (4.13). We find that, for the coherent state to be left invariant under this transformation, one needs

$$Z\to Z'=(u^{\dagger}+Zv^{\dagger}-(1/\sqrt{2})\psi\bar{\theta}^T)^{-1}\times(v^T+Zu^T+(1/\sqrt{2})\psi\bar{\theta}^T), \tag{4.14a}$$

$$\psi\to\psi'=(u^{\dagger}+Zv^{\dagger}-(1/\sqrt{2})\psi\bar{\theta}^T)^{-1}\times(\lambda\psi-\sqrt{2}\,\bar{\chi}-\sqrt{2}Z\chi^T). \tag{4.14b}$$

Alternately, if we define the $N\times(N+1)$ graded rectangular matrix

$$\mathcal{Z}\equiv(Z,\psi/\sqrt{2}), \tag{4.15}$$

and the matrices

$$\alpha\equiv u^{\dagger}, \tag{4.16a}$$

$$\beta\equiv(v^T,-\bar{\chi}), \tag{4.16b}$$

$$\gamma\equiv\begin{pmatrix} v^{\dagger} \\ -\bar{\theta}^T \end{pmatrix}, \tag{4.16c}$$

$$\delta\equiv\begin{pmatrix} u^T & -\chi^T \\ \theta^T & \lambda \end{pmatrix}, \tag{4.16d}$$

we find that Eq. (4.14) can be reexpressed as a linear fractional transformation of the graded vector $\mathcal{Z}$:

$$\mathcal{Z}\to\mathcal{Z}'=(Z',\psi'/\sqrt{2})=(\alpha+\mathcal{Z}\gamma)^{-1}(\beta+\mathcal{Z}\delta). \tag{4.17}$$

Furthermore, using Eqs. (4.4) and (4.5), one finds that the matrix

$$\mathcal{K}=\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}=\begin{pmatrix} u^{\dagger} & v^T & -\bar{\chi}^T \\ v^{\dagger} & u^T & -\chi^T \\ -\bar{\theta}^T & \theta^T & \lambda \end{pmatrix} \tag{4.18}$$

is also an element of the supergroup Osp$(1/2N,\Re)$.

The next step is to express the matrix $\mathcal{K}$ in terms of Z, ψ, and u. Since any complex matrix can be expressed as a product of a Hermitian and a unitary matrix[29] we write

$$u=\hat{U}B, \tag{4.19}$$

where $\hat{U}$ is an $N\times N$ unitary matrix and B is an $N\times N$ Hermitian matrix. Using Eqs. (4.4) and (4.5) we find that the matrix $\mathcal{K}$ can be written as a product of two matrices

$$\mathcal{K}=\mathcal{K}_1\mathcal{K}_2, \tag{4.20a}$$

where $\mathcal{K}_2$ is an element of the U(N) subgroup of Sp$(2N,\Re)$,

$$\mathcal{K}_2=\begin{pmatrix} \hat{U}^{\dagger} & 0 & 0 \\ 0 & \hat{U}^T & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4.20b}$$

and $\mathcal{K}_1$ is an element of the supercoset Osp$(1/2N,\Re)$/U(N),

$$\mathcal{K}_1=\begin{pmatrix} 1 & -Z^T & -\bar{\chi}^T \\ -Z^{\dagger} & 1 & -\chi^T \\ \bar{\psi}^T/\sqrt{2} & -\psi^T/\sqrt{2} & \lambda \end{pmatrix}\begin{pmatrix} B^T & 0 & 0 \\ 0 & B^T & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.20c}$$

In the above equations λ can be obtained from

$$\lambda^2-(1-Z^*Z)_{ij}^{-1}[\bar{\psi}_j-Z_{jk}^*\psi_k]\lambda^2\psi_i=1, \tag{4.21}$$

and $\bar{\chi}$ from

$$\bar{\chi}=(1-ZZ^*)^{-1}[\psi/\sqrt{2}-Z\bar{\psi}/\sqrt{2}]\lambda. \tag{4.22}$$

Finally B satisfies the equation

$$B^TB=(1-ZZ^{\dagger}+\psi\bar{\psi}^T/2)^{-1}. \tag{4.23}$$

We now wish to evaluate an invariant measure $\mu(\mathcal{Z})$ such that

$$\mu(\mathscr{Z}) = \mu(\mathscr{Z}')|J(\mathscr{Z}',\mathscr{Z})| = \mu(0)|J(0,\mathscr{Z})|\,, \tag{4.24}$$

where $J(\mathscr{Z}',\mathscr{Z})$ is the Jacobian of the transformation (4.17) that transforms $\mathscr{Z}$ into $\mathscr{Z}'$. Differentiating Eq. (4.17) we get

$$(\alpha - \beta\delta^{-1}\gamma)\, d\mathscr{Z}' = d\mathscr{Z}\,\delta\,. \tag{4.25}$$

In the above equation the matrix $(\alpha - \beta\delta^{-1}\gamma)$ at the left-hand side transforms bosonic and fermionic coordinates separately, without mixing them, whereas the matrix δ at the right-hand side mixes them as well [cf. Eqs. (4.16)]. Hence Eq. (4.25) yields the Jacobian

$$[\det(\alpha - \beta\delta^{-1}\gamma)\,]^{2(N-1)}|J(0,\mathscr{Z})| = [\mathrm{Sdet}\,\delta]^{2N}\,. \tag{4.26}$$

Hence the invariant measure in the decomposition of unity in terms of the Osp(1/2N) coherent states $|Z,\psi>$ of Eq. (4.12),

$$\int d\mu(Z,\psi)\,|Z,\psi\rangle\langle Z,\psi| = \mathbf{1}\,, \tag{4.27a}$$

is

$$d\mu(Z,\psi) = \mathrm{const}\times d\bar{\theta}\, d\theta\, d^{2N}Z \times [\mathrm{Sdet}\,\delta]^{2N}/[\det(\alpha - \beta\delta^{-1}\gamma)\,]^{2(N-1)}\,. \tag{4.27b}$$

For Osp(1/2,$\Re$) ($N = 1$) this gives Eq. (2.8b).

V. CONCLUSIONS

In this paper we introduced the coherent states for the harmonic oscillator representations of the noncompact supergroup Osp(1/2N,$\Re$), obtained the generalized Bogoliubov transformations for mixed systems of bosons and fermions, and calculated the invariant integration measure by studying transformation properties of the supercoset variables. We also exemplified the use of these coherent states for the simple harmonic oscillator. It would be straightforward to extend our results to the harmonic oscillator representations of the compact version of Osp(1/2N,$\Re$). In a separate article,[30] we investigate the coherent states for Osp(2/2N,$\Re$), in which case one needs to treat typical and atypical representations[19] individually. It would be interesting to investigate Osp(1/2N,$\Re$) coherent states based on the most general lowest-weight representations (e.g., those with a vectorial structure). Such an undertaking would result in orthosymplectic supergroup coherent states generalizing symplectic group coherent states discussed in Ref. 21.

In many-body systems, the pairing-type interactions between particles alter the structure of the vacuum state. For example, in the BCS theory of superconductivity[31] correlated Cooper pairs contribute to the physical vacuum. Bogoliubov transformations and the resulting quasiparticle operators provide a natural framework to study such correlations. The generalized Bogoliubov transformation given in Eq. (4.8) could be used to analyze the possible relationships between the pairing and the boson–fermion exchange forces in mixed systems of bosons and fermions. Such a possibility is presently being investigated.

Orthosymplectic supergroups are the invariance groups of second moments arising in the study of compound nucleus scattering[32] and of the action in the path-integral derivation of symmetry relations for disordered electron systems.[33] Furthermore, one can associate supersymmetric quantum mechanical systems with classical, nonlinear Langevin equations and show that the averages over a Gaussian noise distribution are the same as the quantum mechanical expectation values calculated with a supersymmetric effective action.[34] One can show that the supersymmetry transformations leaving this effective action invariant are also orthosymplectic transformations. In future publications, we plan to study the possible uses of our coherent states in the Grassmann integration approach to the compound nucleus scattering and other problems in stochastic quantum physics.

ACKNOWLEDGMENTS

We thank A. Bincer, C. Goebel, and J. Palmer for useful discussions and A. Bincer for his comments on the manuscript. One of us (H.A.S.) would also like to thank the members of the Theoretical Nuclear Physics group at the University of Wisconsin for their hospitality during his stay.

The research of A.B.B. was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, in part by a Presidential Young Investigator Award administered by the National Science Foundation, and in part by the National Science Foundation Grant No. PHY-8610851. The research of H.A.S. was supported in part by the National Science Foundation Grant No. PHY-8603883.

APPENDIX A: REPRESENTATIONS OF Osp(1/2,$\Re$)

The quadratic Casimir operator of Osp(1/2,$\Re$) is given by

$$C_2(\mathrm{Osp}(1/2,\Re)) = A^2 + \hat{A}/2\,, \tag{A1}$$

where

$$\hat{A} = [F_+,F_-]\,, \tag{A2}$$

and the quadratic Casimir operator of the subalgebra $\mathrm{Sp}(2,\Re) \sim \mathrm{SU}(1,1)$ is given by

$$C_2(\mathrm{Sp}(2,\Re)) = \hat{A}^2 + \hat{A}\,. \tag{A3}$$

The eigenvalues of the Casimir operator of Osp(1/2,$\Re$), Eq. (A1), are denoted by $\tau(\tau - \frac{1}{2})$, where the label τ characterizes irreducible representations. The eigenvalues of the Casimir operator of Sp(2,$\Re$), Eq. (A3), are denoted by $k(k-1)$. There are two irreducible representations of Sp(2,$\Re$), corresponding to $k = \tau$ and $k = \tau + \frac{1}{2}$, embedded in a single representation of Osp(1/2,$\Re$). Here we consider only the discrete series representations and label the states with τ, k, and m, the eigenvalue of K_0:

$$C_2(\mathrm{Osp}(1/2,\Re))|\tau,k,m\rangle = \tau(\tau - \tfrac{1}{2})|\tau,k,m\rangle\,, \tag{A4a}$$

$$C_2(\mathrm{Sp}(2,\Re))|\tau,k,m\rangle = k(k-1)|\tau,k,m\rangle\,, \tag{A4b}$$

$$K_0|\tau,k,m\rangle = m|\tau,k,m\rangle\,. \tag{A4c}$$

In Eq. (A4c), m takes values $k, k+1, k+2, k+3, \ldots$. The result

$$K_\pm|\tau,k,m\rangle = [(m \pm k)(m \mp k \pm 1)\,]^{1/2}|\tau,k,m \pm 1\rangle \tag{A5}$$

follows from the standard representation theory of $Sp(2,\Re)$. Finally imposing $F_+^\dagger = F_-$ and choosing all the phases to be real we obtain

$$F_\pm |\tau,k=\tau,m\rangle = (1/\sqrt{2})[m \pm \tau]^{1/2}|\tau,k=\tau+\tfrac{1}{2},m\pm\tfrac{1}{2}\rangle , \tag{A6a}$$

and

$$F_\pm |\tau,k=\tau+\tfrac{1}{2},m\rangle = (1/\sqrt{2})[m \mp \tau \pm \tfrac{1}{2}]^{1/2}|\tau,k=\tau,m\pm\tfrac{1}{2}\rangle . \tag{A6b}$$

Equations (A4)–(A6) completely specify the representations of $Osp(1/2,\Re)$.

APPENDIX B: COHERENT STATE EXPECTATION VALUES OF Osp(1/2,ℜ) GENERATORS

Expectation values of the $Osp(1/2,\Re)$ generators between the coherent states can easily be calculated using the expansion of the coherent state $|\alpha,\theta\rangle$ into the $Osp(1/2,\Re)$ basis states:

$$\begin{aligned}|\alpha,\theta\rangle = N \sum_{n=0}^{\infty} &\left[\frac{\Gamma(2\tau+n)}{\Gamma(n+1)\Gamma(2\tau)}\right]^{1/2} \\ &\times \alpha^n |\tau,k=\tau,m=\tau+n\rangle \\ + N \sum_{n=1}^{\infty} &\sqrt{\tau}\left[\frac{\Gamma(2\tau+n)}{\Gamma(2\tau+1)\Gamma(n)}\right]^{1/2} \\ &\times \alpha^{n-1}\theta\,|\tau,k=\tau+\tfrac{1}{2},m=\tau+N+\tfrac{1}{2}\rangle ,\end{aligned} \tag{B1}$$

where the normalization factor N is given by Eq. (2.6b). We get

$$\frac{\langle\alpha,\theta|K_+|\alpha',\theta'\rangle}{\langle\alpha,\theta|\alpha',\theta'\rangle} = 2\tau\alpha^*\,\mathrm{Sdet}\,\mathscr{M}(\alpha^*,\alpha';\bar\theta,\theta') , \tag{B2}$$

$$\frac{\langle\alpha,\theta|K_-|\alpha',\theta'\rangle}{\langle\alpha,\theta|\alpha',\theta'\rangle} = 2\tau\alpha'\,\mathrm{Sdet}\,\mathscr{M}(\alpha^*,\alpha';\bar\theta,\theta') , \tag{B3}$$

$$\frac{\langle\alpha,\theta|K_0|\alpha',\theta'\rangle}{\langle\alpha,\theta|\alpha',\theta'\rangle} = \tau(1+\alpha^*\alpha')\,\mathrm{Sdet}\,\mathscr{M}(\alpha^*,\alpha';\bar\theta,\theta') , \tag{B4}$$

$$\frac{\langle\alpha,\theta|F_+|\alpha',\theta'\rangle}{\langle\alpha,\theta|\alpha',\theta'\rangle} = -\tau(\bar\theta+\alpha^*\theta')\,\mathrm{Sdet}\,\mathscr{M}(\alpha^*,\alpha';\bar\theta,\theta') , \tag{B5}$$

$$\frac{\langle\alpha,\theta|F_-|\alpha',\theta'\rangle}{\langle\alpha,\theta|\alpha',\theta'\rangle} = -\tau(\theta'+\alpha'\bar\theta)\,\mathrm{Sdet}\,\mathscr{M}(\alpha^*,\alpha';\bar\theta,\theta') , \tag{B6}$$

where $\mathscr{M}(\alpha^*,\alpha';\bar\theta,\theta')$ is given by Eq. (2.6c).

[1] R. Glauber, Phys. Rev. **130**, 2529 (1963); **131**, 2766 (1963).
[2] For a recent compendium of results, see *Coherent States*, edited by J. R. Klauder and B.-S. Skagerstam (World Scientific, Singapore, 1985).
[3] P. Carruthers and M. N. Nieto, Rev. Mod. Phys. **40**, 411 (1968); L. G. Yaffe, *ibid.* **54**, 407 (1982).
[4] J. M. Radcliffe, J. Phys. A **4**, 313 (1971).
[5] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[6] A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[7] A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972); **44**, 197 (1975); Sov. Phys. Usp. **20**, 703 (1977).
[8] R. Gilmore, Ann. Phys. (NY) **74**, 391 (1972); Rev. Mex. Fis. **23**, 143 (1974); F. T. Arecchi, R. Gilmore, and D. M. Kim, Lett. Nuovo Cimento **6**, 219 (1973); R. Gilmore, C. M. Bowden, and L. M. Narducci, Phys. Rev. A **12**, 1019 (1975). Also see Ref. 4.
[9] A. M. Perelomov, Sov. J. Nucl. Phys. **29**, 867 (1979).
[10] R. Delbourgo and J. R. Fox, J. Phys. A **10**, L233 (1977).
[11] For a recent review, see *Supersymmetry in Physics*, edited by V. A. Kostelecky and D. K. Campbell (North-Holland, Amsterdam, 1985).
[12] For a recent review, see I. Bars, Lect. Appl. Math. **21**, 17 (1985).
[13] A. B. Balantekin, I. Bars, and F. Iachello, Phys. Rev. Lett. **47**, 19 (1981).
[14] I. Bars and M. Günaydin, Commun. Math. Phys. **91**, 31 (1983).
[15] M. J. Bowick and F. Gürsey, Phys. Lett. B **175**, 182 (1986), Nucl. Phys. B **283**, 331 (1987).
[16] C. Becchi, A. Rouet, and R. Stora, Phys. Lett. B **52**, 344 (1974); I. V. Tyutin, Lebedev preprint FIAN No. 39, 1975.
[17] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977); A. Pais and V. Rittenberg, *ibid.* **17**, 868 (1976).
[18] E. D'Hoker and L. Vinet, Phys. Lett. B **137**, 72 (1984); Commun. Math. Phys. **97**, 391 (1985).
[19] A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1149 (1981); R. J. Farmer and P. D. Jarvis, J. Phys. A **16**, 473 (1983); A **17**, 2365 (1984); A. M. Bincer, J. Math. Phys. **24**, 2546 (1983); Q.-Z. Han, H.-Z. Sun, M. Zhang, and D. H. Feng, *ibid.* **26**, 1822 (1985); M. Gourdin, *ibid.* **27**, 61, 1980 (1986), Ann. Inst. H. Poincaré **46**, 281 (1987); V. G. Kac, Commun. Math. Phys. **53**, 31 (1977).
[20] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974); B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[21] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); J. Deenen and C. Quesne, *ibid.* **25**, 2354 (1984); D. J. Rowe, G. Rosensteel, and R. Gilmore, *ibid.* **26**, 2787 (1985).
[22] F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[23] A. B. Balantekin, Ann. Phys. (NY) **164**, 277 (1985); R. J. Hughes, V. A. Kostelecky, and M. M. Nieto, Phys. Lett. B **171**, 226 (1986).
[24] J. Beckers, D. Dehin, and V. Hussin, J. Phys. **20**, 1137 (1987); P. Celka and V. Hussin, Mod. Phys. Lett. A **2**, 391 (1987).
[25] C. Aragone and F. Zypman, J. Phys. **19**, 2267 (1986).
[26] D. Z. Djokovic and G. Hochschild, Ill. J. Math. **20**, 134 (1976).
[27] F. A. Berezin and G. I. Kac, Math. USSR Sb. **11**, 311 (1970); V. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 709 (1978); F. Gürsey and L. Marchildon, *ibid.* **19**, 942 (1978); Phys. Rev. **17**, 2038 (1978).
[28] N. N. Bogoliubov, J. Phys. USSR **11**, 23 (1947).
[29] L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in Classical Dormains* (American Mathematical Society, Providence, RI, 1963).
[30] A. B. Balantekin and H. A. Schmitt, University of Wisconsin at Madison preprint Mad/NT/88-06.
[31] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **106**, 162 (1957).
[32] J. J. M. Verbaarschot, H. A. Weidenmüller, and M. R. Zirnbauer, Phys. Rep. **129**, 367 (1985).
[33] F. Wegner, Z. Phys. B **49**, 297 (1983).
[34] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979), Nucl. Phys. B **206**, 321 (1982); B. Freedman and F. Cooper, Physica D **15**, 138 (1985).

Cartan and relativistic spin fluids in a rotating cylinder

J. P. Krisch
Department of Physics, The University of Michigan, Ann Arbor, Michigan 48109

L. L. Smalley
Department of Physics, The University of Alabama in Huntsville, Huntsville, Alabama 35899 and Space Sciences Laboratory, Marshall Space Flight Center, Alabama 35812

(Received 14 December 1987; accepted for publication 24 February 1988)

Spin-fluid sources and their metric are obtained for an infinite rotating cylinder. The calculation is performed in general relativity and the Cartan theory. The spin fluids are significantly different in the two theories.

I. INTRODUCTION

Rotating fluids are often used in astrophysical and cosmological model calculations. Rotation can be included in relativistic fluid calculations in several ways. The fluid can be contained within a space-time $g_{\mu\nu}$ ($\mu,\nu = 0$–3) that has an off-diagonal component g_{0i} ($i = 1$–3). This can introduce fluid rotation, which is described by the vorticity tensor[1]

$$\omega_{\mu\nu} = U_{[\mu;\nu]} + \dot{U}_{[\mu} U_{\nu]}, \tag{1}$$

where U_μ is the fluid velocity and $\dot{U}_\mu = U_{\mu;\nu} U^\nu$ the acceleration.

In addition to the rotation given by $\omega_{\mu\nu}$, fluids can have an angular velocity that is related to an intrinsic spin density. In these spin fluids, the rotation associated with the spin density can be described by the tensor $\tilde{\omega}_{\mu\nu}$,[2]

$$\tilde{\omega}_{\mu\nu} = \dot{a}^{(i)}_{\mu} a_{\nu(i)}, \tag{2}$$

where $a^{(i)}_\mu$ is an orthonormal tetrad chosen so that $a^\mu{}_{(0)}$ lies along the velocity U^μ. Tetrad indices are in parentheses $(i) = (0$–$3)$ and are raised and lowered with the Minkowski metric $\eta_{(ij)} = (-1,+1,+1,+1)$. Tsoubelis,[3] using the spin-fluid stress-energy tensor of Ray and Smalley,[4,5] has shown that a Cartan spin fluid has the same frame dragging properties as a fluid with conventional vorticity. His result indicates that a spin fluid can serve as a source of angular momentum for astrophysical and cosmological models. Spin-fluid interiors can be matched to stationary vacuum exteriors.

In addition to direct astrophysical applications, spin fluids can also be used to model superfluid rotation detectors like those being developed by Cerdonio[6] for use in experimental relativity. The purpose of this paper is to explore the use of rotating spin-fluid sources. The calculation is performed in both general relativity and the Einstein Cartan self-consistent formalism. We impose cylindrical symmetry since the most immediate application of spin fluids is to rotating detectors. For constant rotational velocity, we find metric solutions of the van Stockum[7] type containing an unaccelerated spin fluid. There are some interesting differences between the Cartan and general relativistic fluid sources. The sources are compared and discussed in the last part of the paper. In the next sections we briefly review the spin-fluid stress-energy tensor and write the field equations leading to our solutions.

II. FORMALISM

A. Metric and tetrad

The space-time we consider is stationary and cylindrically symmetric with metric

$$ds^2 = -f\,dt^2 - 2K\,d\phi\,dt + l\,d\phi^2 + e^\mu(dr^2 + dz^2). \tag{3}$$

The orthonormal tetrad that diagonalizes this metric is

$$\begin{aligned} a^\mu{}_{(0)} &= (1/\sqrt{f},0,0,0), \\ a^\mu{}_{(1)} &= (0,e^{-\mu/2},0,0), \\ a^\mu{}_{(2)} &= (-K/D\sqrt{f},0,0,\sqrt{f}/D), \\ a^\mu{}_{(3)} &= (0,0,0,e^{-\mu/2}), \end{aligned} \tag{4}$$

where $D^2 = fl + K^2$. In terms of this tetrad, the comoving velocity is

$$U^\mu = a^\mu{}_{(0)} = (1/\sqrt{f},0,0,0). \tag{5}$$

B. Stress-energy tensor

A spin fluid is a fluid with an angular momentum density $S_{\mu\nu}$ defined throughout its extent. The spin density is constrained by the Frenkel[8] condition

$$U^\mu S_{\mu\nu} = 0. \tag{6}$$

Ray and Smalley[4] have developed a self-consistent Lagrangian formulation of the Einstein–Cartan theory with spin density. The stress-energy tensor derived from the Lagrangian variation can be written[5] as the sum of two parts:

$$T^{\mu\nu} = T^{\mu\nu}(\text{spin}) + T^{\mu\nu}(\text{fluid}), \tag{7}$$

where $T^{\mu\nu}(\text{fluid})$ is the perfect fluid stress-energy tensor

$$T^{\mu\nu}(\text{fluid}) = (\varepsilon + p) U^\mu U^\nu + p g^{\mu\nu}, \tag{8}$$

where ε is the energy density and p the pressure.

The spin-fluid portion of the stress-energy tensor is

$$T^{\mu\nu}(\text{spin}) = 2U^{(\mu} S^{\nu)\sigma} \dot{U}_\sigma + \nabla^*_\sigma [U^{(\mu} S^{\nu)\sigma}] - \tilde{\omega}^{\sigma(\mu} S^{\nu)}{}_\sigma. \tag{9}$$

The spin density $S^{\mu\nu}$ is related to the proper torsion $\hat{S}_{\mu\nu}{}^\sigma$ by

$$\hat{S}_{\mu\nu}{}^\sigma = \tfrac{1}{2}(8\pi G) S_{\mu\nu} U^\sigma. \tag{10}$$

The proper torsion is the trace-free part of the torsion $S_{\mu\nu}{}^\sigma$,

$$\hat{S}_{\mu\nu}{}^\sigma = S_{\mu\nu}{}^\sigma + \tfrac{2}{3}\delta^\sigma_{[\mu} S_{\nu]\alpha}{}^\alpha. \tag{11}$$

 0022-2488/88/071640-05$02.50

The quantity $S_{\nu\alpha}{}^{\alpha}$ is the torsion vector and describes that part of the torsion that does not satisfy the Frenkel condition. It is zero in our calculation. The ∇_σ^* derivative is $\nabla_\sigma^* = \nabla_\sigma + 2S_{\sigma\alpha}{}^{\alpha}$. It will be the covariant derivative in both calculations. Although the stress-energy tensor was constructed for self-consistent Cartan fluids, it is also valid in a self-consistent general relativity.[5] The only difference is in the derivative operator ∇_σ^*. It is the ordinary covariant derivative in general relativity. It is also the covariant derivative in the Cartan theory but with an additional spin connection calculated from the torsion:

$$\Gamma^{\sigma}{}_{\mu\nu} = S_{\mu\nu}{}^{\sigma} - S_{\nu}{}^{\sigma}{}_{\mu} + S^{\sigma}{}_{\mu\nu}. \tag{12}$$

Before writing the stress-energy tensor, it is useful to define the parameters describing the fluid.

C. Fluid parameters

The fluid parameters of interest are the acceleration and the rotation function $\omega_{\mu\nu}$. The acceleration is the same in general relativity and the Cartan theory,

$$\dot{U}_r = f_r/2\sqrt{f}. \tag{13}$$

We are interested in unaccelerated fluids so that f is a constant. We will choose

$$f = 1. \tag{14}$$

The fluid angular speed in Cartan theory is

$$\omega_{r\phi} = K_r/2\sqrt{f} - S_{r\phi}/2, \quad \omega_{r0} = f_r/2\sqrt{f}. \tag{15}$$

The term $S_{r\phi}$ in $\omega_{r\phi}$ comes from the spin connection. In general relativity these velocities are

$$\omega_{r\phi} = K_r/2\sqrt{f}, \quad \omega_{r0} = f_r/2\sqrt{f}. \tag{16}$$

Only $\omega_{r\phi}$ will be nonzero in the unaccelerated fluid. Because we have an unaccelerated fluid, the angular velocity associated with the spin, $\tilde{\omega}_{\mu\nu}$, is identical to the fluid vorticity $\omega_{\mu\nu}$.

III. FIELD EQUATIONS AND SOLUTIONS—GENERAL RELATIVITY

A. Field equations

Using the fluid parameters discussed in the last section, the stress-energy tensor components are found to be

$$\begin{aligned} T_{00} &= \varepsilon - W_0, \\ T_{rr} &= pe^{\mu} + (S_{r\phi}/D^2)K_r, \\ T_{zz} &= pe^{\mu}, \\ T_{\phi\phi} &= \varepsilon K^2 + D^2 p - KW_\phi + S_{r\phi}e^{-\mu}K_r, \\ T_{0\phi} &= \varepsilon K - \tfrac{1}{2}(W_\phi + KW_0). \end{aligned} \tag{17}$$

The quantity W_μ is the spin divergence $W_\mu = (S_\mu{}^{\nu})_{;\nu}$. The field equations in general relativity are

$$G_{\mu\nu} = T_{\mu\nu}.$$

We have taken $8\pi G = 1$. It is convenient to use the tetrad indexed components in writing the field equations:

$$\begin{aligned} G_{(ij)} &= a_{(i)}{}^{\mu} a_{(j)}{}^{\nu} G_{\mu\nu}, \\ T_{(ij)} &= a_{(i)}{}^{\mu} a_{(j)}{}^{\nu} T_{\mu\nu}, \end{aligned} \tag{18}$$

$$G_{(ij)} = T_{(ij)}. \tag{19}$$

The field equations are

$$(00): \frac{3K_r^2}{4D^2} - \frac{D_{rr}}{D} - \frac{\mu_{rr}}{2} = \varepsilon e^{\mu} + S_{r\phi}\frac{K_r}{D^2}, \tag{20}$$

$$(11): \frac{K_r^2}{4D^2} + \frac{\mu_r D_r}{2D} = pe^{\mu} + S_{r\phi}\frac{K_r}{D^2}, \tag{21}$$

$$(22): \frac{K_r^2}{4D^2} + \frac{\mu_{rr}}{2} = pe^{\mu} + S_{r\phi}\frac{K_r}{D^2}, \tag{22}$$

$$(33): \frac{-K_r^2}{4D^2} - \frac{\mu_r D_r}{2D} + \frac{D_{rr}}{D} = pe^{\mu}, \tag{23}$$

$$(02) + (20): \left(\frac{K_r}{D}\right)_r = \left(\frac{S_{r\phi}}{D}\right)_r. \tag{24}$$

B. Solutions

1. Spin and vorticity

One of the most interesting results of the calculation follows from integrating Eq. (24),

$$K_r/2D = S_{r\phi}/2D + c', \tag{25}$$

where c' is an integration constant, or

$$\omega_{r\phi}/D = S_{r\phi}/2D + c'. \tag{26}$$

This is a driving relation between the spin density and the vorticity or spin angular speed. Associated with the spin and rotation tensors are vector functions

$$\begin{aligned} S^{\mu} &= (\eta^{\mu\nu\alpha\beta}/2\sqrt{-g})\,U_\nu S_{\alpha\beta}, \\ \omega^{\mu} &= (\eta^{\mu\nu\alpha\beta}/2\sqrt{-g})\,U_\nu \omega_{\alpha\beta}. \end{aligned} \tag{27}$$

The driving relation is linear when written in terms of these vectors:

$$\omega_z = S_z/2 + c, \tag{28}$$

where $c = c'\eta^{0r\phi z}$.

2. Metric functions

Comparing Eqs. (21) and (22) the metric potential $\mu(r)$ can be related to D:

$$\mu_r(D_r/D) = \mu_{rr}$$

or

$$\mu_r = AD, \tag{29}$$

where A is a constant of integration.

Using this and eliminating the pressure between Eqs. (20) and (21), a differential equation for D is found:

$$D_{rr}/D - AD_r + 2(S_z^2/4 - c^2) = 0. \tag{30}$$

One solution of this differential equation is

$$D = r. \tag{31}$$

This solution corresponds to a constant spin S_z, and therefore a constant vorticity ω_z,

$$S_z = \text{const}, \quad \omega_z = \text{const}. \tag{32}$$

Equation (30) then gives the constant A:

$$A = 2(S_z^2/4 - c^2). \tag{33}$$

The other metric functions are

$$K = \omega_z r^2 + c_1, \quad u = Ar^2/2 + c_2, \quad l = r^2 - K^2, \tag{34}$$

where c_1 and c_2 are integration constants. This is a metric of the van Stockum type. The comparison will be made in detail in the last part of the paper.

3. Pressure and energy density

Since the choice $D = r$ requires S_z and ω_z to be constant, the driving relation, Eq. (28), can be reparametrized by defining

$$c = \omega_z(2 - n) \tag{35}$$

and

$$S_z = 2\omega_z(n - 1). \tag{36}$$

The pressure and energy density follow from the field equations

$$pe^{\mu} = -\omega_z S_z = \omega_z{}^2 2(1 - n), \tag{37}$$

$$\varepsilon e^{\mu} = 2\omega_z{}^2(5 - 3n). \tag{38}$$

The relation between the pressure and energy density is

$$\varepsilon - 3p = e^{-\mu} 4\omega_z{}^2, \tag{39}$$

where $n = 1$ is a spinless dust solution. Positive pressure and energy require the range of n to be

$$-\infty \leqslant n < 1. \tag{40}$$

IV. FIELD EQUATIONS AND SOLUTIONS—CARTAN

A. Field equations

The tetrad indexed stress-energy tensor is

$$\begin{aligned} T_{(00)} &= \varepsilon - W_0/\sqrt{f}, \\ T_{(11)} &= p + e^{-\mu}(S_{r\phi}/D^2)(K_r - S_{r\phi}), \\ T_{(22)} &= p + (S_{r\phi} e^{-\mu}/D^2)(K_r - S_{r\phi}), \\ T_{(33)} &= p, \\ T_{(02)} &= (KW_0 - W_\phi)/2D. \end{aligned} \tag{41}$$

The field equations in the Cartan theory are

$$G_{\mu\nu} = T_{\mu\nu} + \nabla^*_\sigma(T_{\mu\nu}{}^\sigma + T^\sigma{}_{\mu\nu} - T_\nu{}^\sigma{}_\mu), \tag{42}$$

where $T_{\mu\nu}{}^\sigma$ is the modified torsion

$$T_{\mu\nu}{}^\sigma = S_{\mu\nu}{}^\sigma + 2\delta^\sigma{}_{[\mu} S_{\nu]\alpha}{}^\alpha. \tag{43}$$

Since the torsion vector is zero, $T_{\mu\nu}{}^\sigma$ is just the Cartan spin connection and ∇^*_σ is the Cartan covariant derivative. The field equations are

$$(00): \frac{3K_r{}^2}{4D^2} - \frac{D_{rr}}{D} - \frac{\mu_{rr}}{2} = \varepsilon e^{\mu} + S_{r\phi}\frac{K_r}{D^2} - \frac{S_{r\phi}{}^2}{4D^2}, \tag{44}$$

$$(11): \frac{K_r{}^2}{4D^2} + \frac{\mu_r D_r}{2D} = pe^{\mu} + S_{r\phi}\frac{K_r}{D^2} - \frac{3}{4}\frac{S_{r\phi}{}^2}{D^2}, \tag{45}$$

$$(22): \frac{K_r{}^2}{4D^2} + \frac{\mu_{rr}}{2} = pe^{\mu} + S_{r\phi}\frac{K_r}{D^2} - \frac{3S_{r\phi}{}^2}{4D^2}, \tag{46}$$

$$(33): -\frac{K_r{}^2}{4D^2} - \frac{\mu_r D_r}{2D} + \frac{D_{rr}}{D} = pe^{\mu} - \frac{S_{r\phi}{}^2}{4D^2}, \tag{47}$$

$$(20) + (02): \left(\frac{K_r}{2D}\right)_r = \left(\frac{S_{r\phi}}{2D}\right)_r. \tag{48}$$

B. Solutions

1. Spin and angular speed

Integrating (48), a result similar to Eq. (25) is found:

$$K_r/2D = S_{r\phi}/2D + c'. \tag{49}$$

The Cartan vorticity is

$$\omega_{r\phi} = K_r/2 - S_{r\phi}/2. \tag{50}$$

Equation (50) is not a driving relation between vorticity and spin but a statement that the Cartan vorticity is constant.

The equation satisfied by D follows by eliminating the pressure between (44) and (45):

$$-D_{rr}/D + AD_r + 2c^2 = 0. \tag{51}$$

Because of the constraint on the Cartan vorticity, given by Eq. (50), the solution $D = r$ imposes no restrictions on the functional form of the spin density and rotation. The constant A is

$$A = -2c^2, \tag{52}$$

μ is the same as the general relativistic function

$$\mu_r = AD, \quad \mu = Ar^2/2 + c_2, \tag{53}$$

and A is, of course, different in general relativity and the Cartan theory.

The functional dependence of the metric potential K depends on the spin. Equation (49) can be written

$$K_r/2D = S_z/2 + c, \quad K = \int S_z r\, dr + cr^2 + c_1, \tag{54}$$

and

$$l = r^2 - k^2. \tag{55}$$

A constant spin, for example, gives

$$K = S_z r^2/2 + cr^2 + c_1. \tag{56}$$

A constant spin generates a metric of the van Stockum[7] type.

2. Pressure and energy density

The pressure and energy density can be obtained from the field equations

$$pe^{\mu} = -cS_z, \quad \varepsilon e^{\mu} = 4c^2 + cS_z, \tag{57}$$

with pressure and energy density satisfying

$$\varepsilon + p = 4c^2 e^{-2}. \tag{58}$$

For the special case of constant spin density, the van Stockum example, we can again define $c = \omega_0(2 - n)$, with ω_0 that part of the Cartan vorticity due to $K_r/2D$. This is a useful step to take in order to compare with the general relativity. The results are

$$\begin{aligned} S_z &= 2(n - 1)\omega_0, \\ p &= 2e^{\mu}\omega_0{}^2(n - 2)(n - 1), \\ \varepsilon &= 2e^{\mu}\omega_0{}^2(n - 3)(n - 2). \end{aligned} \tag{59}$$

The range of n for physical solutions is $n \leqslant 1$ and $n > 3$, where $n = 1$ gives a dust solution without spin.

V. COMPARISON TO THE VAN STOCKUM SOLUTION

The van Stockum interior solution can be written[9] as

$$ds^2 = -dt^2 + 2\alpha r^2\, dt\, d\phi + r^2(1-\alpha^2 r^2)d\phi^2 + e^{-\alpha^2 r^2}(dr^2+dz^2). \tag{60}$$

The fluid contained in this space-time is dust with energy density

$$\varepsilon = 4\alpha^2 e^{\alpha^2 r^2}. \tag{61}$$

The general relativistic metric is of this form with $c_1 = c_2 = 0$ in Eq. (34):

$$\begin{aligned} ds^2 &= -dt^2 - 2\omega_z r^2\, dt\, d\phi + r^2(1-\omega_z{}^2 r^2)d\phi^2 \\ &\quad + e^{[S_z^2/4 - c^2]r^2}(dr^2+dz^2) \\ &= -dt^2 - 2\omega_z r^2\, dt\, d\phi + r^2(1-\omega_z{}^2 r^2)d\phi^2 \\ &\quad + (dr^2+dz^2)e^{\omega_z^2(2n-3)r^2}, \end{aligned} \tag{62}$$

where ω_z is the fluid vorticity vector. This is identical to the van Stockum solution for $n=1$ and $|\omega_z| = \alpha$. The fluid is the van Stockum spinless dust. For other n values there is both spin and pressure. If $|n|$ is very large, the fluid equation of state approaches $\varepsilon = 3p$.

The Cartan metric for constant spin density can be written

$$\begin{aligned} ds^2 &= -dt^2 - 2r^2(c + S_z/2)dt\, d\phi \\ &\quad + r^2[1-(c+S_z/2)^2 r^2]d\phi^2 \\ &\quad + (dr^2+dz^2)e^{-c^2r^2}. \end{aligned} \tag{63}$$

The constant c is the Cartan vorticity vector, $n=1$, and $c = -\alpha$ generates the spinless dust van Stockum solution. For other n values, there is spin and pressure with an asymptotic equation of state $p = \varepsilon$. These constant spin van Stockum solutions will not match pressures to an exterior vacuum. This is always true for the general relativistic solution with its required constant spin. The Cartan solution allows nonconstant spins, and a vacuum match is possible in this case.

VI. THE VACUUM MATCH FOR THE CARTAN SOLUTIONS

The general solution for the Cartan calculations is

$$\begin{aligned} &ds^2 = -dt^2 - 2K\, d\phi\, dt + l\, d\phi^2 + e^{\mu}(dr^2+dz^2), \\ &\mu = -c^2r^2, \\ &K_r/2r = S_{r\phi}/2r + c, \\ &pe^{\mu} = -S_{r\phi}c/r, \\ &l = r^2 - K^2, \\ &\varepsilon e^{\mu} = cS_{r\phi}/r + 4c^2, \end{aligned} \tag{64}$$

where $S_{r\phi}$ and K_r are not yet specified.

This metric can be matched to a van Stockum exterior solution found by Bonner[10]:

$$\begin{aligned} &e^{\bar{\mu}} = e^{-1/4}(R/r)^{1/2}, \\ &\bar{l} = (rR/4)[3 + \log(r/R)], \\ &\bar{K} = -\tfrac{1}{2}r[1+\log(r/R)], \\ &\bar{f} = (r/R)[1-\log(r/R)]. \end{aligned} \tag{65}$$

This metric is one of three exterior van Stockum metrics found by Bonner; it is Petrov II. Here R is the boundary between interior and exterior.

The matching conditions[11] identify the first and second fundamental forms of the bounding surface, $r = R$. Matching the metrics, one finds two conditions

$$c^2R^2 = \tfrac{1}{4}, \tag{66}$$

$$K(R) = -R/2. \tag{67}$$

Identifying the second fundamental forms, there is one new relation

$$K_r(R) - S_{r\phi}(R) = \bar{K}_r(R)$$

or

$$K_r(R) - S_{r\phi}(R) = -1. \tag{68}$$

Using the relation requiring the Cartan vorticity to be constant,

$$K_r - S_{r\phi} = 2rc,$$

the sign of c in Eq. (66) is determined as

$$c = -1/2R. \tag{69}$$

The functional form of $K(r)$ and $S_{r\phi}(r)$ is constrained by these matching conditions:

$$K = \int S_{r\phi}\, dr + r^2c + C_0,$$

where C_0 is an integration constant. At $r = R$, this becomes

$$K(R) = \int_{(r=R)} S_{r\phi}\, dr - \frac{R}{2} + C_0. \tag{70}$$

Therefore at the boundary we have the spin constraint

$$\int_{(r=R)} S_{r\phi}\, dr = \text{const.} \tag{71}$$

In addition, if the pressure is to be zero at the vacuum boundary then the spin density is constrained to be zero at the boundary:

$$S_{r\phi}(R) = 0 \tag{72}$$

Any function satisfying (71) and (72) is acceptable. Choosing $S_{r\phi}(r) = rf(r)$, $f(r)$ regular at $r=0$, will produce finite pressure and energy at the origin.

VII. DISCUSSION AND CONCLUSION

General relativity and the Cartan theory both admit physical spin-fluid solutions for a rotating cylinder. The solutions in general relativity are only for constant spin and angular speed. The Cartan theory also has a constant spin solution but it is only one of many possibilities. Both constant spin sources lead to van Stockum-type metrics. The general relativistic source approaches the equation of state $\varepsilon = 3p$, and the Cartan source approaches $\varepsilon = p$, in the high spin limit.

There is another important difference between the Cartan and relativistic solutions in the case of constant spin. In both theories, the spin and $\omega_0 = K_r/2D$ are related through

$$S_z = 2\omega_0(n-1).$$

Both theories allow $n \ll 1$, giving S_z the opposite sense to

ω_0. The Cartan theory also allows $n > 3$; S_z and ω_0 can be parallel.

There is a very interesting driving relation between spin and cylinder speed in general relativity:

$$S_z/2 = \omega_z - c.$$

Vortex formation in rotating superfluids is one of the possible applications of the spin-fluid formalism. This driving relation, although in the continuum limit, could be interpreted in this context. Here c would be the critical speed for vortex formation to begin.

In conclusion, there are some similarities and also some fundamental differences between the two spin-fluid sources used for this calculation. With spin fluids being used as models in astrophysics and condensed matter physics, it is hopeful that a test of the two theories might occur in the near future.

[1]H. Stephani, *General Relativity* (Cambridge U.P., Cambridge, 1982), p. 166.
[2]F. Halbwachs, Prog. Theor. Phys. **24**, 291 (1960).
[3]D. Tsoubelis, Phys. Rev. Lett. **51**, 2235 (1983); Phys. Rev. D **30**, 1632 (1984).
[4]J. R. Ray and L. L. Smalley, Phys. Rev. D **26**, 2615, 2619 (1982).
[5]J. R. Ray and L. L. Smalley, Phys. Rev. D **27**, 1383 (1983); Phys. Rev. Lett. **49**, 1059 (1982); **50**, 626(E) (1983).
[6]M. Cerdonio, ICTP preprint, 1987.
[7]W. van Stockum, Proc. R. Soc. Edinburgh **57**, 135 (1937).
[8]J. Frenkel, Z. Phys. **37**, 243 (1926).
[9]J. N. Islam, *Rotating Fields in General Relativity* (Cambridge U.P., Cambridge, 1985), pp. 35 and 58.
[10]W. B. Bonner, J. Phys. A **13**, 2122 (1980).
[11]W. Arkuszewski, W. Kopczynski, and V. N. Ponomariev, Commun. Math. Phys. **45**, 183 (1975); W. B. Bonner and P. A. Vickers, Gen. Relativ. Gravit. **13**, 1 (1981).

Exact scalar and spinor solutions in some rotating universes

K. D. Krori, P. Borgohain, P. K. Kar, and Dipali Das (Kar)
Mathematical Physics Forum, Cotton College, Gauhati 781 001, India

(Received 27 October 1987; accepted for publication 30 December 1987)

In this paper exact solutions of the Klein–Gordon and Weyl equations in some rotating universes are presented.

I. INTRODUCTION

The study of gravitational interaction on quantum mechanical systems is a field of investigation recently explored by a number of authors. The Klein–Gordon equation and the Dirac equation in covariant form have been used in curved space-times for such studies.[1–6] The electromagnetic equation in Godel's universe have been studied by Cohen *et al.*[7] The Klein–Gordon and the Weyl equations have also been studied in the same universe by Pimentel and Macias.[8] In this paper we consider the same two equations in three other well-known rotating universes, viz., Som–Raychaudhuri, Hoenselaers–Vishveshwara, and Rebouças universes. Exact solutions of the two equations in all three cases are presented in this paper.

II. SCALAR SOLUTIONS

The governing equation for a massive scalar field with arbitrary coupling to the gravitational field can be taken in the form

$$(-\nabla_\alpha \nabla^\alpha + \xi R + m^2)\psi = 0, \tag{2.1}$$

where ξ is a real dimensionless coupling constant and $R = R_{\mu\nu}g^{\mu\nu}$ is the Riemann scalar.

A. Som–Raychaudhuri (SR) metric

The general form of the Som–Raychaudhuri metric is given by

$$ds^2 = dt^2 - dr^2 - dz^2 + 2qr^2\,d\phi\,dt - (r^2 - q^2r^4)d\phi^2. \tag{2.2}$$

On assuming $q = 1$, the metric (2.2) takes the form

$$ds^2 = dt^2 - dr^2 - dz^2 + 2r^2\,d\phi\,dt - r^2(1 - r^2)d\phi^2. \tag{2.3}$$

To calculate the Ricci scalar we consider the following basis one-forms:

$$\theta^1 = dr, \quad \theta^2 = r\,d\phi, \quad \theta^3 = dz, \quad \theta^0 = (dt + r^2\,d\phi), \tag{2.4}$$

so that

$$g_{ij} = \mathrm{diag}\{-1, -1, -1, +1\}.$$

Taking the exterior differentials of (2.4), expressing them in terms of the basis two-forms $\theta^i \wedge \theta^j$, and comparing the expressions with the first equation of structure

$$d\theta^i = -\omega^i_{\ j} \wedge \theta^j, \tag{2.5}$$

we get the following set of connection one-forms:

$$\omega^1{}_2 = \theta^0 - (1/r)\theta^2 = -\omega^2{}_1, \quad \omega^0{}_1 = \theta^2 = \omega^1{}_0, \quad \omega^0{}_2 = -\theta^1 = \omega^2{}_0. \tag{2.6}$$

We get the curvature two-forms $\Omega^i_{\ j}$ from the second equation of structure

$$\Omega^i_{\ j} = d\omega^i_{\ j} + \omega^i_{\ k} \wedge \omega^k_{\ j}, \tag{2.7}$$

and, finally, the components of the Riemann tensor $R^i_{\ jkl}$ from

$$\Omega^i_{\ j} = \tfrac{1}{2}R^i_{\ jkl}\theta^k \wedge \theta^l. \tag{2.8}$$

The nonvanishing components of $\Omega^i_{\ j}$ and $R_{\mu\nu}$ are found to be

$$\Omega^1{}_2 = 3\theta^1 \wedge \theta^2 = -\Omega^{12}, \quad \Omega^0{}_1 = \theta^1 \wedge \theta^0 = -\Omega^{01}, \quad \Omega^0{}_2 = \theta^2 \wedge \theta^0 = -\Omega^{02}, \tag{2.9}$$

and $R_{00} = 2$, $R_{11} = 2$, $R_{22} = 2$, and $R_{02} = 0$. The Ricci scalar is found to be

$$R = -2. \tag{2.10}$$

With Eqs. (2.2) and (2.10), Eq. (2.1) reduces to

$$[(1 - r^2)\partial^2_{tt} - \partial^2_{rr} - (1/r^2)\partial^2_{\phi\phi} - \partial^2_{zz} + 2\partial^2_{\phi t} - (1/r)\partial_r - p^2]\psi = 0, \tag{2.11}$$

where

$$p^2 = m^2 - 2\xi. \tag{2.12}$$

Let us choose ψ in the form

$$\psi = e^{ia(\omega t + l\phi + kz)}R(r), \tag{2.13}$$

where l, ω, and k are constants. Substituting (2.13) in (2.11) we obtain

$$R''(r) + (1/r)R'(r) + a^2[-l^2/r^2 - r^2\omega^2 + \lambda^2]R(r) = 0, \tag{2.14}$$

with

$$\lambda^2 = \omega^2 - k^2 + 2\omega l + m^2/a^2 - 2\xi/a^2. \tag{2.15}$$

Equation (2.14) is reduced to the confluent hypergeometric differential equation

$$yF'' + [(al + 1) - y]F' - [(1 + al)/2 - a\lambda^2/4\omega]F = 0, \tag{2.16}$$

if we make the substitution

$$R(r) = y^{al/2}e^{-y/2}F(y), \tag{2.17}$$

where $y = a\omega r^2$.

Equation (2.16) has two independent solutions given by

$${}_1F_1(b,c,y)$$

and

$$y^{1-c}\,{}_1F_1[(1 + b - c), 2 - c, y],$$

where

$$b = (1 + al)/2 - a\lambda^2/4\omega, \quad c = 1 + al. \tag{2.18}$$

 0022-2488/88/071645-05$02.50

The general solution, which satisfies the usual boundary condition that $R(r)$ will decay when $r\to\infty$, is given by

$$U(b,c,y)=\frac{\Gamma(1-c)}{\Gamma(1+b-c)}\,{}_1F_1(b,c,y)+\frac{\Gamma(c-1)}{\Gamma(b)}\,y^{1-c}\,{}_1F_1[(1+b-c),2-c,y]. \tag{2.19}$$

B. Hoenselaers–Vishveshwara (HV) metric

The Hoenselaers–Vishveshwara metric is given by

$$ds^2=dt^2-dr^2-dz^2+\tfrac{1}{2}A^2(c-1)(c-3)d\phi^2+2(c-1)d\phi\,dt, \tag{2.20}$$

where $c=\cosh\sqrt{K}r$ and $A\sqrt{K}=1$. On assuming $A=1$, $K=1$, the metric (2.20) takes the form

$$ds^2=dt^2-dr^2-dz^2+\tfrac{1}{2}(c-1)(c-3)d\phi^2+2(c-1)d\phi\,dt. \tag{2.21}$$

The Ricci scalar is calculated as in the previous case considering the following basis one-forms:

$$\theta^1=dr,\quad \theta^2=\sqrt{(c^2-1)/2}\,d\phi,$$
$$\theta^3=dz,\quad \theta^0=dt+(c-1)d\phi. \tag{2.22}$$

The nonvanishing components of ω^i_j, Ω^i_j, and $R_{\mu\nu}$ are found to be

$$\omega^1{}_2=\theta^0/\sqrt{2}-(c/\sqrt{c^2-1})\theta^2=-\omega^2{}_1,$$
$$\omega^0{}_1=\tfrac{1}{2}\theta^2=\omega^1{}_0,\quad \omega^0{}_2=-(1/\sqrt{2})\theta^1=\omega^2{}_0,$$
$$\Omega^1{}_2=\tfrac{1}{2}\theta^1\wedge\theta^2=-\Omega^{12},$$
$$\Omega^0{}_1=\tfrac{1}{2}\theta^1\wedge\theta^0=-\Omega^{01},$$
$$\Omega^0{}_2=\tfrac{1}{2}\theta^2\wedge\theta^0=-\Omega^{02},$$
$$R_{11}=0,\quad R_{22}=0,\quad R_{02}=0,\quad R_{00}=1.$$

The Ricci scalar is then

$$R=1. \tag{2.23}$$

Using (2.21) and (2.23) in (2.1) we get

$$\left[\partial^2_{rr}-\frac{2}{1-c^2}\partial^2_{\phi\phi}+\partial^2_{zz}+\frac{c-3}{c+1}\partial^2_{tt}-\frac{4}{c+1}\partial^2_{t\phi}+\left(\frac{1}{\sqrt{c^2-1}}+\frac{\sqrt{c^2-1}}{\sqrt{c+1}}\right)\partial_r+\mu^2\right]\psi=0, \tag{2.24}$$

with $\mu^2=m^2+\xi$. Let us now choose ψ in the form

$$\psi=e^{i\alpha(\omega t+l\phi+kz)}R(r), \tag{2.25}$$

where ω, l, and k are constants. Substituting (2.25) in (2.24) we get the equation

$$R''(r)+R'(r)\coth r+[A_1+A_2\coth r\,\mathrm{csch}\,r+A_3\,\mathrm{csch}^2 r]R(r)=0, \tag{2.26}$$

with

$$A_1=m^2+\xi-a^2\omega^2-a^2k^2,$$
$$A_2=4a^2\omega(\omega+l),$$
$$A_3=-4a^2(l^2/2+\omega l+\omega^2). \tag{2.27}$$

Substituting

$$R(r)=y^{m+n}(1-y)^nU(y),$$

where

$$y=(1+\cosh r)/2,\quad m=-al/\sqrt{2},\quad n=-\sqrt{2}a\omega, \tag{2.28}$$

in Eq. (2.26), we get the following equation:

$$y(1-y)U''(y)+[(1+\sqrt{2}al-2\sqrt{2}a\omega)-(2-2\sqrt{2}al-2\sqrt{2}a\omega)y]U'(y)-[2a^2l^2+a^2\omega^2+4a^2\omega l-\sqrt{2}al-\sqrt{2}a\omega-a^2k^2+m^2+\xi]U(y)=0. \tag{2.29}$$

Equation (2.29) is the Gauss hypergeometric equation. The two linearly independent solutions of (2.29) are

$$F(a',b';a'+b'-c'+1;1-y)$$

and

$$(1-y)^{c'-a'-b'}F(c'-a',c'-b';c'-a'-b'+1;1-y).$$

The general solution, which satisfies the general boundary condition that $R(r)$ will decay when $r\to\infty$, is

$$F(a',b';c';y)=\frac{\Gamma(c')\Gamma(c'-a'-b')}{\Gamma(c'-a')\Gamma(c'-b')}\times F(a',b';a'+b'-c'+1;1-y)+(1-y)^{c'-a'-b'}\frac{\Gamma(c')\Gamma(a'+b'-c')}{\Gamma(a')\Gamma(b')}\times F(c'-a',c'-b';c'-a'-b'+1;1-y), \tag{2.30}$$

with

$$a'=\tfrac{1}{2}[1-2\sqrt{2}al-2\sqrt{2}a\omega+\sqrt{1+4a^2\omega^2+4a^2k^2-4\mu^2}\,],$$
$$b'=\tfrac{1}{2}[1-2\sqrt{2}al-2\sqrt{2}a\omega-\sqrt{1+4a^2\omega^2+4a^2k^2-4\mu^2}\,],$$
$$c'=1-\sqrt{2}al-2\sqrt{2}a\omega. \tag{2.31}$$

C. Rebouças metric

The Rebouças metric is

$$ds^2=dt^2-dr^2-dz^2+4\cosh 2r\,d\phi\,dt+(3\cosh^2 2r+1)d\phi^2. \tag{2.32}$$

The Ricci scalar is calculated as before considering the following basis one-forms:

$$\theta^1=dr,\quad \theta^2=\sinh 2r\,d\phi,\quad \theta^3=dz,$$
$$\theta^0=dt+2\cosh 2r\,d\phi. \tag{2.33}$$

The nonvanishing components of ω^i_j, Ω^i_j, and R_{ij} in this case are found to be

$$\omega^0{}_1=2\theta^2=\omega^1{}_0,\quad \omega^0{}_2=-2\theta^1=\omega^2{}_0,$$
$$\omega^1{}_2=-2\coth 2r\theta^2+2\theta^0=-\omega^2{}_1,$$
$$\Omega^1{}_2=8\theta^1\wedge\theta^2=-\Omega^{12},$$
$$\Omega^0{}_1=4\theta^1\wedge\theta^0=-\Omega^{01},$$
$$\Omega^0{}_2=4\theta^2\wedge\theta^0=-\Omega^{02},$$
$$R_{00}=8,\quad R_{11}=4,\quad R_{22}=4.$$

The Ricci scalar for the metric is found to vanish, i.e.,

$$R = 0. \tag{2.34}$$

Using (2.32) and (2.34) in (2.1) we get

$$\left[\partial_{rr}^2 + \frac{1}{\sinh^2 2r}\partial_{\phi\phi}^2 + \partial_{zz}^2 + \frac{3\cosh^2 2r + 1}{\sinh^2 2r}\partial_{tt}^2 - \frac{4\cosh 2r}{\sinh^2 2r}\partial_{t\phi}^2 + 2\coth 2r\,\partial_r + m^2\right]\psi = 0. \tag{2.35}$$

Let us choose ψ as before in the form

$$\psi = e^{ia(\omega t + l\phi + kz)}R(r), \tag{2.36}$$

where a, ω, l, and k are constants.

Putting (2.36) in (2.35) we get

$$R''(r) + 2\coth 2rR'(r) + [-(a^2l^2 + 4a^2\omega^2)\operatorname{csch}^2 2r + 4a^2\omega l\coth 2r\operatorname{csch} 2r + (m^2 - 3a^2\omega^2 - a^2k^2)] \times R(r) = 0. \tag{2.37}$$

Equation (2.37) is reduced to Gauss' hypergeometric differential equation,

$$y(1-y)U''(y) + [(-a\omega - al/2 + 1) - (2 - 2a\omega)y]U'(y) - \tfrac{1}{4}[a^2\omega^2 - 4a\omega - a^2k^2 + m^2]U(y) = 0, \tag{2.38}$$

if we make the following substitutions:

$$R(r) = y^p(1-y)^qU(y),$$

with

$$y = \tfrac{1}{2}(1 + \cosh 2r), \quad p = -(a/2)(\omega + l/2), \quad q = -(a/2)(\omega - l/2). \tag{2.39}$$

The general solution of Eq. (2.38), which satisfies the usual boundary condition that $R(r)$ will decay when $r \to \infty$, is

$$F(b,c,d,y) = \frac{\Gamma(d)\Gamma(d-b-c)}{\Gamma(d-b)\Gamma(d-c)} \times F(b,c;b+c-d+1;1-y) + (1-y)^{d-b-c}\frac{\Gamma(d)\Gamma(b+c-d)}{\Gamma(b)\Gamma(c)} \times F(d-b,d-c;d-b-c+1;1-y), \tag{2.40}$$

where

$$b = \tfrac{1}{2}\{1 - 2a\omega + \sqrt{1 - m^2 + 3a^2\omega^2 + a^2k^2}\},$$
$$c = \tfrac{1}{2}[1 - 2a\omega - \sqrt{1 - m^2 + 3a^2\omega^2 + a^2k^2}], \tag{2.41}$$
$$d = 1 - a\omega - al/2.$$

III. SPINOR SOLUTIONS

The Weyl equations for a massless spin-$\frac{1}{2}$ field are

$$\gamma^\alpha\nabla_\alpha\psi = 0, \tag{3.1}$$
$$(1 + \gamma^5)\psi = 0, \tag{3.2}$$

where

$$\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$$

and

$$\gamma^\alpha = h_{(a)}{}^\alpha\gamma^{(a)}. \tag{3.3}$$

The γ^α are generalized Dirac matrices and are given in terms of the flat space-time gammas $\gamma^{(a)}$, and $h_{(a)}{}^\alpha$ are vierbeins defined by the relations

$$h_{(a)}{}^\alpha h_{(b)}{}^\beta\eta^{(a)(b)} = g^{\alpha\beta}. \tag{3.4}$$

A. SR metric

First we consider the Weyl equations in the SR metric. For this metric we shall use the following set of vierbeins:

$$h_{(1)}{}^\alpha = \delta^\alpha{}_1, \quad h_{(2)}{}^\alpha = r[\delta^\alpha{}_0 - (1/r^2)\delta^\alpha{}_2], \quad h_{(3)}{}^\alpha = \delta^\alpha{}_3, \quad h_{(0)}{}^\alpha = \delta^\alpha{}_0. \tag{3.5}$$

Now we have

$$\nabla_\alpha\psi = [\partial_\alpha - T_\alpha]\psi, \tag{3.6}$$

where

$$T_\alpha = -\tfrac{1}{4}\gamma^{(a)}\gamma^{(b)}h_{(a)}{}^\nu h_{(b)\nu,\alpha}, \tag{3.7}$$

and

$$h_{(b)\nu,\alpha} = \partial_\alpha h_{(b)\nu} - \Gamma_{\nu\alpha}^\lambda h_{(b)\lambda}. \tag{3.8}$$

Using Eqs. (3.6)–(3.8) we obtain

$$T_1 = -\tfrac{1}{2}\gamma^{(0)}\gamma^{(2)},$$
$$T_2 = (r/2)\gamma^{(1)}\gamma^{(0)} + \tfrac{1}{2}(r^2 - 1)\gamma^{(2)}\gamma^{(1)}, \tag{3.9a}$$
$$T_3 = 0, \quad T_0 = \tfrac{1}{2}\gamma^{(2)}\gamma^{(1)}.$$

We now choose ψ in the form

$$\psi = e^{-ia(\omega t + l\phi + kz)}\begin{pmatrix}\eta_1(r)\\ \eta_2(r)\end{pmatrix}, \tag{3.9b}$$

where η_1 and η_2 are two-component spinors.

Using the standard representations of gamma matrices Eq. (3.2) is reduced to

$$\eta_1(r) = \eta_2(r) = \begin{pmatrix}R_1(r)\\ R_2(r)\end{pmatrix}. \tag{3.10}$$

On using Eqs. (3.6)–(3.10) in Eq. (3.1) we get

$$\left[\partial_r + \frac{al}{r} - a\omega r - \frac{1}{2r}\right]R_2 = ia\left[k + \omega + \frac{1}{2a}\right]R_1, \tag{3.11}$$

$$\left[\partial_r - \frac{al}{r} + a\omega r - \frac{1}{2r}\right]R_1 = ia\left[\omega - k - \frac{1}{2a}\right]R_2. \tag{3.12}$$

Eliminating R_1 from (3.11) and (3.12) we get

$$r^2R_2'' - rR_2' + [(\tfrac{3}{4} - al - a^2l^2) + \lambda^2r^2 - a^2\omega^2r^4]R_2 = 0, \tag{3.13}$$

where

$$\lambda^2 = 2a^2\omega l - a\omega + a^2\omega^2 - a^2k^2 + ak - \tfrac{1}{4}. \tag{3.14}$$

This equation can be reduced to the following confluent hypergeometric differential equation:

$$yF'' + (2k_1 - y)F' - (k_1 - \lambda^2/4a\omega)F = 0, \tag{3.15}$$

if we make the substitutions

$$R_2(r) = y^{k_1}e^{-y/2}F(y), \tag{3.16}$$

where $y = a\omega r^2$, and k_1 is a constant.

Equation (3.15) has two independent solutions given by

$${}_1F_1(k_1 - \lambda^2/4a\omega, 2k_1, y)$$

and

$$y^{1-2k_1}{}_1F_1(1 - k_1 - \lambda^2/4a\omega, 2 - 2k_1, y).$$

The general solution, which satisfies the usual boundary condition that $R(r)$ will decay when $r\to\infty$, is given by

$$U\left(k_1-\frac{\lambda^2}{4a\omega},2k_1,y\right)$$
$$=\frac{\Gamma(1-2k_1)}{\Gamma(1-k_1-\lambda^2/4a\omega)}\,{}_1F_1\left(k_1-\frac{\lambda^2}{4a\omega},2k_1,y\right)$$
$$+\frac{\Gamma(2k_1-1)}{\Gamma(k_1-\lambda^2/4a\omega)}\,y^{1-2k_1}$$
$$\times{}_1F_1\left(1-k_1-\frac{\lambda^2}{4a\omega},2-2k_1,y\right). \tag{3.17}$$

B. HV metric

For the HV metric we choose the following set of vierbeins:

$$h_{(0)}{}^{\alpha}=\delta^{\alpha}{}_0,\quad h_{(1)}{}^{\alpha}=\delta^{\alpha}{}_1,\quad h_{(3)}{}^{\alpha}=\delta^{\alpha}{}_3,$$
$$h_{(2)}{}^{\alpha}=\left[\frac{-\sqrt{2}}{\sqrt{c^2-1}}\,\delta^{\alpha}{}_2+\sqrt{\frac{2(c-1)}{c+1}}\,\delta^{\alpha}{}_0\right]. \tag{3.18}$$

Using Eqs. (3.6)–(3.8) in this case we obtain

$$T_1=-(\sqrt{2}/4)\gamma^{(0)}\gamma^{(2)},$$
$$T_2=(\sqrt{c^2-1}/4)\gamma^{(1)}\gamma^{(0)}-(\sqrt{2}/4)\gamma^{(2)}\gamma^{(1)}, \tag{3.19}$$
$$T_3=0,\quad T_0=(\sqrt{2}/4)\gamma^{(2)}\gamma^{(1)}.$$

As before, we choose

$$\psi=e^{-ia(\omega t+l\phi+kz)}\begin{pmatrix}\eta_1(r)\\ \eta_2(r)\end{pmatrix}, \tag{3.20}$$

where η_1 and η_2 are two-component spinors. Using standard representation of gamma matrices, Eq. (3.2) is reduced to

$$\eta_1(\gamma)=\eta_2(\gamma)=\begin{pmatrix}R_1(\gamma)\\ R_2(\gamma)\end{pmatrix}. \tag{3.21}$$

Using Eqs. (3.6)–(3.8) and (3.18)–(3.21) in Eq. (3.1) we get

$$\left[\partial_r+\frac{\sqrt{2}al}{\sqrt{c^2-1}}-\frac{\sqrt{2}(c-1)}{\sqrt{c^2-1}}\,a\omega-\frac{c}{2\sqrt{c^2-1}}\right]R_2$$
$$=ia\left[\omega+k+\frac{\sqrt{2}}{4a}\right]R_1, \tag{3.22}$$

and

$$\left[\partial_r-\frac{\sqrt{2}al}{\sqrt{c^2-1}}+\frac{\sqrt{2}(c-1)}{\sqrt{c^2-1}}\,a\omega-\frac{c}{2\sqrt{c^2-1}}\right]R_1$$
$$=ia\left[\omega-k-\frac{\sqrt{2}}{4a}\right]R_2. \tag{3.23}$$

Eliminating R_2 from (3.22) and (3.23) we get

$$R_1''-\coth rR_1'$$
$$+[A+B\coth r\,\mathrm{csch}\,r+D\,\mathrm{csch}^2 r]R_1=0, \tag{3.24}$$

where

$$A=-a^2\omega^2-a^2k^2-(1/\sqrt{2})ak+\tfrac{1}{8},$$
$$B=4a^2\omega^2+4a^2\omega l+\sqrt{2}a\omega+\sqrt{2}al, \tag{3.25}$$
$$D=-4a^2\omega^2-4a^2\omega l-2a^2l^2-\sqrt{2}a\omega+\tfrac{3}{4}.$$

Equation (3.24) will be reduced to Gauss' hypergeometric equation,

$$y(1-y)U''(y)+[(\tfrac{1}{2}-\sqrt{2}al-2\sqrt{2}a\omega)$$
$$-(2-2\sqrt{2}al-2\sqrt{2}a\omega)y]U'(y)$$
$$-(\tfrac{1}{8}+2a^2l^2-\sqrt{2}al+a^2\omega^2-\sqrt{2}a\omega$$
$$+4a^2\omega l-a^2k^2-(1/\sqrt{2})ak)U(y)=0, \tag{3.26}$$

if we make the following substitutions:

$$R_1(r)=y^{m+n}(1-y)^mU(y),$$
$$m=-(1/\sqrt{2})al+\tfrac{3}{4},\quad n=-\sqrt{2}a\omega-\tfrac{1}{2},$$
$$y=\tfrac{1}{2}(1+\cosh r). \tag{3.27}$$

The two linearly independent solutions of (3.26) are

$$F(a,b;a+b-c+1;1-y)$$

and

$$(1-y)^{c-a-b}F(c-a,c-b;c-a-b+1;1-y).$$

The general solution, which satisfies the usual boundary conditions that $R_1(r)$ will decay when $r\to\infty$, is

$$F(a,b;c;y)=\frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$$
$$\times F(a,b;a+b-c+1;1-y)$$
$$+(1-y)^{c-a-b}\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(c)}$$
$$\times F(c-a,c-b;c-a-b+1;1-y), \tag{3.28}$$

where

$$a=\tfrac{1}{2}[1-2\sqrt{2}al-2\sqrt{2}a\omega$$
$$+\sqrt{\tfrac{1}{2}+4a^2\omega^2+4a^2k^2+2\sqrt{2}ak}\,],$$
$$b=\tfrac{1}{2}[1-2\sqrt{2}al-2\sqrt{2}a\omega \tag{3.29}$$
$$-\sqrt{\tfrac{1}{2}+4a^2\omega^2+4a^2k^2+2\sqrt{2}ak}\,],$$
$$c=\tfrac{1}{2}-\sqrt{2}al-2\sqrt{2}a\omega.$$

C. Rebouças metric

For the Rebouças metric we choose the following set of vierbeins:

$$h_{(1)}{}^{\alpha}=\delta^{\alpha}{}_1,\quad h_{(3)}{}^{\alpha}=\delta^{\alpha}{}_3,\quad h_{(0)}{}^{\alpha}=\delta^{\alpha}{}_0,$$
$$h_{(2)}{}^{\alpha}=(1/\sqrt{c^2-1})[2c\delta^{\alpha}{}_0-\delta^{\alpha}{}_2]. \tag{3.30}$$

We have as before

$$\nabla_{\alpha}\psi=[\partial_{\alpha}-T_{\alpha}]\psi,$$

with

$$T_{\alpha}=-\tfrac{1}{4}\gamma^{(a)}\gamma^{(b)}h_{(a)}{}^{\nu}h_{(b)\nu,\alpha}$$

and

$$h_{(b)\nu,\alpha}=\partial_{\alpha}h_{(b)\nu}-\Gamma^{\lambda}_{\nu\alpha}h_{(b)\lambda},$$

whence we get

$$T_1=-\gamma^{(0)}\gamma^{(2)},$$
$$T_2=\sqrt{c^2-1}\,\gamma^{(1)}\gamma^{(0)}+2c\gamma^{(2)}\gamma^{(1)}, \tag{3.31}$$
$$T_3=0,\quad T_0=\gamma^{(2)}\gamma^{(1)}.$$

As before we consider

$$\psi = e^{-ia(\omega t + l\phi + kz)} \begin{pmatrix} \eta_1(r) \\ \eta_2(r) \end{pmatrix} \tag{3.32}$$

where η_1 and η_2 are two-component spinors. We get from (3.2)

$$\eta_1(r) = \eta_2(r) = \begin{pmatrix} R_1(r) \\ R_2(r) \end{pmatrix}. \tag{3.33}$$

Using (3.31)–(3.33) in (3.1) we get

$$\left[\partial_r + \frac{al}{\sqrt{c^2-1}} - \frac{2ca\omega}{\sqrt{c^2-1}} - \frac{c}{\sqrt{c^2-1}}\right] R_2 = ia\left[\omega + k + \frac{1}{a}\right] R_1, \tag{3.34}$$

$$\left[\partial_r - \frac{al}{\sqrt{c^2-1}} + \frac{2ca\omega}{\sqrt{c^2-1}} - \frac{c}{\sqrt{c^2-1}}\right] R_1 = ia\left[\omega - k - \frac{1}{a}\right] R_2. \tag{3.35}$$

Eliminating R_2 from (3.34) and (3.35) we get

$$[\partial_{rr}^2 - 2\coth 2r\,\partial_r + (3 - a^2l^2 - 4a\omega - 4a^2\omega^2)\operatorname{csch}^2 2r + (2al + 4a^2\omega l)\coth 2r \operatorname{csch} 2r - (3a^2\omega^2 + a^2k^2 + 2ak)]R_1 = 0. \tag{3.36}$$

Equation (3.36) is now reduced to Gauss' hypergeometric equation,

$$y(1-y)U''(y) + \left[\left(\frac{1}{2} - \frac{al}{2} - a\omega\right) - (1 - 2a\omega)y\right] U'(y) - \left[\frac{a^2\omega^2}{4} - \frac{a^2k^2}{4} - \frac{ak}{2} - \frac{1}{4}\right] U(y) = 0, \tag{3.37}$$

if we make the following substitutions:

$$R_1(r) = y^p(1-y)^q U(y),$$

where

$$y = \tfrac{1}{2}(1 + \cosh 2r),$$
$$p = -a\omega/2 - al/4 + \tfrac{1}{4}, \tag{3.38}$$
$$q = -a\omega/2 + al/4 + \tfrac{1}{4}.$$

The general solution of Eq. (3.37), which satisfies the usual boundary condition that $R_1(r)$ will decay when $r \to \infty$, is

$$F(a',b',c';y) = \frac{\Gamma(c')\Gamma(c'-a'-b')}{\Gamma(c'-a')\Gamma(c'-b')} \times F(a',b';a'+b'-c'+1;1-y) + (1-y)^{c'-a'-b'} \frac{\Gamma(c')\Gamma(a'+b'-c')}{\Gamma(a')\Gamma(b')} \times F(c'-a',c'-b';c'-a'-b'+1;1-y), \tag{3.39}$$

with

$$a' = -a\omega + \tfrac{1}{2}\sqrt{3a^2\omega^2 + a^2k^2 + 2ak + 1}$$
$$b' = -a\omega - \tfrac{1}{2}\sqrt{3a^2\omega^2 + a^2k^2 + 2ak + 1} \tag{3.40}$$
$$c' = \tfrac{1}{2} - al/2 - a\omega.$$

ACKNOWLEDGMENTS

The authors express their profound gratitude to the Government of Assam for all facilities provided at Cotton College, Guahati 781 001, India, to carry out the investigations reported in this paper.

[1] J. Audretsch and G. Schafer, Gen. Relativ. Gravit. **9**, 243 (1978).
[2] E. Fischback, B. S. Freeman, and W. Cheng, Phys. Rev. D **23**, 2157 (1981).
[3] L. Parker and L. O. Pimentel, Phys. Rev. D **25**, 3180 (1982).
[4] G. Y. Li and H. Y. Guo, Acta. Phys. Sin. **31**, 1501 (1982).
[5] M. Parsic, Phys. Lett. A **90**, 175 (1982).
[6] P. Bureev, Gen. Relativ. Gravit. **17**, 799 (1985).
[7] J. M. Cohen, C. V. Vishveshwara, and S. V. Dhurandhar, J. Phys. **13**, 933 (1980).
[8] L. O. Pimentel and A. Macias, Phys. Lett. A **117**, 325 (1986).

Symmetries of Einstein's field equations with a perfect fluid source as examples of Lie–Bäcklund symmetries

H. Stephani
Sektion Physik der Friedrich-Schiller-Universität, Max-Wien-Platz 1, DDR-6900 Jena, German Democratic Republic

(Received 29 January 1988; accepted for publication 2 March 1988)

The framework of Lie–Bäcklund (or generalized) symmetries is used to give a unifying view of some of the known symmetries of Einstein's field equations for the vacuum or perfect fluid case (with a $\mu = p$ or a $\mu + 3p = 0$ equation of state). These symmetries occur if space-time admits one or two Killing vectors (orthogonal or parallel, respectively, to the four-velocity in the perfect fluid case).

I. INTRODUCTION

When looking for exact solutions of Einstein's field equations with a perfect fluid source,

$$R_{ab} = T_{ab} - Tg_{ab}/2 = (\mu + p)u_a u_b + (\mu - p)g_{ab}/2 , \tag{1.1}$$

it certainly is good advice to use the symmetries of the problem as much as possible. Although an infinite number of those symmetries does always exist, it is very difficult and in most cases even practically impossible to find and to use them. The symmetries we have in mind are the Lie–Bäcklund, or generalized, symmetries[1–3]; we shall discuss them briefly in Sec. II.

Symmetries of the vacuum field equations are known if one or more Killing vectors are assumed. Then it is also well known how to exploit these symmetries for generating new solutions. For the case of one Killing vector, we will formulate the relevant results in terms of Lie–Bäcklund symmetries in Sec. III. These results can easily be generalized to perfect fluid solutions if they obey an equation of state $\mu = p$ (or $\mu + 3p = 0$), and if the Killing vector is orthogonal (or parallel) to the fluid's four-velocity u^a. We will do this in Sec. IV.

If there are two commuting Killing vectors that possess two-surfaces orthogonal to the group orbits, then in the vacuum case there exist numerous techniques to use the many symmetries explicitly known. We will sketch briefly in Sec. V how they are related to Lie–Bäcklund symmetries. If the Killing vectors are orthogonal to u^a, then all this can be carried over to the perfect fluid case. But an additional symmetry also exists that can be used to generate the general perfect fluid solution (with $\mu = p$) from the vacuum (Sec. VI).

We close with some remarks on possible generalization to other symmetries and/or sources (energy-momentum tensors) in Sec. VII.

II. LIE–BÄCKLUND SYMMETRIES OF EINSTEIN'S FIELD EQUATIONS

By definition, Lie–Bäcklund symmetries map solutions of ordinary or partial differential equations into solutions. If such symmetries exist and form a group depending on (at least) one arbitrary parameter ϵ, then any (general) solution $\{g_{ab}, u_a, \mu, p\}$ of Einstein's equations (1.1) will be a member of a one-parameter set $\{\tilde{g}_{ab}(\epsilon), \tilde{u}_a(\epsilon), \tilde{\mu}(\epsilon), \tilde{p}(\epsilon)\}$ of solutions connected by the action of the symmetry.

To any such symmetry, an infinitesimal generator X defined by

$$X = f_{ab}\frac{\partial}{\partial g_{ab}} + v_a\frac{\partial}{\partial u_a} + m\frac{\partial}{\partial \mu} + \pi\frac{\partial}{\partial p} ,$$

$$f_{ab} = Xg_{ab} = \left.\frac{\partial \tilde{g}_{ab}(\epsilon)}{\partial \epsilon}\right|_{\epsilon=0} , \quad v_a = Xu_a = \left.\frac{\partial \tilde{u}_a}{\partial \epsilon}\right|_{\epsilon=0} ,$$

$$m = X\mu = \left.\frac{\partial \tilde{\mu}}{\partial \epsilon}\right|_{\epsilon=0} , \quad \pi = Xp = \left.\frac{\partial \tilde{p}}{\partial \epsilon}\right|_{\epsilon=0} , \tag{2.1}$$

can be associated. Since the symmetry under consideration maps solutions into solutions, the field equations have to be invariant under the action of X, i.e., we must have

$$XR_{ab} = X(T_{ab} - Tg_{ab}/2) . \tag{2.2}$$

To evaluate the left-hand side we have, of course, to extend X to the derivatives of the metric. Since X commutes with taking the partial derivative,

$$Xg_{ab,i} = (Xg_{ab})_{,i} = f_{ab,i} , \tag{2.3}$$

we find after a short calculation that

$$X\Gamma^r_{st} = \tfrac{1}{2}g^{rn}(f_{ns;t} + f_{nt;s} - f_{st;n}) \tag{2.4}$$

and

$$XR_{ab} = \tfrac{1}{2}(f^n{}_{a;bn} + f^n{}_{b;an} - f_{ab}{}^{;n}{}_{;n} - f^n{}_{n;ab}) \tag{2.5}$$

hold.

Conversely, if we can find functions f_{ab}, v_a, m, π so that the symmetry condition (2.2) is satisfied, then these functions determine the generator of a symmetry, and the finite symmetry transformation can be constructed by exponentiating this generator. Application of this transformation to a known solution will yield a one-parameter family of solutions.

The difficulties hidden in this appearingly simple approach to symmetries are, first of all, that the symmetry condition (2.2) has to be fulfilled for *all* solutions of the field equations, i.e., for *all* solutions to which the symmetry will be applied. That means that (2.2) must hold identically in

 0022-2488/88/071650-05$02.50

the field variables $\{g_{ab},u_a,\mu,p\}$ and all their derivatives [occurring, e.g., in the covariant derivatives in (2.5)]. We therefore have to expect, e.g., the f_{ab} to depend on the metric g_{ab} (and on u_a, μ, and p) in a very complicated, nonlocal way, possibly by depending on the curvature tensor and its derivatives of arbitrary high order. To make the symmetry condition less stringent one therefore often narrows the class of solutions for which it has to be satisfied by imposing an extra constraint, mostly by assuming the existence of a Killing vector. This constraint has to be invariant under X, too, but it can be used to eliminate terms from the symmetry condition.

The second difficulty arises when one tries to exponentiate the infinitesimal generator X. If for brevity we denote the set of independent functions by V_A, then the equations we have to exponentiate are of the form

$$XV_A = W_A\,, \tag{2.6}$$

where the W_A depend on the V_A *and their derivatives*. In fact, this means that the system

$$\frac{\partial \tilde{V}_A}{\partial \epsilon} = W_A(\tilde{V}_A,\tilde{V}_{A,a,\ldots})\,, \quad \tilde{V}_A|_{\epsilon=0} = V_A \tag{2.7}$$

we have to solve to obtain the finite symmetry transformations $\tilde{V}_A = \tilde{V}_A(V_B,\epsilon)$ is a system of partial differential equations of perhaps greater complexity than the original field equations it was invented to be used for. The way out here is to find new variables (or generators) for which no derivatives will appear on the right-hand sides of (2.6) and (2.7), so that the integration is easier to perform (the Lie–Bäcklund symmetries then reduce to Lie point symmetries).

People often speak of "hidden" symmetries of Einstein's equations when they have found an unexpectedly rich set of solutions to these equations. In particular the symmetries on which the generation techniques for stationary axisymmetric vacuum solutions rest have earned this attribute. Because of the very definition of a Lie–Bäcklund symmetry as sketched above, it is clear that any generation method must rest on them; *they* are those hidden symmetries. It is the aim of this paper to show that this is indeed the case, and to indicate how the two difficulties mentioned above have been overcome.

For a detailed treatment of Lie–Bäcklund symmetries and their applications we refer the reader to textbooks on this subject.[1–3]

III. A SYMMETRY OF THE VACUUM EQUATIONS WITH ONE KILLING VECTOR

Finding symmetries of the vacuum field equations $R_{ab} = 0$ means finding solutions f_{ab} of the symmetry condition

$$f^n{}_{a;bn} + f^n{}_{b;an} - f_{ab}{}^{;n}{}_{;n} - f^n{}_{n;ab} = 0\,. \tag{3.1}$$

As pointed out in Sec. II, it is very difficult to find solutions to this equation. In the general case, only the trivial solution

$$f_{ab} = \underset{\xi}{\pounds} g_{ab} + Mg_{ab}\,, \quad M = \text{const}\,, \tag{3.2}$$

is known. It corresponds to arbitrary coordinate transformations

$$\tilde{x}^i = f^i(x^n,\epsilon)\,, \quad \xi^i = \left.\frac{\partial f^i}{\partial \epsilon}\right|_{\epsilon=0} \tag{3.3}$$

and to the multiplication of the metric by a constant factor.

If, however, one assumes the existence of a non-null Killing vector ξ^a ($\xi_a\xi^a = \lambda \neq 0$), then a Lie–Bäcklund symmetry can be constructed that leaves the system

$$R_{ab} = 0\,, \quad \underset{\xi}{\pounds} g_{ab} = \xi_{a;b} + \xi_{b;a} = 0 \tag{3.4}$$

of field equations and Killing equations invariant. This symmetry was found by Kramer and Neugebauer[4] and rediscovered by Geroch.[5] In the notation of Geroch, it rests on the existence [guaranteed by (3.4)] of a scalar ω (the twist potential) and two vectors α_a and β_a, which can be determined from the system of partial differential equations

$$\omega_{,a} = \epsilon_{abcd}\xi^{d;c}\xi^b\,, \tag{3.5}$$

$$\alpha_{b,a} - \alpha_{a,b} = \epsilon_{abcd}\xi^{d;c}\,, \tag{3.6}$$

$$\beta_{b,a} - \beta_{a,b} = 4\lambda\xi_{b;a} + 2\omega\epsilon_{abcd}\xi^{d;c}\,. \tag{3.7}$$

Since these equations determine α_a and β_a only up to gradients, we can impose the gauge conditions

$$\xi^a\alpha_a = \omega\,, \quad \xi^a\beta_a = \omega^2 + \lambda^2\,, \quad \lambda \equiv \xi_a\xi^a \tag{3.8}$$

(which slightly differ from Geroch's conditions). In terms of these quantities the generator of the Lie–Bäcklund symmetry is then given by

$$\begin{aligned} Xg_{ab} &= f_{ab} = -2\omega g_{ab} + 2(\xi_a\alpha_b + \xi_b\alpha_a)\,, \\ X\xi^a &= 0\,, \quad X\xi_a = 2\lambda\alpha_a\,. \end{aligned} \tag{3.9}$$

One can easily check that f_{ab} satisfies (3.1) so that $XR_{ab} = 0$ holds, and that also $\underset{\xi}{\pounds} g_{ab} = 0$ is invariant under X. Moreover one can derive from (3.9) and (3.5)–(3.8) how X acts on λ, ω, α_a, and β_a. The result is

$$X\lambda = 2\omega\lambda\,, \quad X\omega = \omega^2 - \lambda^2\,, \tag{3.10}$$

$$X\alpha_a = 2\omega\alpha_a - \beta_a\,, \quad X\beta_a = 2(\omega^2 + \lambda^2)\alpha_a\,. \tag{3.11}$$

In particular, Eq. (3.9) shows the (expected) complicated dependence of the generator f_{ab} on the metric quantities and the Killing vector field.

To construct the finite symmetry transformations, the general advice was to introduce variables such that in terms of these variables the generator X does not depend on their derivatives. As Eq. (3.10) shows, the variables λ and ω are of this type, i.e., no derivatives appear on the right-hand side (and no new variables either). To take them means to take the Ernst potential

$$\mathscr{E} = \lambda + i\omega \tag{3.12}$$

as variable. Also α_a and β_a may fit into this framework, but an even better choice is the tensor h_{ab} defined by

$$h_{ab} = g_{ab} - \xi_a\xi_b/\lambda\,, \tag{3.13}$$

which, because of (3.9), obeys

$$Xh_{ab} = -2\omega h_{ab}\,. \tag{3.14}$$

It is best to start with (3.10) and (3.14) in order to determine the finite symmetry transformations. The final result is

$$\tilde{\lambda} = \lambda / [1 - 2\epsilon\omega + \epsilon^2(\lambda^2 + \omega^2)] , \tag{3.15}$$

$$\tilde{\omega} = \tilde{\lambda} [\omega - \epsilon(\lambda^2 + \omega^2)] / \lambda , \tag{3.16}$$

$$\tilde{g}_{ab} = \lambda h_{ab} / \tilde{\lambda} + \tilde{\xi}_a \tilde{\xi}_b / \tilde{\lambda} , \tag{3.17}$$

$$\tilde{\xi}_a = \xi_a + \tilde{\lambda}(2\epsilon\alpha_a - \epsilon^2 \beta_a) .$$

It says how to construct the "new" metric $\tilde{g}_{ab}$ if the "old" metric g_{ab} (and λ, ω, α_a, and β_a constructed from it) are known. The Killing vector components ξ^a remain unchanged.

When inspecting all the formulas given above, one immediately sees that most of them are valid only for $\lambda \neq 0$. But it has been noticed already by Geroch[5] that the final transformation (3.17) of the metric is valid also for null Killing vectors with $\xi_a \xi^a = \lambda = 0$. Here we want to sketch a different approach, which runs as follows.

It is known that for a null Killing vector not only λ, but—for vacuum solutions—also the twist ω vanishes. Moreover it follows from the definitions (3.6) and (3.7) of α_a and β_a that both are parallel to ξ_a. So in generalizing (3.9) to the null Killing vector case we may suspect that the symmetry generator has the form

$$X g_{ab} = N \xi_a \xi_b , \quad \xi_a \xi^a = 0 , \quad \xi_{a;b} + \xi_{b;a} = 0 , \tag{3.18}$$

with some (unknown) function N. When starting with this ansatz, it turns out that the symmetry condition (3.1) leads to

$$N_{,a} \xi^a = 0 , \tag{3.19}$$

$$N_{,r}{}^{;r} \xi_i \xi_k + 4N \xi^m \xi^r R_{mirk} - 2N^{,r}(\xi_{r;k}\xi_i + \xi_{r;i}\xi_k) = 0 \tag{3.20}$$

(also the second set is in fact only one equation, all terms being proportional to $\xi_i \xi_k$). Since these two equations are compatible with $XN = 0$, $X\xi_a = 0$, $X\xi^m \xi^r R_{mirk} = 0$, there exist functions N which do *not* depend on those components of the metric that change under the symmetry operation. That is, (3.18) generates a simple Lie-point transformation, the finite transformation being given by

$$\tilde{g}_{ab} = g_{ab} + \epsilon N(x^i) \xi_a \xi_b , \tag{3.21}$$

where $N(x^i)$ is a solution of (3.19) and (3.20). These transformations can be used to generate the *pp* waves from flat space-time and all other vacuum solutions (with a null Killing vector) from a special type D metric.[6]

IV. A SYMMETRY OF THE PERFECT FLUID EQUATIONS WITH ONE KILLING VECTOR

If one analyzes the conditions for the existence of the functions ω and α_a defined by Eqs. (3.5) and (3.6), i.e., if one checks when the right-hand sides of these equations satisfy the necessary integrability conditions, then one arrives at

$$\epsilon^{an}{}_{pq} \omega_{,a;n} = -2\xi^m (\xi_p R_{mq} - \xi_q R_{mp}) = 0 , \tag{4.1}$$

$$\epsilon_{abn}{}^m \alpha_{a;b;m} = 4\xi^a R_{an} = 0 . \tag{4.2}$$

In deriving these equations use has been made of the identity

$$\xi_{a;b;c} = R^i{}_{cba} \xi_i , \tag{4.3}$$

which is valid for Killing vectors ξ_a. If (4.2) holds, β_a also can be introduced.

It is evident from the integrability conditions (4.1) and (4.2) that functions ω, α_a, and β_a exist also in the perfect fluid case if only

$$R_{ab} \xi^b = 0 \tag{4.4}$$

holds. Since the existence of these functions was crucial for the construction of the Lie–Bäcklund symmetry (3.9), one may suspect that a symmetry will exist in the perfect fluid case, too. Indeed it turns out that if only $R_{ab}\xi^a = 0$ holds, then also $XR_{ab} = 0$ follows with X given by (3.9), and the generator X can be extended to the variables u_a, μ, and p so that $X(T_{ab} - Tg_{ab}/2)$ is satisfied and both sides of the perfect fluid equations (1.1) are invariant under X. In detail this runs as follows.

The condition $R_{ab}\xi^a = 0$ can be satisfied in two distinct cases: the Killing vector ξ^a has either to be parallel or orthogonal to the fluid's four-velocity u^a.

In the first case we have

$$\xi_a = u_a \sqrt{-\lambda} , \quad \mu + 3p = 0 ,$$
$$R_{ab} = \tfrac{2}{3}\mu(g_{ab} + u_a u_b) = \tfrac{2}{3}\mu h_{ab} , \tag{4.5}$$

and from the symmetry condition $X\mu h_{ab} = 0$ and the structure of X given in Sec. III we obtain

$$X\mu = 2\omega\mu , \quad Xu^a = -\omega u^a ,$$
$$Xu_a = -\omega u_a - 2\alpha_a \sqrt{-\lambda} . \tag{4.6}$$

In the second case we have

$$\xi_a u^a = 0 , \quad \mu = p , \quad R_{ab} = 2\mu u_a u_b , \tag{4.7}$$

and the symmetry condition $X\mu u_a u_b = 0$ yields

$$X\mu = \omega , \quad Xu_a = -\omega u_a , \quad Xu^a = \omega u^a - 2\xi^a (u^n \alpha_n) . \tag{4.8}$$

The finite symmetry transformations are given by (3.15)–(3.17) for the metric and the two Killing vector invariants λ and ω, and —upon integrating (4.6) and (4.8)— by

$$\tilde{\mu} = \mu / F , \quad F = 1 - 2\epsilon\omega + \epsilon^2(\lambda^2 + \omega^2) , \tag{4.9}$$

and

$$\tilde{u}^a = u^a F^{1/2} , \quad \text{for } \mu + 3p = 0 ,$$
$$\tilde{u}_a = u_a F^{1/2} , \quad \text{for } \mu = p . \tag{4.10}$$

They associate a one-parameter family of perfect fluid solutions to any given solution (with $\mu + 3p = 0$ or $\mu = p$), each of them admitting the Killing vector ξ^a. Since R_{ab} and $T_{ab} - Tg_{ab}/2$ are separately invariant under X, it is not possible to generate perfect fluid solutions from the vacuum by means of this specific Lie–Bäcklund symmetry.

V. SYMMETRIES OF THE VACUUM EQUATIONS WITH TWO COMMUTING KILLING VECTORS

If there are two Killing vectors, then it is obvious that a symmetry of the kind discussed in Sec. III is connected with both of them. Moreover it turns out that an infinite number of Lie–Bäcklund symmetries can be found (the Geroch

group) if these two Killing vectors commute and if there are two-spaces orthogonal to the group orbits of these Killing vectors.[7] This infinite-dimensional Lie group is the basis of the many generating techniques developed during the last 20 years.

We do not want, and are not able, to give a detailed account of how the many thousands of pages written on this subject can be translated into the language of Lie–Bäcklund symmetries. Any transformation that generates (new) solutions from given solutions is a Lie–Bäcklund symmetry, since it maps solutions into solutions, although it need not be a member of a continuous group in every case. So such a translation is always possible. But it may prove difficult to find the underlying group explicitly. The two typical questions one should always try to answer when discussing a special generating technique are (1) do the symmetries form a continuous group, and which is the generator X, and (2) how did the author manage to write this symmetry as a Lie point symmetry? A third question (which we have not pursued in this paper so far) is which are the similarity solutions of a given symmetry, i.e., which are the fixed points in the space of solutions defined by $Xg_{ab} = 0$.

VI. SYMMETRIES OF THE PERFECT FLUID EQUATIONS (FOR $\mu = p$) WITH TWO KILLING VECTORS

It is clear from the results of Sec. IV that all the Lie–Bäcklund symmetries and all the generation techniques that rest on the existence of two Killing vectors exist and can be used also in the perfect fluid case if only (4.4) holds for both Killing vectors. For two Killing vectors these conditions can be satisfied only if both are orthogonal to the four-velocity, i.e., for

$$u_a \xi^a = u_a \eta^a = 0, \quad \mu = p, \quad R_{ab} = 2\mu u_a u_b \,. \tag{6.1}$$

As u_a is timelike, the Killing vectors have to be spacelike.

Thus one can pursue the same program as in the vacuum case, repeating everything for the perfect fluid case (the first to do this was Belinski[8]). But in a certain sense this is unnecessary, since all these perfect fluid solutions can be generated from the vacuum by means of a Lie–Bäcklund symmetry. The existence of this additional symmetry is mainly due to the fact[8] that because of the Bianchi identities the vector $s_a = u_a\sqrt{2\mu}$ has to obey

$$s^a{}_{;a} = 0, \quad s^a(s_{a,b} - s_{b,a}) = 0 \,, \tag{6.2}$$

and that in two dimensions (orthogonal to ξ^a and η^a) the second part of these equations implies that s_a is a gradient. The field equations

$$R_{ab} = s_{,a}s_{,b}, \quad s^a{}_{;a} = 0 \tag{6.3}$$

then admit the symmetry with generator

$$Xg_{ab} = \Omega[g_{ab} - (1/W^2)\{\eta_a\eta_b\xi^n\xi_n + \xi_a\xi_b\eta^n\eta_n - (\eta_a\xi_b + \xi_a\eta_b)\xi^n\eta_n\}] \,, \tag{6.4}$$

$$Xs_a = \tau_{,a} \,, \tag{6.5}$$

where the functions τ and Ω have to satisfy

$$\tau^a{}_{;a} = 0 \,, \tag{6.6}$$

$$\Omega_{,a} = (4W/W_{,b}W^{,b})(\tau^{,n}W_{,n}s_{,a} + s^nW_{,n}\tau_{,a} - \tau^{,n}s_{,n}W_{,a}) \tag{6.7}$$

and the invariant W is defined by

$$W^2 = \xi^n\xi_n\eta^a\eta_a - (\xi^n\eta_n)^2 \,. \tag{6.8}$$

This symmetry is in fact a Lie point symmetry, and it can easily be exponentiated. The finite transformation is best given in coordinates adapted to the Killing vectors $\xi = \partial_x$ and $\eta = \partial_y$, i.e., in terms of the line element

$$ds^2 = e^M(dz^2 - dt^2) + W[e^{\Psi}(dx + A\,dy)^2 + e^{-\Psi}dy^2] \,, \tag{6.9}$$

cf. Ref. 6. All metric functions except M are invariant, and the finite symmetry transformation is

$$\tilde{M} = M + \hat{\Omega}, \quad \tilde{s} = s + \epsilon\tau \,, \tag{6.10}$$

where τ satisfies (6.6) and $\hat{\Omega}$ is to be determined from

$$\hat{\Omega}_{,a} = \frac{2W}{W_{,b}W^{,b}}[2\epsilon(\tau^{,n}W_{,n}s_{,a} + s^nW_{,n}\tau_{,a} - \tau^{,n}s_{,n}W_{,a}) + \epsilon^2(\tau^{,n}W_{,n}\tau_{,a} - \tau^{,n}\tau_{,n}W_{,a})] \,. \tag{6.11}$$

(The existence of the possibility of generating the vacuum from the perfect fluid by choosing $\epsilon\tau = -s$ was first noticed by Tabensky and Taub[9] and Wainwright *et al.*[10])

VII. CONCLUDING REMARKS

It is clear from the examples given above that the Lie–Bäcklund symmetries provide a framework to understand and to classify other classes of solutions and generating techniques, too, e.g., for Einstein–Maxwell fields. Here we want to give an example of a different character.

If one sets out to linearize, e.g., the Einstein–Maxwell (or vacuum) equations on a given background (g_{ab}, F_{ab}), then one usually starts with an ansatz

$$\tilde{g}_{ab} = g_{ab} + \epsilon f_{ab}, \quad \tilde{F}_{ab} = F_{ab} + \epsilon\varphi_{ab} \,, \tag{7.1}$$

where the parameter ϵ indicates that f_{ab} and φ_{ab} are small and that products of these terms are to be neglected. Obviously the linearized field equations then coincide with the conditions that f_{ab} and φ_{ab} are components of a symmetry generator X, the difference being only that in the context of symmetries the fields g_{ab} and F_{ab} are *arbitrary* whereas when linearizing one usually starts from a *fixed* background.

One sometimes is interested to find out whether (and for which background fields) the solutions of the linearized field equations are also solutions to the exact field equations, i.e., whether $\tilde{g}_{ab}$ and $\tilde{F}_{ab}$ are exact solutions. From the symmetries' point of view the answer is simple. In the notation of Sec. II, the question is when

$$\tilde{V}_A = V_A + \epsilon W_A, \quad W_A \equiv XV_A \tag{7.2}$$

is the solution to Eq. (2.7), and the answer is that (7.2) is the solution only if W_A does not depend on ϵ. That is, the solutions f_{ab}, φ_{ab} are solutions also of the exact equations with background g_{ab}, F_{ab} if

$$XXg_{ab} = Xf_{ab} = 0 \,, \quad XXF_{ab} = X\varphi_{ab} = 0 \tag{7.3}$$

holds. An example where this problem is discussed (without

referring to symmetries) can be found in Xanthopoulos.[11]

The result that so many solutions can be understood, and could have been found, in terms of Lie–Bäcklund symmetries makes one hope that more new solutions can be obtained by exploiting this method.

ACKNOWLEDGMENT

I thank Dr. D. Kramer for valuable discussions on Bäcklund transformations.

[1]P. J. Olver, *Applications of Lie Groups to Differential Equations* (*Graduate Texts in Mathematics*) (Springer, New York, 1986).
[2]N. Ibragimov, *Transformation Groups Applied to Mathematical Physics* (Reidel, Boston, 1985).
[3]H. Stephani (to be published).
[4]D. Kramer and G. Neugebauer, Commun. Math. Phys. **10**, 132 (1968).
[5]R. Geroch, J. Math. Phys. **12**, 918 (1971).
[6]D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[7]R. Geroch, J. Math. Phys. **13**, 394 (1972).
[8]V. A. Belinski, Zh. Eksp. Teor. Fiz **77**, 1239 (1979) [Sov. Phys. JETP **50**, 623 (1979)].
[9]R. Tabensky and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).
[10]J. Wainwright, W. C. W. Ince, and B. J. Marshman, Gen. Relativ. Gravit. **10**, 259 (1979).
[11]B. C. Xanthopoulos, Class. Quant. Grav. **3**, 158 (1986).

Spinor fields and the GL(4,R) gauge structure in the nonsymmetric theory of gravitation

J. W. Moffat
Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada

(Received 19 May 1987; accepted for publication 27 January 1988)

The spinor structure associated with the local gauge group GL(4,R) of the nonsymmetric gravitation theory (NGT) is based on a spinor wave equation constructed from a *vierbein*, a GL(4,R) spin connection, and the infinite-dimensional irreducible representations of the universal covering group $\mathscr{SL}$ (4,R) of the noncompact group SL(4,R). The multiplicity-free irreducible representations of $\mathscr{SL}$ (4,R) correspond to bivalued spinorial representations of SL(4,R) that contain an infinite number of half-odd integer spin particles. By adjoining the translations T_4, the extended group $\mathscr{A} = T_4 \times \mathrm{GL}(4,R)$ replaces the Poincaré group $\mathscr{P}$. The properties of the mass spectrum are obtained from an infinite-component wave equation and the physical spinor field consists of an infinite sum of finite, nonunitary representations of the Lorentz group.

I. INTRODUCTION

The classical nonsymmetric gravitation theory (NGT) has been developed in a series of papers[1-5] in which it has been shown that the macroscopic predictions of the theory are consistent with observational data. The local gauge group of general relativity (GR), associated with the four-dimensional fiber bundle, is the homogeneous Lorentz group SO(3,1). In NGT, the gauge group SO(3,1) is extended to a larger gauge group U(3,1,Ω), based on the hyperbolic complex ring of numbers Ω, which is isomorphic to GL(4,R). Moreover, we find that a *complete* dynamical solution of space-time can be achieved within the classical framework of NGT. It is mainly for these reasons that an intensive study of the consequences of NGT has been pursued. In the following, we shall investigate the role played by spinor fields in NGT.

In GR the introduction of a pseudo-Riemannian metric reduces the structure group of the frame bundle from GL(4,R) to the group of transformations SO(3,1) that preserves the metric. In NGT, the underlying mathematical structure of the theory can be formulated in a real eight-dimensional fiber bundle space which has a hyperbolic complex structure imposed upon it, generated by a complex operator E satisfying $E^2 = +1$.[6-9] The hyperbolic complex number system forms a *ring* and not a field as with ordinary complex numbers. The subgroup of GL(8,R) of the eight-dimensional fiber bundle space that preserves E is isomorphic to GL(4,R)$\times$GL(4,R); this group reduces to GL(4,R) when a metric is introduced. Thus the group GL(8,R) reduces to GL(4,R)$\times$GL(4,R) when it is required that $\nabla E = 0$, where ∇ denotes the covariant derivative with respect to the connection in the fiber bundle. The introduction of a metric g in the fiber bundle, which satisfies $\nabla g = 0$, then reduces GL(4,R)$\times$GL(4,R) to GL(4,R). This latter group preserves the nonsymmetric fundamental tensor $g_{\mu\nu}$ ($\mu,\nu = 0,1,2,3$), which decomposes according to

$$g_{\mu\nu} = g_{(\mu\nu)} + g_{[\mu\nu]}, \tag{1.1}$$

where

$$g_{(\mu\nu)} = \tfrac{1}{2}(g_{\mu\nu} + g_{\nu\mu}), \quad g_{[\mu\nu]} = \tfrac{1}{2}(g_{\mu\nu} - g_{\nu\mu}). \tag{1.2}$$

In (1.1), $g_{[\mu\nu]}$ is a pure imaginary skew tensor $g_{[\mu\nu]} = \epsilon a_{\mu\nu}$, where $a_{[\mu\nu]}$ is a real quantity and $\epsilon^2 = +1$. The tensor $g_{\mu\nu}$ is hyperbolic complex Hermitian $g_{\mu\nu} = \bar{g}_{\nu\mu}$. The inverse of $g_{\mu\nu}$ is defined by

$$g^{\mu\nu}g_{\sigma\nu} = g^{\nu\mu}g_{\nu\sigma} = \delta^{\mu}_{\sigma}. \tag{1.3}$$

For a sesquilinear, hyperbolic complex-valued $g_{\mu\nu}$, there exists a *local* GL(4,R) gauge symmetry, which corresponds to "orthogonal" (i.e., $g_{\mu\nu}$ preserving) rotations of the generalized frames. It should be stressed that, in NGT, the GL(4,R) appears as an *internal* symmetry group, which is not *a priori* related to the linear subgroup GL(4,R) of the group of general coordinate transformations $\mathscr{C}$ over the four-dimensional space-time manifold M_4 (group of diffeomorphisms) under which NGT is invariant.

It is only in NGT that the group GL(4,R) appears correctly as the local gauge group of the fiber bundle in the space of anholonomic coordinates. This is in complete analogy with the existence of the Lorentz group SO(3,1) as the local gauge group of GR.

We can define hyperbolic complex *vierbeins* by means of

$$e^a_\mu = \mathrm{Re}(e^a_\mu) + \epsilon\,\mathrm{Im}(e^a_\mu), \tag{1.4}$$

where $a = 0,1,2,3$. The sesquilinear form of $g_{\mu\nu}$ is given by

$$g_{\mu\nu} = e^a_\mu \bar{e}^b_\nu \eta_{ab}. \tag{1.5}$$

Here $\eta_{ab} = \mathrm{diag}(1,-1,-1,-1)$ is the Minkowski flat-space metric and $\bar{e}^a_\mu$ is the complex conjugate of the hyperbolic complex *vierbein* e^a_μ. This formalism has been extended to an n-dimensional space.[7,10] It has been proved that in the linear approximation to the field equations of NGT, the hyperbolic complex structure is free of ghost poles.[11,12]

II. GEOMETRICAL STRUCTURE OF NGT

The hyperbolic complex *vierbeins* e^a_μ obey

$$e^\mu_c e^b_\mu = \delta^b_c, \quad e^a_\sigma e^\rho_a = \delta^\rho_\sigma \tag{2.1}$$

and satisfy ($e^a_{\mu,\sigma} = \partial_\sigma e^a_\mu$)

$$e^a_{\mu,\sigma} + (\omega_\sigma)^a_c e^c_\mu - W^\rho_{\sigma\mu} e^a_\rho = 0, \tag{2.2}$$

 0022-2488/88/071655-06$02.50

where ω_σ is the spin connection in NGT and $W^\lambda_{\mu\nu}$ is the nonsymmetric affine connection defined by

$$W^\lambda_{\mu\nu} = W^\lambda_{(\mu\nu)} + W^\lambda_{[\mu\nu]}. \tag{2.3}$$

The skew part $W^\lambda_{[\mu\nu]}$ is pure imaginary, $W^\lambda_{[\mu\nu]} = \epsilon L^\lambda_{[\mu\nu]}$ ($L^\lambda_{[\mu\nu]}$ is a real skew-symmetric tensor).

We can solve for W in terms of e and ω:

$$W_{\sigma\lambda\rho} = g_{\delta\rho} W^\delta_{\sigma\lambda} = \eta_{ab}(D_\sigma e^a_\lambda)\tilde{e}^b_\rho, \tag{2.4}$$

where D_σ is the covariant derivative operator defined by

$$D_\sigma e^a_\mu = e^a_{\mu,\sigma} + (\omega_\sigma)^a_c e^c_\mu. \tag{2.5}$$

By differentiating (1.5), we obtain

$$g_{\mu\nu,\sigma} - g_{\rho\nu} W^\rho_{\mu\sigma} - g_{\mu\rho}\tilde{W}^\rho_{\nu\sigma} = 0, \tag{2.6}$$

where we have used the condition that

$$(\omega_\sigma)_{ca} = -(\tilde{\omega}_\sigma)_{ac}. \tag{2.7}$$

Thus the spin connection $(\omega_\sigma)_{ab}$ is skew-Hermitian in the indices a and b. The connection $W^\lambda_{\mu\nu}$ is assumed to be Hermitian: $\tilde{W}^\lambda_{\mu\nu} = W^\lambda_{\nu\mu}$, giving the compatibility condition

$$g_{\mu\nu,\sigma} - g_{\rho\nu} W^\rho_{\mu\sigma} - g_{\mu\rho} W^\rho_{\sigma\nu} = 0. \tag{2.8}$$

A group of isometries is defined by

$$e^a_\sigma = e'^b_\sigma (U)^a_b, \tag{2.9}$$

where U is an element of GL(4,R) that leaves the fundamental form $g_{\mu\nu}$ invariant. Moreover, the connection W will remain invariant under the transformation (2.9) provided we satisfy the non-Abelian transformation

$$(\omega_\sigma)^a_b \to [U\omega_\sigma U^{-1} - (\partial_\sigma U)U^{-1}]^a_b. \tag{2.10}$$

A curvature tensor can be defined by

$$([D_\mu, D_\nu])^a_b = (R_{\mu\nu})^a_b, \tag{2.11}$$

where

$$(R_{\mu\nu})^a_b = (\omega_\nu)^a_{b,\mu} - (\omega_\mu)^a_{b,\nu} + ([\omega_\mu,\omega_\nu])^a_b. \tag{2.12}$$

In terms of the transformation (2.10) with

$$(U^{-1})^a_b = \eta_{bd}(\tilde{U})^d_e \eta^{ea}, \tag{2.13}$$

we have

$$(R_{\mu\nu})^a_b \to U^a_c (R_{\mu\nu})^c_d (U^{-1})^d_b. \tag{2.14}$$

The curvature tensor in holonomic coordinates is given by

$$R^\lambda{}_{\sigma\mu\nu} = (R_{\mu\nu})^a_b e^\lambda_a e^b_\sigma \tag{2.15}$$

and the scalar curvature takes the form

$$R = e^{\mu a}\tilde{e}^{\nu b}(R_{\mu\nu})_{ab}. \tag{2.16}$$

The action for the theory in the absence of matter is

$$S_R = -\frac{1}{16\pi G}\int d^4x |e(x)| R(x), \tag{2.17}$$

where $|e| = (e\tilde{e})^{1/2}$ with $e = \det(e^a_\mu)$ and G is the Newtonian constant of gravitation. A variation of the action (2.17) with respect to ω and e yields the field equations

$$[|e|(e^{\mu a}\tilde{e}^{\nu b} - e^{\nu a}\tilde{e}^{\mu b})]_{,\nu} + |e|[(\omega_\mu)^b_c(e^{\mu a}\tilde{e}^{\nu c} - e^{\nu a}\tilde{e}^{\mu c}) + (\omega_\mu)^a_c(e^{\nu c}\tilde{e}^{\mu b} - e^{\mu c}\tilde{e}^{\nu b})] = 0, \tag{2.18}$$

$$R_{\mu a} = 0, \tag{2.19}$$

where

$$R_{\mu a} = \tilde{e}^{\nu b}(R_{\mu\nu})_{ab}. \tag{2.20}$$

By contracting (2.18) over the indices a and b, we obtain

$$[|e|(e^{\mu a}\tilde{e}^\nu_a - e^{\nu a}\tilde{e}^\mu_a)]_{,\nu} = 0, \tag{2.21}$$

which is equivalent to

$$((-g)^{1/2} g^{[\mu\nu]})_{,\nu} = 0. \tag{2.22}$$

III. SPINORS IN GL(4,R) SPACE

In GR a spinor transforms according to the law

$$\psi \to \psi + \delta_\xi \psi, \quad \delta_\xi \psi = -\xi^\sigma \partial_\sigma \psi, \tag{3.1}$$

where $\xi^\sigma(x)$ is defined by the infinitesimal transformations

$$x^\sigma \to x^\sigma + \xi^\sigma(x). \tag{3.2}$$

The ψ belongs to a finite-component nonunitary spinor representation of the universal covering group $\mathscr{SO}(3,1) \simeq \mathrm{SL}(2,\mathrm{C})$ of the Lorentz group. Under infinitesimal Lorentz transformations with the coefficients $a_{ab}(x) = -a_{ba}(x)$, a Dirac spinor transforms under the infinitesimal Lorentz transformations according to

$$\delta_L \psi = -\tfrac{1}{2} a_{ab}\sigma^{[ab]}\psi, \tag{3.3}$$

where the six generators $\sigma^{[ab]}$ can be constructed from the Dirac matrices $\gamma^a = V^a_\mu \gamma^\mu$ using the (real) *vierbein* V^a_μ defined at every point $x = x'$ by

$$V^a_\mu = \left(\frac{\partial X^a(x)}{\partial x^\mu}\right)_{x=x'}. \tag{3.4}$$

The X^a correspond to a set of anholonomic coordinates that are locally inertial at $x = x'$. The Dirac matrices satisfy the rules

$$\{\gamma^a,\gamma^b\} = 2\eta^{ab}, \quad \sigma^{[ab]} = -\tfrac{1}{4}[\gamma^a,\gamma^b]. \tag{3.5}$$

In NGT, the covariant derivative operating on a spinor ψ is defined by

$$\mathscr{D}_\mu \psi(x) = [\partial_\mu + (\omega_\mu(x))_{ab}\Sigma^{ab}]\psi(x), \tag{3.6}$$

where $\psi(x)$ transforms under the group of transformations of GL(4,R) as an *infinite*-component spinor. The GL(4,R) representations whose O(4) reduction yields multivalued representations become single-valued for the covering groups $\mathscr{GL}(4,R)$ and $\mathscr{O}(4)$. The double covering of the group $\mathscr{GL}(4,R)$, taken as a matrix group, exists only for *infinite matrices*.

The generalized Dirac scalar operator is defined by

$$\mathscr{D}(x)\psi(x) = Y^\sigma \mathscr{D}_\sigma(x)\psi(x) = g^{\sigma\nu} e_{\nu a} Y^a \mathscr{D}_\sigma(x)\psi(x). \tag{3.7}$$

In (3.6) we have

$$(\omega_\sigma(x))_{ab} = (\omega_\sigma(x))_{(ab)} + (\omega_\sigma(x))_{[ab]}, \quad \Sigma^{ab} = \Sigma^{(ab)} + \Sigma^{[ab]}. \tag{3.8}$$

The Y^a are algebraic operators determined by an infinite irreducible representation of the algebra GL(4,R) and the Σ^{ab} are the 16 generators of the noncompact group GL(4,R).

For a generic infinite-component spinor Ψ, which can represent a spinor ψ or a Rarita–Schwinger spinor carrying a

"world" suffix ψ_μ, the infinite-component wave equation, in NGT, takes the form[10]

$$(i\mathscr{D}(x)-\kappa)\Psi(x)=0, \tag{3.9}$$

where κ is a constant. The action for NGT now takes the form

$$S=S_R+S_\Psi, \tag{3.10}$$

where S_Ψ is given by

$$S_\Psi=\int d^4x\,\bar{\Psi}(x)(i\mathscr{D}(x)-\kappa)\Psi(x). \tag{3.11}$$

IV. PROPERTIES OF THE GROUP GL(4,R) AND THE ALGEBRA SL(4,R)

The generators Σ_{ab} of the noncompact group GL(4,R) can be split into the one-parameter group of dilations and the SL(4,R) group, with the latter being the group of volume preserving transformations in a nonsimply connected parameter space. We picture the group elements of GL(4,R) as being described by 4×4 matrices. Then the subgroup of dilations consists of constant, diagonal matrices that commute with those of the semisimple noncompact Lie group GL(4,R). The maximal compact subgroup of GL(4,R) is SO(4) and the universal covering group of SL(4,R) is $\mathscr{SL}(4,R)$, which has the same Lie algebra as SL(4,R). Here $\mathscr{SL}(4,R)$ is simply connected and contains the maximal compact subgroup $\mathscr{SO}(4)$, which is isomorphic to SU(2)$\times$SU(2). The factor group of $\mathscr{SL}(4,R)$ with respect to Z_2 is isomorphic to SL(4,R):

$$\mathscr{SL}(4,R)/Z_2\simeq \mathrm{SL}(4,R). \tag{4.1}$$

Moreover, $\mathscr{SL}(4,R)$ and $\mathscr{SO}(4)$ are the double-covering groups of SL(4,R) and SO(4), respectively. The complete center Z_4 of $\mathscr{SL}(4,R)$ gives the factoring $\mathscr{SL}(4,R)/Z_4\simeq$SO(3,3). The spinor representations, in NGT, are the infinite-dimensional irreducible representations, which are double-valued representations of SL(4,R). The fact that infinite-dimensional multivalued spinor representations of SL(4,R) exist was proved by Ne'eman.[13] The homogeneous Lorentz group SO(3,1) is a subgroup of SL(4,R) and, consequently, the Lorentz double-covering group $\mathscr{SO}(3,1)\simeq$SL(2,C) is a subgroup of $\mathscr{SL}(4,R)$.

We identify $\Sigma_{[ab]}=M_{[ab]}$ as the six generators $M_{[ab]}$ of the homogeneous Lorentz group formed from the angular momentum operators J_i and the boost operators K_i ($i=1,2,3$). The remaining nine generators of SL(4,R) are the shear tensor $S_{(ab)}$ with $\mathrm{Tr}(S_{(ab)})=0$. The dilation generator S and the nine generators $S_{(ab)}$ together determine the ten generators $\Sigma_{(ab)}$. The commutation relations of the algebra SL(4,R) are[14,15]

$$\begin{aligned}
[M_{[ab]},M_{[cd]}] &= -i(\eta_{ac}M_{[bd]}-\eta_{ad}M_{[bc]}\\
&\quad -\eta_{bc}M_{[ad]}+\eta_{bd}M_{[ac]}),\\
[M_{[ab]},S_{(cd)}] &= -i(\eta_{ac}S_{(bd)}+\eta_{ad}S_{(bc)}\\
&\quad -\eta_{bc}S_{(ad)}-\eta_{bd}S_{(ac)}),\\
[S_{(ab)},S_{(cd)}] &= i(\eta_{ac}M_{[bd]}+\eta_{ad}M_{[bc]}\\
&\quad +\eta_{bc}M_{[ad]}+\eta_{bd}M_{[ac]}).
\end{aligned} \tag{4.2}$$

We can write the commutation relations (4.2) in the more compact form

$$[K_{ab},K_{cd}]=i\eta_{bc}K_{ad}-i\eta_{ad}K_{cb}, \tag{4.3}$$

where $K_{(ab)}=S_{(ab)}$ and $K_{[ab]}=\Sigma_{[ab]}=M_{[ab]}$. From the commutators with the angular momentum generators $\mathrm{J}_i=\frac{1}{2}\epsilon_{ijk}M_{[jk]}$ ($i,j,k=1,2,3$) that form the subalgebra SU(2), we find that the $S_{(ab)}$ separate into $J=2$ associated with the $S_{(jk)}$, $J=1$ associated with the $S_{(0j)}$, and $J=0$ associated with the S_{00}. The generators Σ^{ab} of the algebra GL(4,R) are given by

$$\Sigma^a_b=\tfrac{1}{2}(M^a_b+S^a_b+\tfrac{1}{2}\delta^a_bS) \tag{4.4}$$

and satisfy the commutation relations

$$[\Sigma_{ab},\Sigma_{cd}]=i\eta_{bc}\Sigma_{ad}-i\eta_{ad}\Sigma_{cb}. \tag{4.5}$$

SL(4,R) has the subgroup SL(3,R) that exists in the spatial subspace of Minowski space. The latter group has eight generators formed from the angular momentum operators $M_{[ij]}$ ($i,j=1,2,3$) and the five shear operators $S_{(ij)}$, which transform under the subgroup SO(3) of SL(3,R) as a quadrupole operator. The noncompact subgroup SO(3,3) possesses the compact subgroup SO(3)$\times$SO(3) with the double covering SU(2)$\times$SU(2). The generators of the latter group can be constructed as follows:

$$\begin{aligned}
J^{(1)}_i &= \tfrac{1}{4}\epsilon_{ijk}M_{[jk]}+\tfrac{1}{2}S_{(0i)},\\
J^{(2)}_i &= \tfrac{1}{4}\epsilon_{ijk}M_{[jk]}-\tfrac{1}{2}S_{(0i)},
\end{aligned} \tag{4.6}$$

The Lie algebra SL(4,R) is isomorphic to the algebra SO(3,3). The generators of SO(3,3) can be written in terms of Dirac matrices.[16] Let us introduce the notation

$$\gamma^m=\gamma^1,\gamma^2,\gamma^3,-i\gamma^5,\gamma^0,-iI,\quad m=1,2,3,4,5(\equiv0),6, \tag{4.7}$$

and set

$$Q^{[mn]}=\tfrac{1}{2}\gamma^{[m}\gamma^{n]}. \tag{4.8}$$

The commutation relations for the algebra SO(3,3) read as

$$\begin{aligned}
[Q^{[mn]},Q^{[pq]}] &= h^{np}Q^{[mq]}-h^{mp}Q^{[nq]}\\
&\quad -h^{nq}Q^{[mp]}+h^{mq}Q^{[np]},
\end{aligned} \tag{4.9}$$

where the metric h^{mn} is given by $h^{mn}=\mathrm{diag}(-1,-1,-1,+1,+1,+1)$. The SL(4,$R$) generators K^{ab} that satisfy the commutation relations (4.3) can be identified with the 15 generators $Q^{[mn]}$.

In NGT, the Poincaré algebra $\mathscr{P}$ is replaced by the algebra $\mathscr{A}=T_4\times$GL(4,R), where T_4 are the translations in space-time. However, the algebra $\mathscr{A}$ cannot be gauged directly by the fiber bundle structure in NGT. This situation is analogous to GR, in which the Poincaré algebra $\mathscr{P}$ cannot be obtained directly as a local gauge structure. The algebra $\mathscr{A}$ can be used to obtain kinematical information about masses in NGT since the momentum operator P_μ satisfies $P_\mu P^\mu=m^2$. The algebra of $\mathscr{A}$ is obtained by supplementing the commutation relations (4.2) with

$$\begin{aligned}
&[M_{[ab]},P_c]=i(\eta_{bc}P_a-\eta_{ac}P_b),\\
&[S_{(ab)},P_c]=-i(\eta_{ac}P_b+\eta_{bc}P_a-\tfrac{1}{2}\eta_{ab}P_c),\\
&[P_a,P_b]=0,\quad [S,M_{[ab]}]=[S,S_{(ab)}]=0,\\
&[S,P_a]=-iP_a.
\end{aligned} \tag{4.10}$$

V. UNITARY REPRESENTATIONS OF $\mathscr{SL}(4,R)$

Ne'eman and Šijački[17] have constructed a classification of the multiplicity-free irreducible unitary representations of $SL(4,R)$ using the representations of the maximal compact subgroup $SU(2)\times SU(2)$. Work on this problem has also been done by Kihlberg,[18] Speh,[19] and Friedman and Sorkin.[20] At present there does not exist a complete classification of all the unirreps of $SL(4,R)$. We shall list the classification for the multiplicity-free set obtained by Ne'eman and Šijački[17]:

(i) the principal series,

$\mathscr{D}^{\text{prin}}(e_2;j_1j_2)$: $e_2\in R$,

$j_1=0,\ j_2=0,\quad \{|J_1|+|J_2|\}\equiv 0 \pmod 2$,

$j_1=1,\ j_2=0,\ \{|J_1|+|J_2|\}\equiv 1 \pmod 2$;

(ii) the supplementary series,

$$\mathscr{D}^{\text{sup}}(e_1;j_1,j_2),\quad 0\langle|e_1|\rangle 1,\quad j_1=1,\quad j_2=0; \tag{5.1}$$

(iii) the discrete series,

$\mathscr{D}^{\text{disc}}(j,0)$ and $\mathscr{D}^{\text{disc}}(0,j)$,

$\{j\}=\frac{1}{2},1,\frac{3}{2},2,\ldots\ |J_1|-|J_2|\geqslant j,\quad |J_1|+|J_2|\equiv j \pmod 2$;

(iv) the ladder series,

$\mathscr{D}^{\text{ladd}}(J),\quad \{j\}=0,\frac{1}{2},\quad |J_1|=|J_2|=j+|Z|$.

The j_1,j_2 denote the coordinates of the lowest $\mathscr{SO}(4)$ submultiplets. Also, $e=e_1+ie_2$, $e_1,e_2\in R$, where e denotes the $\mathscr{SL}(4,R)$ representation label.

The second-order Casimir operator for $\mathscr{SL}(4,R)$ is given by

$$C_2=K_{ab}K^{ab}=-4+\tfrac{1}{4}(e_1+ie_2)^2. \tag{5.2}$$

The C_2 for the principal and ladder series is

$$C_2=-4-\tfrac{1}{2}e_2^2, \tag{5.3}$$

while for the discrete and supplementary series, we obtain

$$C_2=-4+\tfrac{1}{4}(j_0-1)^2 \tag{5.4}$$

and

$$C_2=-4+\tfrac{1}{4}e_1^2, \tag{5.5}$$

respectively, where $j_0=\frac{1}{2},1,\frac{3}{2},2,\ldots$.

It has been noted by Ne'eman and Šijački[21] that the $\mathscr{SL}(4,R)$ algebra commutation relations (4.2) are invariant under the automorphism

$$J_i'=J_i,\quad K_i'=iN_i,\quad N_i'=iK_i, \tag{5.6}$$

where $N_i=S_{(0i)}$ and the (J_i,iK_i) make up a new compact algebra $SO(4)'$ and the (J_i,iN_i) make up the algebra $SL(2,C)'$.

VI. DYNAMICAL FIELD EQUATIONS OF NGT

The field equations, in NGT, obtained from a variation of the Lagrangian density, including a matter Lagrangian, take the form in holonomic coordinates,[2]

$$G_{\mu\nu}(W)=8\pi GT_{\mu\nu}, \tag{6.1}$$

$$\mathbf{g}^{[\mu\nu]}{}_{,\nu}=4\pi\mathbf{S}^{\mu}, \tag{6.2}$$

where $G_{\mu\nu}(W)$ is the Einstein tensor, formed from the W connection

$$G_{\mu\nu}(W)=R_{\mu\nu}(W)-\tfrac{1}{2}g_{\mu\nu}R(W), \tag{6.3}$$

$\mathbf{g}^{[\mu\nu]}=(-g)^{1/2}g^{[\mu\nu]}$, and $\mathbf{S}^{\mu}=(-g)^{1/2}S^{\mu}$ is a conserved current density

$$\mathbf{S}^{\mu}{}_{,\mu}=0. \tag{6.4}$$

The field equations (6.1) and (6.2) are invariant under the group $\mathscr{C}$ of nonlinear coordinate transformations (diffeomorphism group) in the base manifold, as well as the Abelian U(1) (or R_+) gauge transformation

$$W_{\mu}'=W_{\mu}+\lambda_{,\mu}, \tag{6.5}$$

where W_{μ} is the torsion vector defined by $W_{\mu}=W^{\lambda}_{[\mu\lambda]}$. This Abelian gauge invariance gives rise to the conservation law (6.4) by virtue of Noether's theorem. The four conservation laws obeyed by the nonsymmetric energy-momentum tensor $T^{\mu\nu}$ follow from the invariance of the Lagrangian density with respect to the group of transformations $\mathscr{C}$.

A possible model for S^{μ} is to interpret it as a fermion number current density

$$S^{\mu}=\sum_i f_i^2 n_i, \tag{6.6}$$

where the f_i are coupling constants for the i species of fermions and n_i is the number density of fermions; f has the dimensions of a length. The coupling constants f and G set two mass scales in NGT, $f^{-1}\geqslant 10^6$ GeV and $G^{-1/2}=1.2\times 10^{19}$ GeV, where the latter mass scale corresponds to the Planck mass. The precise value of the constant f will be determined by gravitational experiments.[5] Thus we can anticipate that, in NGT, the lowest mass scale is at $m=0$, corresponding to the standard hadron and weak interaction mass scale, while there exists an intermediate mass scale at $\geqslant 10^6$ GeV before we reach the Planck mass scale $\simeq 10^{19}$ GeV.

In NGT, we can choose a local inertial frame at a point $x=x'$ such that the *metric tensor* $g_{(\mu\nu)}$ takes its Minkowski values $g_{(\mu\nu)}=\eta_{\mu\nu}$. Accordingly, in the definition of the metric tensor in terms of the (real) *vierbeins* E^a_{μ},

$$g_{(\mu\nu)}=E^a_{\mu}E^b_{\nu}\eta_{ab}, \tag{6.7}$$

we have $E^a_{\mu}=\delta^a_{\mu}$ at the point $x=x'$. The equivalence of inertial and gravitational masses still holds in NGT,[5] although a new composition-dependent force occurs in the post-Newtonian order of approximation that behaves like $1/r^5$, which will cause test particles to fall at different rates in the gravitational field of a spherically symmetric body. In the local inertial frame, the NGT Christoffel symbol defined by

$$\begin{Bmatrix}\lambda\\ \mu\nu\end{Bmatrix}=\frac{1}{2}\gamma^{(\lambda\sigma)}(g_{(\mu\sigma),\nu}+g_{(\sigma\nu),\mu}-g_{(\mu\nu),\sigma}), \tag{6.8}$$

where $\gamma^{(\mu\nu)}$ satisfies $\gamma^{(\mu\nu)}g_{(\sigma\mu)}=\delta^{\nu}_{\sigma}$, will vanish at the space-time point $x=x'$.

VII. THE MASS SPECTRUM AND $\mathscr{SL}(4,R)$

In standard relativistic quantum field theory, the physical particle states are characterized by their three-momentum, spin, and mass and are Wigner's basis states for a unitary irreducible representation of the Poincaré group. *The fields transform as finite and nonunitary representations* of

$GL(4,R)$ when tensorial and of $SO(3,1)$—or its double cover $SL(2,C)$—when spinorial. We require that a Lorentz boost acting on, say, a proton with $J=\frac{1}{2}$ should increase its linear momentum or velocity, but not change its nature. If the proton field were associated with a unitary representation of $SL(2,C)$, then the Lorentz boost would transform the $J=\frac{1}{2}$ state partly into a $J=\frac{3}{2}$ state which is unphysical. Thus if we use unitary representations of $SL(4,R)$, then the Lorentz subgroup $SL(2,C)$ will also be represented unitarily, leading to unphysical results.

Infinite-component fields were first introduced by Majorana,[22,23] who implemented the two irreducible representations of $SL(2,C)$ to introduce an invariant, linear wave equation

$$(iX^{\mu}\partial_{\mu}-\kappa)\psi(x)=0, \tag{7.1}$$

where the operators X^{μ} close on the algebra $\mathcal{SP}(4,R)\simeq\mathcal{SO}(3,2)$ and the ladder representation of $Sp(4,R)$ is unitary and separates into the direct sum of the two Majorana[22,23] representations. Because of Majorana's[22,23] use of unitary infinite-dimensional representations of $SL(2,C)$, Eq. (7.1) leads to unphysical results. The mass spectrum predicted by Eq. (7.1) decreases in mass with increasing spin, contrary to our experience. Infinite-component wave equations were studied by Barut,[24] Fronsdal,[25] and Nambu[26] in the context of the hydrogen atom using representations of $SO(4,2)$ and the correct hydrogen spectrum was obtained by fixing certain parameters in the nonrelativistic limit.[27] Relativistic field theories of infinite-component fields have been studied[25,28,29] and various potential difficulties inherent to a particular form of these theories have been noted: (i) problems with spin and statistics and the associated PCT theorem, and (ii) spacelike solutions with $P_{\mu}P^{\mu}<0$. These difficulties can all be associated with the use of unitary infinite-dimensional representations of the Lorentz group.

A new motivation for studying infinite-component fields and noncompact groups comes from recent developments in string theories,[30,31] in which the boson string or superstring spectrum is generated by infinitely many excited particle states with all spins.

Let us now consider the $GL(4,R)$ invariant wave equation

$$(iY^{\mu}\partial_{\mu}+K(x)-\kappa)\psi(x)=0, \tag{7.2}$$

where the operators Y^{μ} and $\psi(x)$ take their values in the Hilbert space $\mathcal{H}$ of the representations of $\mathcal{SL}(4,R)$ and the operator-valued function $K(x)$ is

$$K(x)=Y^{\mu}(\omega_{\mu}(x))_{ab}\Sigma^{ab}. \tag{7.3}$$

The mass spectrum and its mass scales depend on the form of $K(x)$. The specific dependence of $K(x)$ depends on the solution for $(\omega(x)_{\mu})_{ab}$—the NGT spin connection—which is determined by the field equations. This is analogous to the situation in quantum electrodynamics, in which the minimal coupling assumption

$$\partial_{\mu}\rightarrow\partial_{\mu}+ieA_{\mu}(x) \tag{7.4}$$

is used in the Dirac equation, where the potential $A^{\mu}(x)$ is determined by Maxwell's equations.

The skew part of the spin connection $(\omega_{\mu}(x))_{[ab]}$, in (3.8), is associated with the gravitational field whose coupling constant G possesses the Planck mass scale $G^{-1/2}=1.2\times10^{19}$ GeV, while the symmetric part $(\omega_{\mu}(x))_{(ab)}$ is proportional to the NGT coupling constant f that measures the strength of the coupling of $g_{[\mu\nu]}$ to matter. The latter coupling constant has the mass scale $f^{-1}\geqslant10^{6}$ GeV. We expect that there will be three mass scales in NGT, namely, $m=0$, $m\geqslant10^{6}$ GeV, and $m\simeq10^{19}$ GeV.

In the case of the standard treatment of the Dirac equation, each particle is labeled with a mass m and a spin J in terms of the four-component Dirac spinor. Each particle, in NGT, is labeled by the two spin variables j_1, j_2 associated with the compact subgroup $\mathcal{SO}(4)\simeq SU(2)\times SU(2)$. It has been shown that a Lorentz invariant infinite-component wave equation with a local spinor field that is an infinite *unitary* irreducible representation of $SO(3,1)\simeq SL(2,C)$ has a degenerate mass spectrum.[32–35] A field theory with an infinite-component field that is associated with a unitary representation of the Lorentz group that is invariant under Lorentz transformations violates locality unless the masses in the towers of particles are all degenerate. This theorem will also hold true in our case if we extend the transformation group $SO(3,1)$ to that of $GL(4,R)$. This difficulty can be circumvented in the following way. As noted in Sec. V, the commutation relations (4.2) are invariant under the automorphism (5.6). We can use the "Weyl unitary trick" (5.6) to convert the irredicible representations of $\mathcal{SO}(4)\simeq SU(2)\times SU(2)$ into *nonunitary* finite representations of the Lorentz group $SO(3,1)$.[21] The boost operators K_i become anti-Hermitian operators under the unitary trick analytic continuation. Once the finite nonunitary representations of $SO(3,1)$ are identified correctly, the physical spinor field representations will be described by infinite sums of finite, nonunitary representations of the Lorentz group. Only nonunitary spinor representations in $SO(3,1)$ have a physical meaning because the scalar density $\psi^{\dagger}\psi=E/m$ has the correct boosting property. The anti-Hermitian intrinsic spin pieces cancel in the angular momentum operator, so that the boost acts only orbitally, as in the case of standard finite-component spinors. This avoids the unphysical excitation of a given spin state to other spins and masses.

We are required to decompose the infinite-dimensional representations of $SL(4,R)$ into those of $\mathcal{SO}(4)$, while at the same time we implement the unitary trick to obtain the infinite sums of *physical* nonunitary, finite spinor representations of the Lorentz group. The unitary trick corresponds to the transition from an Euclidean space to the Minkowski space-time with the signature $(1,-1,-1,-1)$. We could, as an alternative, have started with *nonunitary* infinite-dimensional representations of $SL(4,R)$ and proceeded directly to the infinite sum of nonunitary, finite representations of the Lorentz group. However, working with the unitary trick accomplishes the same results and the classification of unitary irreducible representations of $SL(4,R)$ is more complete at this time.

This decomposition scheme avoids all the standard problems associated with infinite-component spinor equations; the infinite-component field theory will be free of any

fundamental difficulties such as spacelike solutions and spin-statistics problems.

Cant and Ne'eman[16] have presented an SL(2,C) invariant, infinite-component wave equation with a spinor that is a unitary irreducible representation of $\mathscr{SL}(4,R)$ given by $\mathscr{D}^{\text{disc}}(j_1, j_2)$, for $j_1 = p_1 + 1$, $j_2 = 0$ or $j_1 = 0$, $j_2 = p_1 + 1$, where $p_1 = -\frac{1}{2},\frac{1}{2},\frac{3}{2},\frac{5}{2},\ldots$, $p_2 = 0$, and $|j_1 - j_2| \geqslant p_1 + 1$. The physical spinor particles will be described by $\psi(x) \simeq \mathscr{D}^{\text{disc}}(\frac{1}{2},0) \oplus \mathscr{D}^{\text{disc}}(0,\frac{1}{2})$.

The Y^μ operators that take values in the Hilbert space $\mathscr{H}$ of the representations of SL(4,R) transform as an SL(2,C) vector, i.e., we demand that Y^μ be a covariant vector only under Lorentz transformations. With this requirement the $\mathscr{SL}(4,R)$ towers of particles that make up the representations describing ψ are not mass degenerate and the field theory underlying this formalism should display locality.

Ne'eman and Šijački[36] have shown that the spin content described by (j_1, j_2) in the SL(4,R) unirreps in (5.1) can be used to provide a shell-model-like description of the baryon and meson resonances for each flavor. The infinite-dimensional representations of the subgroup SL(3,R) have been shown to be associated with a Regge behavior.[37,38] This scheme provides a phenomenological basis for the towers of particle states at the lowest mass scale. The hadron description that emerges from this program complements the dynamical calculations based on quantum chromodynamics.

VIII. CONCLUSIONS

The nonsymmetric gravitation theory possesses the local gauge symmetry GL(4,R) with a spin connection that is constructed from the 16 generators Σ^{ab} of GL(4,R). Multivalued spinor representations of the unimodular group SL(4,R) exist *only* for infinite irreducible representations of SL(4,R) and its covering group $\mathscr{SL}(4,R)$. This represents a fundamental departure from the physics of Einstein's theory of gravitation. The particle spectrum of NGT consists of infinite towers of particles with all spins, whereas GR is consistent with a finite number of particles that are described by nonunitary representations of SL(2,C). This picture of the particle spectrum in NGT, which is forced upon us by the nature of the covering group of $\mathscr{SL}(4,R)$, is similar to the one predicted by string theories. The GL(4,R) symmetry of NGT is broken down to the experimentally observed Lorentz invariance corresponding to SO(3,1), with the associated finite nonunitary particle representations displaying mass splittings. The infinite-component wave equation is Lorentz invariant, with spinors that are described by infinite sums of finite nonunitary representations of $\mathscr{SL}(4,R)$.

ACKNOWLEDGMENTS

I thank Mark Sutherland and Utpal Sarkar for helpful and stimulating conversations.

This work was supported by the National Sciences and Engineering Research Council of Canada.

[1] J. W. Moffat, Phys. Rev. D **19**, 3554, 3562 (1979).
[2] J. W. Moffat, J. Math. Phys. **21**, 1978 (1980).
[3] J. W. Moffat, Phys. Rev. Lett. **50**, 709 (1983).
[4] J. W. Moffat, Found. Phys. **14**, 1217 (1984).
[5] J. W. Moffat, Phys. Rev. D **15**, 3733 (1987); J. W. Moffat and E. Woolgar, *ibid.* **37**, 918 (1988).
[6] G. Kunstatter, J. W. Moffat, and J. Malzan, J. Math. Phys. **24**, 886 (1983).
[7] J. W. Moffat, J. Math. Phys. **25**, 347 (1984).
[8] R. B. Mann, Class. Quantum Grav. **1**, 561 (1984).
[9] W. K. Clifford, Proc. London Math. Soc. **4**, 381 (1873); Am. J. Math. **1**, 350 (1878); A. Crumeyrolle, Riv. Mat. Univ. Parma **8**, (2) 27 (1967); Z.-Z. Zhong, J. Math. Phys. **26**, 404 (1985); N. Salingaros, *ibid.* **22**, 2096 (1981).
[10] J. W. Moffat, J. Math. Phys. **26**, 528 (1985). Equations (7), (10), and (13) in this reference should be replaced by Eqs. (2.4), (2.10), and (2.16), respectively, in the present paper.
[11] R. B. Mann and J. W. Moffat, Phys. Rev. D **26**, 1858 (1982).
[12] P. F. Kelly and R. B. Mann, Class. Quantum Grav. **3**, 705 (1986).
[13] Y. Ne'eman, Ann. Inst. H. Poincaré A **28**, 369 (1978).
[14] Y. Ne'eman and Dj. Šijački, Ann. Phys. (NY) **120**, 292 (1979); Proc. Natl. Acad. Sci. **76**, 561 (1979).
[15] Y. Ne'eman, Found. Phys. **13**, 467 (1983).
[16] A. Cant and Y. Ne'eman, J. Math. Phys. **26**, 3180 (1985).
[17] Dj. Šijački and Y. Ne'eman, J. Math. Phys. **26**, 2457 (1985).
[18] A. Kihlberg, Ark. Fys. **32**, 241 (1966).
[19] B. Speh, Mat. Ann. **258**, 113 (1981).
[20] J. L. Friedman and R. D. Sorkin, J. Math. Phys. **21**, 1269 (1980).
[21] Y. Ne'eman and Dj. Šijački, Phys. Lett. B **157**, 275 (1985).
[22] E. Majorana, Nuovo Cimento **9**, 335 (1932).
[23] D. T. Stoyanov and I. T. Todorov, J. Math. Phys. **9**, 2146 (1968).
[24] A. O. Barut, *Springer Tracts in Modern Physics, Ergebnisse der Exakten Naturwiss,* Vol. 50 (Springer, Berlin, 1969), pp. 1–28.
[25] C. Fronsdal, Phys. Rev. **156**, 1653, 1665 (1967).
[26] Y. Nambu, Phys. Rev. **160**, 1171 (1967).
[27] B. G. Wybourne, *Classical Groups for Physicists* (Wiley–Interscience, New York, 1974).
[28] G. Feldman and P. T. Matthews, Ann. Phys. (NY) **40**, 19 (1966); Phys. Rev. **151**, 1176 (1966); **154**, 1241 (1967).
[29] E. Abers, L. T. Grodsky, and R. E. Norton, Phys. Rev. **159**, 1222 (1967).
[30] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U.P. Cambridge, 1987), Vols. I and II.
[31] For a review, see J. W. Moffat, Can. J. Phys. **64**, 561 (1986).
[32] H. D. J. Abarbanel and Y. Frishman, Phys. Rev. **171**, 1442 (1968).
[33] I. T. Grodsky and R. F. Streater, Phys. Rev. Lett. **20**, 695 (1968).
[34] A. I. Oksak and I. T. Todorov, Phys. Rev. D **1**, 3511 (1970).
[35] A. Chodos, Phys. Rev. D **1**, 2937 (1970).
[36] Y. Ne'eman and Dj. Šijački, Phys. Lett. B **157**, 267 (1985).
[37] Dj. Šijački, J. Math. Phys. **16**, 298 (1975).
[38] L. C. Biedenharn, R. Y. Cusson, M. Y. Han, and O. L. Weaver, Phys. Lett. **42B**, 257 (1972).

Gravity as a gauge theory with Cartan connection

Andrew Zardecki
Theoretical Division, MS K723, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 11 December 1987; accepted for publication 9 March 1988)

A gauge formulation of gravity, based on the notion of the connection of Cartan, is given. The Cartan connection involves two principal fiber bundles P and P' with groups G and G', respectively; G' is a subgroup of G and can be regarded as a symmetry group to which G is broken. When the differential form defining the connection in P gives absolute parallelism in P', one speaks about the Cartan connection. General geometric framework is specialized to the case in which G is the de Sitter group SO(4,1) and G' is the Lorentz group SO(3,1). The action of the Yang–Mills type is similar to, but not identical with, the action derived earlier by Townsend. The field equations in P are translated into a system of coupled equations for curvature and torsion in P'. Under contraction of SO(4,1) to the Poincaré group ISO(3,1), and for vanishing torsion, the equations become Yang's equation and Einstein's equation in vacuum. The BRS invariance of the theory, supplemented by Faddeev–Popov and gauge fixing terms, is analyzed in some detail.

I. INTRODUCTION

Early attempts to formulate gravity as a gauge theory emphasized the similarity of the gravitational field and the non-Abelian field of Yang and Mills.[1] Geometrically, the Yang–Mills field potential can be identified with the connection form in a principal fiber bundle $P(M,G)$ with structure group G over space-time M. Since all known gauge formulations of gravity assume that the Lorentz group O(3,1) is contained in G, the bundle P is then the bundle $L(M)$ of frames, and the connection in P is a linear connection in L. In contrast to P, in which only connection form ω is defined, L is also endowed with the canonical form (solder form) θ. That the gravitation, regarded as a theory based on ω and θ, is richer than other gauge theories has on various occasions been stressed by Trautman.[2] A comprehensive discussion of gauge treatment of gravity can be found in the review of Ivanenko and Sardanashvily.[3]

In a different yet related context, gauge invariance is obtained as a consequence of the spontaneous breaking of a larger symmetry.[4] The massless gravitational field emerges then automatically as the corresponding Goldstone field. Typical examples are provided by breaking of the general linear group GL(4) to SO(3,1),[5] the Poincaré group ISO(3,1) to SO(3,1),[6] and the de Sitter groups SO(3,2) and SO(4,1) to SO(3,1).[7] The last two schemes, involving the Poincaré and de Sitter groups, enjoy a special feature, not shared by GL(4). To explain it, denote the initial symmetry group as G and the subgroup to which G is broken as G'. When G is identified with ISO(3,1), SO(3,2), or SO(4,1), and G' is identified with SO(3,1), then $\dim(G/G') = \dim M$, the dimension of space-time. This, however, is one of the ingredients of the soldering condition to M of a fiber bundle $E(M,G/G',G,P)$ with fiber G/G' associated with $P(M,G)$.[8] Corresponding to two structural groups G and G', one obtains two principal fiber bundles P and P'; $P' \subset P$. A connection in P is called a connection of Cartan[8,9] if it satisfies the following condition: the differential form $\tilde{\omega}$ defining the connection in P gives absolute parallelism in P'. [The last phrase simply means that $\dim G' + \dim(G/G')$ vectors form a basis of $T_u(P')$ for every $u \in P'$.] The notion of a fiber bundle with Cartan connection has become of considerable interest in physics as a basis of geometric description of gravity[10] and strong interactions.[11]

We note that the de Sitter gauge theory of gravity was already discussed earlier, starting with the papers of Townsend[12] and MacDowell and Mansouri.[13] From a geometric viewpoint, however, Townsend works with the bundle $P(M,\mathrm{SO}(4,1))$ facing interpretive problems of the form θ, which exists only in $P(M,\mathrm{SO}(3,1))$. Furthermore, Townsend's action, in which the group indices are contracted with the aid of the unit metric in the tangent plane, is not of the Yang–Mills type, as advocated in this paper. MacDowell and Mansouri, on the other hand, construct an action invariant under SO(3,1), and not under SO(4,1).

The purpose of this paper is to set up a general framework to describe gravity in terms of the gauge formalism on a fiber bundle with Cartan connection. In this description, the metric tensor will be regarded as an auxiliary quantity to satisfy general covariance[7,14] and is considered as a certain background without dynamical role. Such an approach, in which only the connection forms are regarded as dynamical fields, allows us to bypass the problems of ghosts and tachyons that appear in metric theories linearized about the Minkowski metric.[15] On the other hand, when the theory is supplemented by the Faddeev–Popov (FP) ghosts and auxiliary gauge fixing fields, it becomes a quantum gauge theory satisfying the unitarity postulate for the S matrix.

The remainder of this paper is organized as follows. In Sec. II, we outline the formalism of Cartan connection. The specification to fiber bundles with the de Sitter structure group is the subject of Sec. III. This is followed, in Sec. IV, by a formulation of the gauge theory based on SO(4,1). Although SO(4,1) yields ISO(3,1) under contraction, the resulting theory is quite different from the existing theories[16] based on gauging of the Poincaré group. Rather, it is reminiscent of the theory of López[17] with a quadratic Lagrangian in curvature and torsion. The field equations of our theory

 0022-2488/88/071661-06$02.50

are obtained from a Lagrangian having the Yang–Mills form. One can contrast this approach with the Yang theory of gravity.[18] In both formulations, the Lagrangian is quadratic in curvature. While Yang's theory uses the curvature of the connection in P', we use the curvature of the Cartan connection. In Sec. V, we introduce the FP fields and discuss the Becchi, Rouet, and Stora (BRS) transformation. In the spirit of the findings of Kugo and Ojima,[19] the BRS invariance is regarded as an underlying principle for constructing a quantum gauge theory.

II. CONNECTIONS OF CARTAN

A convenient starting point of our discussion is the definition of soldering of a fiber bundle $E(M,F,G,P)$ with fiber F associated with the principal fiber bundle $P(M,G)$.[20] In conformity with Ehresmann[21] and Kobayashi,[8] $E(M,F,G,P)$ is soldered to M, if the following conditions are satisfied.

(s.1) G acts on F transitively; F can be identified with the homogeneous space G/G', where G' is the stability group at a point o of F.

(s.2) $\dim F = \dim M$.

(s.3) The structure group G of E can be reduced to G'. (When E is considered as the fiber bundle with the structure group G', it will be denoted as E'; similarly, P' denotes the principal fiber bundle associated to E'.)

(s.4) If $T_x(M)$ is the tangent space of M at x and $T^*(M)$ is the space of all tangent vectors to F_x, then for x running through M one can identify $T_x(M)$ and $T^*(M)$ by an isomorphism.

Let $\mathcal{G}$ and $\mathcal{G}'$ be the Lie algebras of G and G', respectively. A homogeneous space $F = G/G'$ is called weakly reductive, if there is a vector subspace $\mathcal{F}$ satisfying the conditions

$$\mathcal{G} = \mathcal{G}' + \mathcal{F} \tag{1}$$

and

$$[\mathcal{G}',\mathcal{F}] \subset \mathcal{F} . \tag{2}$$

For a weakly reductive homogeneous space $F = G/G'$ a fiber bundle E is soldered to M, if and only if there exists an $\mathcal{F}$-valued linear differential form θ on P' with the following properties.

(θ.1) If $X \in T(P')$ and $\theta(X) = 0$, then $\pi(X) = 0$; π being the projection of P' onto M.

(θ.2) $\theta(Xa) = a^{-1}\theta(X)a, \quad X \in T(P'), \quad a \in G'$.

(θ.3) $\theta(Xa) = 0, \quad X \in P', \quad a \in T(G')$.

Let us now assume that a connection in P is defined by a one-form $\tilde{\omega}$ with values in $\mathcal{G}$. Since P is an extension of P', there exists an injective map γ: $P' \to P$. Then the pullback $\overline{\omega} = \gamma^* \tilde{\omega}$ (the restriction of the form $\tilde{\omega}$ to P') is a $\mathcal{G}$-valued linear differential form; $\overline{\omega}$ does not give a connection in P', because it is not $\mathcal{G}'$ valued. A connection in P defined by $\tilde{\omega}$ is called a connection of Cartan, if the restricted form $\overline{\omega}$ satisfies the following condition: If $X' \in T(P')$ and $\overline{\omega}(x') = 0$, then X' is the zero vector. This implies that $\overline{\omega}$ defines an absolute parallelism in P'.

Let

$$\overline{\omega} = \omega + \theta \tag{3}$$

be the decomposition corresponding to Eq. (1), so that ω is a $\mathcal{G}'$-valued form on P' and θ is a $\mathcal{F}$-valued form on P'. If $F = G/G'$ is weakly reductive, then Eq. (3) expresses the correspondence between a connection of Cartan in P and a unique pair of a connection form ω and a solder form θ in P'.

The structure equation of the Cartan connection reads

$$\tilde{\Omega} = d\tilde{\omega} + \tfrac{1}{2}[\tilde{\omega},\tilde{\omega}] . \tag{4}$$

By restricting both sides of Eq. (4) to P' and by comparing the $\mathcal{G}'$-components and the $\mathcal{F}$ components, we obtain

$$\Omega = \overline{\Omega}_{\mathcal{G}'} - \tfrac{1}{2}[\theta,\theta]_{\mathcal{G}'} , \tag{5}$$

$$\Theta = d\theta + \tfrac{1}{2}[\theta,\omega] + \tfrac{1}{2}[\omega,\theta] . \tag{6}$$

Here

$$\Omega = d\omega + \tfrac{1}{2}[\omega,\omega] \tag{7}$$

is the curvature form of the connection in P', and

$$\Theta = \overline{\Omega}_{\mathcal{G}'} - \tfrac{1}{2}[\theta,\theta]_{\mathcal{G}'} \tag{8}$$

is the torsion form of the connection of Cartan.

If, in addition to Eqs. (1) and (2), the condition

$$[\mathcal{F},\mathcal{F}] \subset \mathcal{G}' \tag{9}$$

also holds, then $\mathcal{F}$ is called a symmetric space. In this case—to be dealt with in Sec. III—Eq. (8) simplifies to $\Theta = \overline{\Omega}_{\mathcal{G}'}$, whereas Ω is given by $\Omega = \overline{\Omega}_{\mathcal{G}'} - \tfrac{1}{2}[\theta,\theta]$. For the sake of completeness, we mention that if the homogeneous space $F = G/G'$ satisfies the stronger condition $[\mathcal{F},\mathcal{F}] = 0$, then $\Theta = \overline{\Omega}_{\mathcal{F}}$ and $\Omega = \overline{\Omega}_{\mathcal{G}'}$. In that case, a connection in P' is called a linear connection.

We shall denote by $\{X_A\}$ the set of generators of the Lie algebra $\mathcal{G}$, with the Lie brackets

$$[X_A,X_B] = f_{AB}^C X_C . \tag{10}$$

Following the notation introduced by MacDowell,[22] the generators of the subgroup G' will be denoted by $\{X_{A_0}\}$ and the generators of the coset space G/G' will be denoted by $\{X_{A_1}\}$. In terms of the set $\{X_A\}$, the connection form and the curvature form of the Cartan connection in P can be written as

$$\tilde{\omega} = \overline{\omega}^A X_A, \quad \tilde{\Omega} = \tilde{\Omega}^A X_A , \tag{11}$$

where the one-forms $\tilde{\omega}^A$ and the two-forms $\tilde{\Omega}^A$ are related by

$$\tilde{\Omega}^A = d\tilde{\omega} + \tfrac{1}{2} f_{BC}^A \tilde{\omega}^B \wedge \tilde{\omega}^C . \tag{12}$$

In P', Eq. (12) can be rewritten in terms of the pulled back forms $\overline{\omega}^A$ and $\overline{\Omega}^A$. As in Eq. (3), we first write

$$\overline{\omega}^A = \omega^{A_0} X_{A_0} + \theta^{A_1} X_{A_1} , \tag{13}$$

where ω^{A_0} and θ^{A_1} are the connection forms and solder forms on P'. Furthermore, we set

$$\overline{\Omega}_{\mathcal{G}'} = \overline{\Omega}^{A_0} X_{A_0}, \quad \overline{\Omega}_{\mathcal{F}} = \overline{\Omega}^{A_1} X_{A_1} . \tag{14}$$

Then, in P', after splitting into the components A_0 and A_1, Eq. (12) becomes

$$\overline{\Omega}^{A_0} = d\omega^{A_0} + \tfrac{1}{2}(f^{A_0}_{B_0C_0}\omega^{B_0}\wedge\omega^{C_0} + f^{A_0}_{B_1C_1}\theta^{B_1}\wedge\theta^{C_1}) , \quad (15)$$

$$\overline{\Omega}^{A_1} = d\theta^{A_1} + \tfrac{1}{2}(f^{A_1}_{B_0C_1}\omega^{B_0}\wedge\theta^{C_1} + f^{A_1}_{B_1C_0}\theta^{B_1}\wedge\omega^{C_0}) . \quad (16)$$

Equations (15) and (16) will allow us to relate the components of the curvature of the Cartan connection in P to the components of the curvature and torsion in P'.

III. FIBER BUNDLES WITH THE de SITTER STRUCTURE GROUP

The de Sitter Lie algebra $\mathcal{SO}(4,1)$ has generators $M_{ab} = -M_{ba}$, $a,b = 0,...,4$, satisfying

$$[M_{ab},M_{cd}] = \eta_{ac}M_{bd} + \eta_{bd}M_{ac} - \eta_{ad}M_{bc} - \eta_{bc}M_{ad} , \quad (17)$$

where

$$\eta_{ab} = \mathrm{diag}(1,-1,-1,-1,-1) . \quad (18)$$

Let now $M_{4i} = \ell\Pi_i$, $i = 0,...,3$, where ℓ is the de Sitter length. Then

$$[M_{ij},M_{kl}] = \eta_{ik}M_{jl} + \eta_{jl}M_{ik} - \eta_{il}M_{jk} - \eta_{jk}M_{il}, \quad (19a)$$

$$[\Pi_i,M_{jk}] = \eta_{ik}\Pi_j - \eta_{ij}\Pi_k , \quad (19b)$$

$$[\Pi_i,\Pi_j] = -(1/\ell^2)M_{ij} , \quad (19c)$$

where $\eta_{ij} = \mathrm{diag}(1,-1,-1,-1,-1)$. In the limit, $\ell\to\infty$, the algebra $\mathcal{SO}(4,1)$ contracts to the Poincaré algebra $\mathcal{ISO}(3,1)$. Equations (19) show that SO(4,1)/SO(3,1) is a symmetric space. The structure constants f^{C}_{AB} of $\mathcal{SO}(4,1)$ are seen to have the form

$$f^{mn}_{ij,kl} = \eta_{ik}\delta^m_j\delta^n_l + \eta_{jl}\delta^m_i\delta^n_k - \eta_{il}\delta^m_j\delta^n_k - \eta_{jk}\delta^m_i\delta^n_l , \quad (20a)$$

$$f^{m}_{i,jk} = \eta_{ik}\delta^m_j - \eta_{ij}\delta^m_k , \quad (20b)$$

$$f^{mn}_{i,j} = -(1/\ell^2)\delta^m_{[i}\delta^n_{j]} . \quad (20c)$$

Therefore, the Cartan metric tensor, defined as

$$C_{AB} = f^{D}_{AC}f^{C}_{BD} , \quad (21)$$

is given as[23]

$$C = \begin{bmatrix} C_{ij,kl} & C_{ij,k} \\ C_{k,ij} & C_{i,j} \end{bmatrix} = \begin{bmatrix} 6(\eta_{il}\eta_{jk} - \eta_{jl}\eta_{ik}) & 0 \\ 0 & (6/\ell^2)\eta_{ij} \end{bmatrix} . \quad (22)$$

Let x^μ, $\mu = 0,...,3$, denote local coordinates in M, the space-time, and y^A denote local coordinates in SO(4,1). Then, since the bundle $P(M,\mathrm{SO}(4,1))$ is locally trivial, the pair (x^μ,y^A) defines local coordinates in P. In the (x,y) space the one-forms $\tilde{\omega}^A$ of the Cartan connection can be written as

$$\tilde{\omega}^A = \tilde{\omega}^A_\mu\, dx^\mu + \tilde{C}^A_B\, dy^B . \quad (23)$$

Here the forms dx^μ and dy^A locally span the cotangent bundle $T^*(P)$. Similarly, the curvature forms $\tilde{\Omega}^A$ are

$$\tilde{\Omega}^A = \tfrac{1}{2}(\tilde{R}^A_{\mu\nu}\, dx^\mu\wedge dx^\nu + \tilde{R}^A_{\mu B}\, dx^\mu\wedge dy^B + \tilde{R}^A_{B\nu}\, dy^B\wedge dx^\nu + \tilde{R}^A_{BC}\, dy^B\wedge dy^C) . \quad (24)$$

Because the curvature form is horizontal, the last three terms in Eq. (24) vanish.[24] This will be of importance in Sec. V, where we derive the BRS transformation for the gravitational field. By choosing a section σ: $M\to P$, one can pull back $\tilde{\omega}$ and $\tilde{\Omega}$ to M. For simplicity, the pullbacks $\sigma^*\tilde{\omega}$ and $\sigma^*\tilde{\Omega}$ will be denoted as $\tilde{\omega}$ and $\tilde{\Omega}$.

We now restrict the forms $\tilde{\omega}$ and $\tilde{\Omega}$, defined on P, to the bundle $P' = P'(M,\mathrm{SO}(3,1))$. The restricted forms, after being pulled back to M, become $\mathcal{SO}(4,1)$-valued one-forms defined on M; they are denoted $\overline{\omega}$ and $\overline{\Omega}$, respectively. Equation (1), specified to the de Sitter case, is

$$\mathcal{SO}(4,1) = \mathcal{SO}(3,1) + \mathbf{R}^4 , \quad (25)$$

whereas Eqs. (15) and (16) become

$$\overline{\Omega}^{ij} = d\omega^{ij} + \omega^i_k\wedge\omega^{jk} - \omega^i_k\wedge\omega^{jk} - (1/2\ell^2)\theta^i\wedge\theta^j , \quad (26)$$

$$\overline{\Omega}^{i} = d\theta^{i} - \omega^i_k\wedge\theta^k + \theta^k\wedge\omega^i_k . \quad (27)$$

The curvature and torsion tensors of the Cartan connection (reduced to P') are introduced through the equations

$$\overline{\Omega}^{ij} = \tfrac{1}{2}\overline{R}^{ij}_{\mu\nu}\, dx^\mu\wedge dx^\nu , \quad (28)$$

$$\overline{\Omega}^{i} = \tfrac{1}{2}\overline{T}^{i}_{\mu\nu}\, dx^\mu\wedge dx^\nu . \quad (29)$$

If we write

$$\omega^{ij} = -\tfrac{1}{2}\omega^{ij}_\mu\, dx^\mu, \quad \theta^i = e^i_\mu\, dx^\mu , \quad (30)$$

then Eqs. (26) and (27) yield

$$\overline{R}^{ij}_{\mu\nu} = -\tfrac{1}{2}(R^{ij}_{\mu\nu} + (2/\ell^2)e^{[i}_\mu e^{j]}_\nu) , \quad (31)$$

$$\overline{T}^{i}_{\mu\nu} = T^{i}_{\mu\nu} . \quad (32)$$

Here

$$R^{ij}_{\mu\nu} = \partial_\mu\omega^{ij}_\nu - \partial_\nu\omega^{ij}_\mu + \omega^i_{\mu k}\omega^{kj}_\nu - \omega^{kj}_\mu\omega^i_{\nu k} \quad (33)$$

and

$$T^i_{\mu\nu} = \partial_\mu e^i_\nu - \partial_\nu e^i_\mu + \omega^i_{\mu k}e^k_\nu - \omega_\nu{}^i{}_k e^k_\mu \quad (34)$$

are the curvature and torsion tensors of the connection ω in P'.

IV. SO(4,1) GAUGE SYMMETRY

In this section, we consider a model of the classical gravity, based on the local gauge group SO(4,1). By classical we mean a theory without FP and gauge fixing terms. We define a gauge configuration to be a Cartan connection $\tilde{\omega}$ on the principal fiber bundle $P = P(M,\mathrm{SO}(4,1))$ over space-time M. Once a section σ: $M\to P$ of P is chosen, the pulled back forms $\tilde{\omega}^A_\mu$, related to $\tilde{\omega}$ through $\tilde{\omega} = \tilde{\omega}^A_\mu X_A$, define the potentials in gauge σ. For each point $u\in P$, the vectors $\{\partial_\mu,D_A\}$ span $T_u(P)$; $\{\partial_\mu\}$ are coordinate basis vectors in M, whereas $\{D_A\}$ are the fundamental vector fields corresponding to X_A.[20] Alternatively, another basis is provided by $\{D_\mu,D_A\}$, where

$$D_\mu = \partial_\mu + \tilde{\omega}^A_\mu D_A \quad (35)$$

are the horizontal lifts of ∂_μ.

Under the action of an infinitesimal group element

$$U = \exp(\epsilon^A D_A) \approx 1 + \epsilon^A D_A , \quad (36)$$

the connection coefficients $\tilde{\omega}^A_\mu$ (gauge potentials) transform as

$$\delta\tilde{\omega}_\mu^A = -(D_\mu)_B^A \epsilon^B = -(\partial_\mu \delta_B^A + \tilde{\omega}_\mu^C f_{CB}^A)\epsilon^B, \quad (37)$$

where the second line of Eq. (37) defines D_μ in the adjoint representation of the group. On the other hand, the curvature coefficients (gauge fields)

$$\tilde{R}_{\mu\nu}^A = \partial_\mu \tilde{\omega}_\nu^A - \partial_\nu \tilde{\omega}_\mu^A + f_{BC}^A \tilde{\omega}_\mu^B \tilde{\omega}_\nu^C \quad (38)$$

transform tensorially under U:

$$\delta\tilde{R}_{\mu\nu}^A = \epsilon^B f_{BC}^A \tilde{R}_{\mu\nu}^C . \quad (39)$$

To write the field equations for $\tilde{\omega}_\mu^A$, we use the Yang–Mills Lagrangian density

$$L = -(\kappa/2) C_{AB} \tilde{\Omega}^A \wedge *\tilde{\Omega}^B , \quad (40)$$

where κ is a coupling constant and where the Hodge $*$ operator defines the dual form $*\tilde{\Omega}$. In curved space, the Hodge operation involves the metric $g_{\mu\nu}$, in terms of which Eq. (40) becomes

$$L = -(\kappa/4) C_{AB} g^{\mu\lambda} g^{\nu\sigma} \tilde{R}_{\mu\nu}^A \tilde{R}_{\lambda\sigma}^B |g|^{1/2} dx^0 \wedge \cdots \wedge dx^3 . \quad (41)$$

We stress that, as in Ref. 7, the dynamical variables are the gauge potentials $\tilde{\omega}_\mu^A$ that make up $\tilde{R}_{\mu\nu}^A$, while $g_{\mu\nu}$ is a spectator and does not participate in the dynamics. This viewpoint also implies that a possible link between the metric and the solder form in P' can be established only *after*[25] the variation of the action based on the Lagrangian density given by Eq. (41).

By making use of Eqs. (33) and (34) and of the explicit form of the Cartan metric, Eq. (22), the Lagrangian density (41) can be expressed in terms of the connection and solder forms in the fiber bundle $P'(M,\mathrm{SO}(3,1))$. Dropping the invariant volume element, we obtain

$$\begin{aligned} L = 3\kappa\big[& (1/2\ell^2) R_{ij\mu\nu} e^{i\mu} e^{j\nu} + \tfrac{1}{4} R_{ij\mu\nu} R^{ij\mu\nu} \\ & + (1/2\ell^4) g^{\mu\lambda} g^{\nu\sigma} (\eta_{il}\eta_{jk} - \eta_{jl}\eta_{ik}) e_\mu^i e_\nu^j e_\lambda^k e_\sigma^l \\ & - (1/2\ell^2) T_{i\mu\nu} T^{i\mu\nu} \big] . \end{aligned} \quad (42)$$

The interpretation of different terms in Eq. (42) was given by Townsend.[12] As mentioned in Sec. I, Townsend does not use the Cartan metric to contract the group indices in Eq. (41), but rather the diagonal metric η_{ab}; this results in different coefficients of the terms in the Lagrangian.

The Euler–Lagrange equations deduced by varying ω_μ^A in the action (41) are

$$\partial_\mu |g|^{1/2} \tilde{R}^{A\mu\nu} + \tilde{\omega}_\mu^C f_{CB}^A |g|^{1/2} \tilde{R}^{B\mu\nu} = 0 . \quad (43)$$

According to the decomposition of the Lie algebra $\mathcal{SO}(4,1)$, described by Eq. (25), we can convert Eq. (43) to a pair of equations in the bundle $P'(M,\mathrm{SO}(3,1))$. By splitting Eq. (43) into components along the generators X_{A_0} and X_{A_1}, we obtain

$$\nabla_\mu (|g|^{1/2} R^{ij\mu\nu} + (2/\ell^2) e^{[i|\mu|} e^{j]\nu}) + (2/\ell^2) |g|^{1/2} T^{ji\nu} = 0 , \quad (44)$$

$$\nabla_\mu (|g|^{1/2} T^{i\mu\nu}) + |g|^{1/2} (R^{i\nu} + (3/\ell^2) e^{i\nu}) = 0 , \quad (45)$$

where ∇_μ denotes the covariant derivative in P' defined in terms of ω_μ^{ij}. Explicit formulas are

$$\nabla_\mu R^{ij\mu\nu} = \partial_\mu R^{ij\mu\nu} + \omega_\mu{}^i{}_k R^{kj\mu\nu} + \omega_{\mu k}{}^i R^{jk\mu\nu} , \quad (46)$$

$$\nabla_\mu T^{i\mu\nu} = \partial_\mu T^{i\mu\nu} + \omega_\mu{}^i{}_k T^{k\mu\nu} . \quad (47)$$

As a consistency check, the same field equations can be derived from Eq. (42), when the action is varied with respect to ω_μ^{ij} and e_μ^i.

Let us now assume that $g_{\mu\nu} = \eta_{ij} e_\mu^i e_\nu^j$ and that the connection ω_μ^{ij} is a metric connection, related to the vierbein field in a standard way. If SO(4,1) is contracted to ISO(3,1), and the torsion is required to vanish, Eq. (45) yields the Einstein limit in vacuum. Equation (44), on the other hand, becomes the Yang equation, the physical content of which was discussed by Fairchild.[26] The theory described by Eqs. (44) and (45) is therefore equivalent to Einstein's gravity when the de Sitter length is large. The differences, which emerge only for small ℓ, can be significant in quantum gravity.

V. THE BRS SYMMETRY

We now supplement the Lagrangian of the Yang–Mills (YM) theory by gauge fixing (GF) and FP gauge-compensating terms. The total Lagrangian density for the quantum gauge theory of gravity thus becomes

$$L = L_{\mathrm{YM}} + L_{\mathrm{GF}} + L_{\mathrm{FP}} , \quad (48)$$

where

$$L_{\mathrm{YM}} = -(\kappa/4) \tilde{R}_{A\mu\nu} \tilde{R}^{A\mu\nu} , \quad (49a)$$

$$L_{\mathrm{GF}} = \tilde{\omega}_\mu^A \partial^\mu \tilde{B}_A + (\alpha/2) \tilde{B}_A B^A , \quad (49b)$$

$$L_{\mathrm{FP}} = -i \partial^\mu \tilde{\bar{c}}_A (D_\mu)_B^A \tilde{c}^B . \quad (49c)$$

Here $\tilde{c}$ and $\tilde{\bar{c}}$ are the FP ghost fields, with Hermitian assignment:

$$\tilde{c}^\dagger = \tilde{c}, \quad \tilde{\bar{c}}^\dagger = \tilde{\bar{c}}, \quad (50)$$

whereas $\tilde{B}^A$ are the Kugo–Ojima[19] auxiliary fields. The parameter α in Eq. (49b) specifies the choice of gauge.

The presence of the L_{GF} and L_{FP} terms results in the BRS symmetry of L, replacing the original gauge symmetry given by Eq. (36). A geometric derivation of the BRS transformation relies on vanishing of the last three terms in Eq. (24). If, in addition, one imposes a constraint $\delta\tilde{\bar{c}}^A = \tilde{B}^A$, where $\delta = dy^A \partial_A$, then the BRS equations read

$$\delta\tilde{\omega}_\mu^A = (D_\mu)_B^A \tilde{c}^A , \quad (51a)$$

$$\delta\tilde{c}^A = -\tfrac{1}{2} f_{EF}^A \tilde{c}^E \wedge \tilde{c}^F , \quad (51b)$$

$$\delta\tilde{\bar{c}}^A = i\tilde{B}^A , \quad (51c)$$

$$\delta\tilde{B}^A = 0 . \quad (51d)$$

In a similar way, the anti-BRS transformation is obtained by considering a fiber bundle locally homeomorphic to $M \times \mathrm{SO}(4,1) \times \mathrm{SO}(4,1)$.[27] The requirement that the curvature form be horizontal then yields three more conditions involving the anti-BRS operator $\delta' = dy'^A \partial_A$,

$$\delta'\tilde{\omega}_\mu^A = (D_\mu)_B^A \tilde{\bar{c}}^A , \tag{52a}$$

$$\delta'\tilde{\bar{c}}^A = -\tfrac{1}{2} f_{EF}^A \tilde{\bar{c}}^E \wedge \tilde{\bar{c}}^F , \tag{52b}$$

$$\delta\tilde{\bar{c}}^A + \delta'\tilde{c}^A + f_{EF}^A \tilde{\bar{c}}^E \wedge \tilde{c}^F = 0 . \tag{52c}$$

Equivalently, Eq. (52c) can be written as

$$\delta'\tilde{c}^A = -f_{EF}^A \tilde{\bar{c}}^E \wedge \tilde{c}^F - i\tilde{B}^A , \quad \delta'\tilde{B}^A = -f_{EF}^A \tilde{\bar{c}}^E \wedge \tilde{B}^F . \tag{53}$$

As usual, the operators δ and δ' satisfy the nilpotency conditions

$$\delta^2 = \delta\delta' + \delta'\delta = \delta'^2 = 0 . \tag{54}$$

There are more general methods, based entirely on the BRS symmetry, to construct the correct quantum Lagrangian from the classical one.[28] The Lagrangian given by Eqs. (48) and (49) is known, however, to be invariant under both the BRS and anti-BRS transformations.[29]

Thus far, both the Lagrangian of the theory and the BRS transformations have been written in the fiber bundle $P(M,\mathrm{SO}(4,1))$. To obtain the BRS equations in the bundle $P'(M,\mathrm{SO}(3,1))$, we follow the by now familiar procedure. First, we restrict the FP forms $\tilde{c}$ and $\tilde{\bar{c}}$ to P' to get $\bar{c}$ and $\bar{\bar{c}}$. Second, similarly to Eq. (30), we write

$$\bar{\bar{c}}^{ij} = -\tfrac{1}{2} c^{ij}, \quad \bar{c}^i = c^i . \tag{55}$$

Then, after projecting Eqs. (51a) and (51b) along the generators of $\mathscr{SO}(3,1)$ and $\mathbb{R}^4$, we obtain

$$\delta\omega_\mu^{ij} = \nabla c^{ij} + (2/\ell^2) e_\mu^{[j} c^{i]} , \tag{56a}$$

$$\delta e_\mu^i = \nabla c^i + e_\mu^k c_k^i , \tag{56b}$$

$$\delta c^{ij} = 2c_k^i \wedge c^{kj} + (1/2\ell^2) c^i \wedge c^j , \tag{56c}$$

$$\delta c^i = c_k^i \wedge c^k . \tag{56d}$$

Similar equations can be written in place of the remaining BRS equations and the anti-BRS equations.

Finally, we note that the Noether currents $\tilde{J}_\mu^B$ and $\tilde{J}_\mu^{B'}$ of the BRS and anti-BRS transformations, as well as the ghost current $\tilde{J}_\mu^c$,[19,29] can be introduced in a standard way:

$$\tilde{J}_\mu^B = \frac{\partial L}{\partial(\partial^\mu \tilde{\omega}_\nu^A)} (D_\nu)_E^A \tilde{c}^E - \frac{1}{2} \frac{\partial L}{\partial(\partial^\mu \tilde{\bar{c}}^A)} f_{EF}^A \tilde{c}^E \wedge \tilde{c}^F + i \frac{\partial L}{\partial(\partial^\mu \tilde{\bar{c}}^A)} \tilde{B}^A , \tag{57}$$

$$\tilde{J}_\mu^{B'} = \frac{\partial L}{\partial(\partial^\mu \tilde{\omega}_\nu^A)} (D_\nu)_E^A \tilde{\bar{c}}^E - \frac{1}{2} \frac{\partial L}{\partial(\partial^\mu \tilde{\bar{c}}^A)} f_{EF}^A \tilde{\bar{c}}^E \wedge \tilde{\bar{c}}^F - \frac{\partial L}{\partial(\partial^\mu \tilde{c}^A)} (f_{EF}^A \tilde{\bar{c}}^E \wedge \tilde{c}^F + i\tilde{B}^A) - \frac{\partial L}{\partial(\partial^\mu \tilde{B}^A)} f_{EF}^A \tilde{\bar{c}}^E \wedge \tilde{B}^F , \tag{58}$$

$$\tilde{J}_\mu^c = i[\tilde{\bar{c}}_A (D_\mu)_B^A \tilde{\bar{c}}^B - \partial_\mu \tilde{\bar{c}}_A \tilde{c}^A] . \tag{59}$$

As has been shown by Kugo and Ojima,[19,29] the unitarity of the S matrix results when physical states are subject to the following subsidiary conditions:

$$\tilde{Q}_B |\mathrm{phys}\rangle = \tilde{Q}_c |\mathrm{phys}\rangle = \tilde{Q}'_B |\mathrm{phys}\rangle = 0 , \tag{60}$$

where the Q's are the charges associated with the Noether currents. The initial proof, carried out for compact groups, was already extended in Ref. 29 to the noncompact group $\mathrm{Sl}(2,C)$. In the absence of a formal proof for the simple, but noncompact, SO(4,1), we follow the trend to recognize the BRS invariance principle as a general procedure for gauge theory.[28]

An argument in favor of unitarity goes as follows. If $\tilde{\pi}_\mu^A$ denote the canonical momenta conjugate to $\tilde{\omega}_\mu^A$, the equal-time commutation relations

$$[\tilde{\omega}_\mu^A(\mathbf{x},t), \tilde{\pi}_\nu^B(\mathbf{y},t)] = iC^{AB} g_{\mu\nu} \delta(\mathbf{x}-\mathbf{y}) \tag{61}$$

imply that the fields $\tilde{\omega}_\mu^A$ for $\mu = 0$ and $A = 1,2,3,12,23,31$, and for $\mu = 1,2,3$ and $A = 0,01,02,03$ are nonphysical negative-metric fields; this is similar to the situation encountered in the quantization of the $\mathrm{Sl}(2,C)$ Yang–Mills field.[30] Assume now that by using the gauge freedom, one can eliminate the fields $\tilde{\omega}_\mu^A$, where $\mu = 1,2$ and where $A = 0,3,01,02,03$. Then, by following the proof of Kugo and Ojima,[19] one checks that the quartet mechanism—implied by the BRS equations—leads to exact cancellation between the nonphysical classical modes (longitudinal and scalar) and the ghost modes. The remaining ten fields $\tilde{\omega}_\mu^A$, where $\mu = 1,2$ and $A = 1,2,12,23,31$ are all positive metric. We also note that a different construction of gauge theory of noncompact groups was proposed by Cahill.[31]

VI. CONCLUSIONS

We have formulated the gravity as a gauge theory with the structural group SO(4,1) or its contraction[32] ISO(3,1). The natural setting for this theory is a fiber bundle endowed with a connection of Cartan, and not a bundle of frames in which one can define the forms of linear connection and of soldering. In the bundle $P(M,\mathrm{SO}(4,1))$ the torsion is thus undefined; the only gauge fields of the theory being the components of the curvature tensor. By postulating a quadratic Lagrangian, we are thus led to Yang's formulation of gravity expressed in terms of the Cartan connection. In the restricted bundle $P'(M,\mathrm{SO}(3,1))$, we recover a theory with torsion, thus extending Einstein's gravity. The quantum gauge theory—with the gauge symmetry taken over by the BRS symmetry—is also formulated naturally in $P(M,\mathrm{SO}(4,1))$. To include the matter fields, one can set up a gauge theory with sources. A better method, however, seems to be provided by supergravity; the theory would then remain sourceless.

ACKNOWLEDGMENTS

Many friends and colleagues helped me in this work through useful conversations and correspondence. I am especially grateful to K. Cahill, S. J. Gates, Jr., T. Goldman, R. J. Hughes, V. A. Kostelecký, F. Mansouri, E. Mottola, H. T. Nieh, S. Raby, and R. Slansky.

[1] R. Utiyama, Phys. Rev. **101**, 1597 (1956); D. W. Sciama, in *Recent Developments in General Relativity*, Infeld's Festschrift (Pergamon, New York, 1962), p. 415; T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).

[2]A. Trautman, Rep. Math. Phys. **1**, 29 (1970); *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 1, p. 287; *Differential Geometry for Physicists* (Bibliopolis, Naples, 1984).
[3]D. Ivanenko and G. Sardanashvily, Phys. Rep. **94**, 1 (1983); N. Nissani, *ibid.* **109**, 95 (1984).
[4]S. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969); C. G. Callan, Jr., S. Coleman, J. Wess, and B. Zumino, *ibid.* **177**, 2247 (1969); L. Alvarez-Gaumé and P. Ginsparg, Nucl. Phys. B **262**, 439 (1985).
[5]Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975); N. Nakanishi and I. Ojima, Phys. Rev. Lett. **43**, 91 (1979).
[6]L. N. Chang and F. Mansouri, Phys. Rev. D **17**, 3168 (1978).
[7]F. Mansouri and C. Schaer, in *Supergravity*, edited by P. van Nieuwenhuizen and D. Z. Friedman (North-Holland, Amsterdam, 1979), p. 169.
[8]S. Kobayashi, Can. J. Math. **8**, 145 (1956); S. Kobayashi, Ann. Math. **43**, 119 (1957).
[9]W. Drechsler, J. Math. Phys. **18**, 1358 (1977).
[10]A. Trautman, Symp. Math. **12**, 139 (1973); W. Drechsler, Ann. Inst. H. Poincaré **37**, 155 (1982).
[11]W. Drechsler, Fortschr. Phys. **23**, 649 (1975); W. Drechsler, in *Fiber Bundle Techniques in Gauge Theories, Lecture Notes in Physics*, Vol. 67 (Springer, Berlin, 1977), p. 145; W. Drechsler and W. Thacker, Class. Quantum Gravit. **4**, 291 (1987).
[12]P. K. Townsend, Phys. Rev. D **10**, 2795 (1977).
[13]S. W. MacDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[14]M. Martellini and P. Sodano, Phys. Rev. D **22**, 1325 (1980).
[15]K. S. Stelle, Phys. Rev. D **16**, 953 (1977); E. Sezgin and P. van Nieuwenhuizen, *ibid.* **21**, 3269 (1980); D. E. Sezgin, *ibid.* **24**, 1677 (1981); D. E. Neville, *ibid.* **18**, 3535 (1978); **21**, 867 (1980); R. Rauch and H. T. Nieh, *ibid.* **24**, 2029 (1981); R. Rauch, J. C. Shaw, and H. T. Nieh, Gen. Relativ. Gravit. **14**, 331 (1982); H. T. Nieh, in *Proceedings of the Third Marcel Grossmann Meeting on General Relativity*, edited by H. Ning (North-Holland, Amsterdam, 1983), p. 129.
[16]F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976); D. Grensing and G. Grensing, Phys. Rev. D **28**, 286 (1983); J. M. Nester, in *An Introduction to Kaluza–Klein Theories* (World Scientific, Singapore, 1984), p. 83.
[17]C. A. López, Int. J. Theor. Phys. **16**, 167 (1977). For a recent treatment of gauge theory of gravity with quadratic Lagrangian see V. Szczyrba, Phys. Rev. D **36**, 351 (1987).
[18]C. N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[19]T. Kugo and I. Ojima, Suppl. Progr. Theor. Phys. **66**, 1 (1979).
[20]When specifying various budles of interest, we essentially follow the notation of S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963).
[21]C. Ehresmann, in *Colloque de Topologie* (Centre Belge de Recherches Mathématiques, Louvain, 1950), p. 29.
[22]S. W. MacDowell, Phys. Lett. B **80**, 212 (1979).
[23]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (World Scientific, Singapore, 1986), p. 50.
[24]J. Thierry-Mieg and Y. Ne'eman, Ann. Phys. (NY) **123**, 247 (1979); Y. Ne'eman and J. Thierry-Mieg, Proc. Natl. Acad. Sci. USA **77**, 720 (1980); J. Thierry-Mieg, J. Math. Phys. **21**, 2834 (1980).
[25]A similar viewpoint, in which the connection is assumed to be Christoffel connection after the variation of the action integral, was advanced by G. Stephenson, Nuovo Cimento **9**, 263 (1958).
[26]E. E. Fairchild, Jr., Phys. Rev. D **14**, 384 (1976); **16**, 2438 (1977).
[27]M. Quirós, F. J. De Uries, J. Hoyos, M. L. Mazon, and E. Rodriguez, J. Math. Phys. **22**, 1767 (1981).
[28]T. Kugo and S. Uehara, Nucl. Phys. B **197**, 378 (1982); L. Baulieu and J. Thierry Mieg, *ibid.* B **197**, 477 (1982); L. Alvarez-Gaumé and L. Baulieu, *ibid.* B **212**, 255 (1983).
[29]I. Ojima, Prog. Theor. Phys. **64**, 625 (1980).
[30]J. P. Hsu and D. Xin, Phys. Rev. D **24**, 471 (1981).
[31]K. Cahill, Phys. Rev. D **18**, 2930 (1978); **20**, 2636 (1979); K. Cahill and S. Özenli, *ibid.* **27**, 1396 (1983).
[32]In this context, a different model was proposed by R. Aldovandi and J. G. Pereira, Phys. Rev. D **33**, 2788 (1986).

Exact solutions for discrete kinetic models with ternary collisions

Henri Cornille
Service de Physique Théorique, CEN–Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 21 July 1987; accepted for publication 9 March 1988)

Discrete velocity Boltzmann models with three or four independent densities and only one independent binary collision term are considered. As has been suggested, multiplying the binary collision term by $1+\tau M$, M being the total mass (the sum of all densities), one could include ternary collisions. However, a phenomenological point of view is adopted, τ being an effective parameter for multiparticle collisions, while the restriction $1+\tau M>0$ ensures the validity of the H theorem. Exact solutions for three $(1+1)$-dimensional models, the six-velocity Broadwell model and the four- and the six-velocity planar models, are discussed. Both similarity shock-wave solutions and $(1+1)$-dimensional solutions with two exponential variables are obtained. These last solutions are either periodic solutions or infinitely weak shock solutions.

I. INTRODUCTION

As a result of many difficulties occurring in the study of the continuous Boltzmann equation, physicists and mathematicians have suggested simpler models where the velocity can only take the discrete values $\mathbf{v}_i$. The oldest model, the four-velocity planar model, is attributed to Maxwell. Since the discovery of the popular six-velocity Broadwell[1] model many others have been proposed.[2] To each discrete velocity $\mathbf{v}_i$, $|\mathbf{v}_i|=1$, is associated a density N_i and in $1+1$ dimensions (space x, time t) the number of independent densities is in general less than the number of velocities.

For the Boltzmann collision term it is assumed that only binary collisions between the particles can occur. Consequently for the Boltzmann discrete models the N_i satisfy a system of semilinear coupled partial differential equations with quadratic nonlinearities. The similarity shock-wave solutions are rational solutions with one exponential variable, $\exp(\gamma u)$, with $u=x+ct$ and c, $|c|<1$, being the shock velocity:

$$N_i^B = n_{0i} + n_i/D\,, \quad D = 1 + d\exp(\gamma u)\,, \quad u = x + ct\,. \tag{1.1}$$

It has recently been understood[3] that the $(1+1)$-dimensional rational solutions are simply the sums of two such similarity waves and four classes of solutions were found: (i) $(1+1)$-dimensional shock waves,[3] (ii) periodic solutions in space propagating when the time is growing,[3] (iii) periodic nonpropagating solutions,[3–5] and (iv) densities N_i not relaxing towards constant Maxwellians.[3]

The first motivation of the present work is to investigate whether or not similar exact solutions exist when the collision term includes ternary collisions. Then the semilinear discrete model equations have a mixing of cubic and quadratic nonlinearities.

The problem of the inclusion of triple collisions (more generally, of multiple collisions) is not simple; however, in $1+1$ dimensions a recipe has been proposed by Harris,[2] Gatignol,[2] and Platkowski.[6] Let us consider discrete models (more than two $\mathbf{v}_i$) having only one independent collision term. The system of equations being of the type $N_{it}+N_{ix}=\mathrm{Col}_i^B$, it was suggested that ternary collisions be included by the substitution

$$\mathrm{Col}_i^B \to \mathrm{Col}_i^T = (1+\tau M)\mathrm{Col}_i^B\,, \quad M=\sum N_i\,, \tag{1.2}$$

with M being the total mass and $\tau>0$ being interpreted as the ternary cross section. For such cubic nonlinearities the exact similarity shock-wave solution is no longer a rational solution with pole singularity but a solution with algebraic branch point of the square-root type

$$N_i^T = n_{0i} + n_i D^{-1/2}\,,$$
$$D = 1 + d\exp(\gamma u)\,, \quad u = x + ct\,. \tag{1.3}$$

To my knowledge, similarity solutions (1.4) were not previously known and their study is the second motivation of this work.

As in the binary case, the exact $(1+1)$-dimensional solutions are the sums of two similarity waves:

$$N_i^B = n_{0i} + \sum_{j=1,2} n_{ji} D_j^{-1}\,,$$
$$N_i^T = n_{0i} + \sum_{j=1,2} n_{ji} D_j^{-1/2}\,,$$
$$D_j = 1 + d_i \exp(\gamma_j x + \rho_j t)\,, \tag{1.4}$$

$n_{0i}, n_{ji}, \gamma_j, \rho_j, d_i$ being constants.

As recalled above τ has been considered as a positive quantity. However, here we adopt a phenomenological viewpoint and the collision term (1.2) is more like an effective collision term including the multiple collision process, τ being a phenomenological parameter. If corrections to the binary collisions have to be included, there is no physical reason why the additional factor $1+\tau M$ would be always larger than 1. Of course the character of gain minus loss terms must be respected and we must require $1+\tau M\geqslant 0$ for the H theorem. However, it seems reasonable that the corrective factor be sometimes larger than 1 and sometimes smaller than 1. So the third motivation, a phenomenological one, is to compare both the $\tau>0$ and $\tau<0$ possibilities.

We study two $(1+1)$-dimensional models with three

 0022-2488/88/071667-11$02.50

independent N_i: first, the six-velocity Broadwell model with equations

$$N_{0t} + N_{0x} = N_{2t} - N_{2x}$$
$$= -bN_{1t} = (1+\tau M)(N_1^2 - N_0N_2)\,, \quad (1.5)$$

with $M = N_0 + N_2 + 2bN_1$ being the total mass and $b = 2$; second, the four-velocity (Gatignol[2]) planar model with $b = 1$ in (1.5). We report also some results for the hexagonal six-velocity planar model with four independent N_i,

$$N_{0t} + N_{0x} = N_{3t} - N_{3x} = -2N_{1t} - N_{1x}$$
$$= -2N_{2t} + N_{2x} = (1+\tau M)(N_1N_2 - N_0N_3)\,, \quad (1.6)$$

with the total mass $M = N_0 + N_3 + 2(N_1 + N_2)$. For these models the conservation laws of mass and momentum are satisfied; however, for the H theorem with $H = \Sigma N_i \log N_i$, the condition $1 + \tau M \geqslant 0$ is necessary:

$$H_t + \partial_x(\cdots) = (1+\tau M)(N_1^2 - N_0N_2)\log(N_0N_2/N_1^2)\,,$$
$$H_t + \partial_x(\cdots) = (1+\tau M)\log(N_0N_3/N_1N_2)\,. \quad (1.7)$$

In Sec. II, for (1.5) we study the exact similarity shock waves, both from the macroscopic and microscopic viewpoints. We are mainly interested in the ratio of the two shock thicknesses whether ternary collisions are included or not. For the macroscopic study we find that the nonlinear differential equation satisfied by the total mass M is quadratic in the binary case (M^B) and cubic in the ternary (M^T) one. Assuming in both cases the same shock limits and the same shock velocity we find, in the variable $u = x + ct$,

$$M^B = m_0 + m/D^B\,, \quad D^B = 1 + \exp(\gamma^B u), \quad \tau = 0;$$
$$M^T = m_0 + m/\sqrt{D^T}\,, \quad D^T = 1 + \exp(\gamma^T u)\,, \quad \tau \neq 0\,. \quad (1.8)$$

We notice the change of analytical structure at $\tau = 0$. Furthermore τ is not arbitrary, being fixed by the macroscopic shock limits. We determine the shock thickness $w = |m|/\mathrm{Max}|M_u|$ and compute the ratio $R = w^T/w^B$. We find $R < 3\sqrt{3}/8 < 1$ if $\tau > 0$ [compare with the Gatignol result[2] for the model (1.6)] while if $\tau < 0$ we cannot draw that conclusion, both $R > 1$ and $R < 1$ being allowed. Then starting from the same knowledge of macroscopic quantities we construct the exact microscopic similarity solutions and build up their total mass. The results of the microscopic study confirm those of the macroscopic one. From explicit examples we show that in the $\tau < 0$ case we can have both ratios R larger or smaller than 1.

In Sec. III, always for the models (1.5), we build up the (1 + 1)-dimensional solutions of the type (1.4). We have obtained two classes of exact solutions for $\tau < 0$. In the first class the two similarity components are real and we find solutions with infinitely weak shocks. In the second class, the two components are complex conjugate and we obtain periodic solutions

$$N_i = n_{0i} + 2\,\mathrm{Re}\, n_i D^{-1/2}\,, \quad D = 1 + d\exp(i\gamma_I x + \rho_R t)\,, \quad (1.9)$$

with d, n_i complex and γ_I, ρ_R real which means that the solutions are nonpropagating.

In Sec. IV we sketch briefly some results for the model (1.6) with four independent N_i. We still construct solutions that are periodic in the space variable but nonpropagating when the time is growing.

II. EXACT SIMILARITY SHOCK WAVE SOLUTIONS FOR THE MODELS WITH THREE INDEPENDENT N_i

Equation (1.5) with the $u = x + ct$ variable becomes a nonlinear differential system

$$(c+1)N_{0u} = (c-1)N_{2u}$$
$$= -bN_{1u} = (1+\tau M)(N_1^2 - N_0N_2)\,. \quad (2.1)$$

From (2.1) we can deduce the nonlinear differential equation satisfied by M^B (binary collisions) and M^T (ternary included). All N_{iu} as well as their linear combination M_u are proportional. Consequently all N_i and M are proportional (up to constants) and the rhs of (2.1) becomes a quadratic polynomial for M^B and a cubic one for M^T. In order to compare the shock profiles in both binary and ternary collisions included we assume the same shock limits and the same shock speed. Here M_u must vanish at the shock limits and we deduce from (2.1),

$$M_u^B = \lambda(M^B - \alpha)(M^B - \beta)\,,$$
$$M_u^T = \lambda(1+\tau M^T)(M^T - \alpha)(M^T - \beta)\,, \quad (2.2)$$

with λ, α, β being constants. We notice that for the nonlinear (2.2) equation, M is not necessarily analytic at $\tau = 0$, and for the exact solutions (1.8), $\tau = 0$ is a singular value. From both a macroscopic and a microscopic studies we build up M^B, M^T and compare the ratio of their shock thickness.

A. Macroscopic study

We seek exact solutions of (2.2) depending only on the variable $\exp(\gamma u)$ (in the following sometimes we suppress the superscript B or T when no confusion is possible). Assuming in (2.2) denominators D^p, then the balance of linear versus nonlinear parts gives $p = 1$ in the binary and $p = 1/2$ in the ternary cases. We find the ansatz (1.8) that we substitute into (2.2) and get in the binary $\tau = 0$ case

$$m_0 = \alpha\,, \quad m + m_0 = \beta\,, \quad \gamma^B = \lambda m\,, \quad w^B = 4/|\gamma^B|\,, \quad (2.3)$$

with, as usual, $w = |m|/\mathrm{Max}|M_u|$, $\mathrm{Max}|M_u|$ being obtained at $u = 0$ and $D^B = 2$. In the ternary case $\mathrm{Max}|M_u|$ is for $\gamma^T u = \log 2$ or $D^T = 3$; substituting M^T into (2.2) we find a constant term $(1+\tau m_0)(m_0 - \alpha)(m_0 - \beta) = 0$ which leads to two different classes A and B depending on whether $1 + \tau m_0 = 0$ or not. Further, for class B we find two subclasses $1 + \tau m_0 = \zeta\tau m$, $\zeta = \mp 1$:

Class A: $\tau = -m_0^{-1}$, $m_0 + m = \alpha$, $m_0 - m = \beta$;

Class B: $\tau = (\zeta m - m_0)^{-1}$, $m_0 = \alpha$, $m_0 + m\zeta = \beta$;

Classes A and B: $\gamma^T = 2\lambda\tau m^2$, $w^T = 3\sqrt{3}/|\gamma^T|\,.$ (2.4)

We find for the ratio $R = w^T/w^B = |3\sqrt{3}\gamma^B/4\gamma^T|$,

$$R = 3\sqrt{3}(8|\tau m|)^{-1}\,. \quad (2.5)$$

We note the different τ values in classes A and B determined by the shock limits $m_0, m_0 + m$. We must have $M \geqslant 0$ or (due to $D \geqslant 1$) $m_0 \geqslant 0$, $m_0 + m \geqslant 0$. Further, for the H theorem we

require $1+\tau M\geqslant 0$ or $1+\tau m_0\geqslant 0$, $1+\tau(m+m_0)\geqslant 0$. Taking into account these constraints and (2.2)–(2.5) we find the following.

(i) If $\tau>0$, then M^T belongs to class B, $\zeta=1$ with $m\geqslant m_0\geqslant 0$, $|\tau m|^{-1}=1-m_0/m<1$, leading to $R<3\sqrt{3}/8<1$. The ternary collision thickness is smaller than the binary one.

(ii) If $\tau<0$ we find for class A $-m_0\leqslant m\leqslant 0$ or $|\tau m|^{-1}>1$. Then $R>3\sqrt{3}/8$ which can be either larger or smaller than 1. For the subclass B, $\zeta=1$ with $-m_0\leqslant m\leqslant 0$, $|\tau m|^{-1}=1+m_0/|m|>2$ or $R>3\sqrt{3}/4>1$, it is the only case where we can conclude that the ternary thickness is larger than the binary one. For the other subcase B, $\zeta=-1$, then $m>0$, and $R=3\sqrt{3}(1+m_0/m)/8$ can be either larger or smaller than 1. We summarize the results:

$$\begin{aligned}&\tau>0\text{: class B, } \zeta=1\text{: } && R<3\sqrt{3}/8<1;\\ &\tau<0\text{: class A or class B, } \zeta=-1\text{: } && R>3\sqrt{3}/8;\\ &\tau<0\text{: class B, } \zeta=1\text{: } && R>3\sqrt{3}/4>1\,.\end{aligned}\tag{2.6}$$

In conclusion, if $\tau>0$ then necessarily $R>1$ while if $R<1$ we cannot draw that conclusion.

B. Microscopic study

We assume that m_0, m_0+m positive shock limits and the shock velocity c are given macroscopic quantities. We wish to construct the associated classes of positive microscopic densities N_i. For the binary collisions alone and the ternary collisions included we start with

$$\begin{aligned}&N_i^B=n_{0i}^B+n_i^B/D^B,\quad D^B=1+\exp(\gamma^B u)\,,\\ &N_i^T=n_{0i}^T+n_i^T/(D^T)^{-1/2},\quad D^T=1+\exp(\gamma^T u)\,,\\ &u=x+ct\,,\end{aligned}\tag{2.7}$$

and build up M^B and M^T of the (1.8) type. In both cases $M=N_0+N_2+2bN_1$, $m_0=n_{00}+n_{02}+2bn_{01}$, $m=n_0+n_2+2bn_1$ (where the omitted superscripts B and T must be understood). We get, if we substitute (2.7) into (2.1), the relations between the parameters in both binary alone or not alone collisions. In the binary case, the constant present only in the collision term gives $n_{00}n_{02}-n_{01}^2=0$, while in the ternary case $(1+\tau m_0)(n_{01}^2-n_{00}n_{02})=0$ still leads to two classes A and B. As in the macroscopic formalism for the class B we get $1+\tau m_0=\zeta\tau m$ with two subclasses $\zeta=\mp 1$ (we drop the superscripts B and T):

$$\text{Binary: } bn_1(n_0+n_2)+2n_0n_2=0,\quad n_2+n_0=c(n_2-n_0),\quad -2bc\gamma=2n_1+b(n_0+n_2),$$
$$2n_1n_{01}=n_0n_{02}+n_2n_{00}+n_0n_2-n_1^2,\quad n_{01}^2=n_{00}n_{02}\,;\tag{2.8a}$$

$$\text{Ternary class A: } \tau=-m_0^{-1},\quad bn_1(n_0+n_2)=-2n_0n_2,\quad n_2+n_0=c(n_2-n_0)\,,$$
$$cb\gamma=-\tau m(2n_1+b(n_0+n_2)),\quad 2n_1n_{01}=n_{00}n_2+n_{02}n_0,\quad n_{01}^2=n_{00}n_{02}-n_1^2+n_0n_2\,;\tag{2.8b}$$

$$\text{Ternary class B: } \tau=(\zeta m-m_0)^{-1},\quad bn_1(n_0+n_2)=-2n_0n_2,\quad n_2+n_0=c(n_2-n_0)$$
$$cb\gamma=-\tau m(2n_1+b(n_0+n_2)),\quad 2n_1n_{01}=n_0n_{02}+n_2n_{00}+\zeta(n_0n_2-n_1^2),\quad n_{01}^2=n_{00}n_{02}\,.\tag{2.8c}$$

As in the macroscopic formalism, τ has the same values fixed by the shock limits m_0 and $m+m_0$. For the ten remaining parameters n_{0i}, n_i, γ, c, m_0, m we have seven relations, so that we can choose (m_0,m,c) as the three arbitrary macroscopic quantities. We notice for the subclass B, $\zeta=1$, that the relations are the same as the binary ones (except for γ).

From the second and third (2.8a)–(2.8c) relations we find $\gamma^T=2m\tau\gamma^B$, leading to a thickness ratio $R=w^T/w^B=3\sqrt{3}(8|\tau m|)^{-1}$, which is the same as the macroscopic one (2.5).

We want to show that the shock-limit difference m and the shock speed c give the same n_i values or $n_i^B=n_i^T$. For this result we remark that the first and second (2.8a)–(2.8c) relations are the same. We introduce intermediate parameters $y=n_0/n_2$ and $\bar{n}_1=n_1/n_2$, functions of c alone while n_2 can be expressed in terms of m and c. Finally, multiplying by n_2, we recover the original parameters n_0,n_1, functions of c,m ($n_0=yn_2$, $n_1=\bar{n}_1n_2$):

$$\begin{aligned}&y=n_0/n_2=(1+c)/(1-c),\\ &\bar{n}_1=n_1/n_2=-2y/b(1+y),\\ &n_2=m(1+y)/(1-y)^2\,.\end{aligned}\tag{2.9}$$

When n_i,τ are known as functions of (m,m_0,c), from the third (2.8a)–(2.8c) relation we determine γ^B and γ^T as functions of these free parameters.

We determine the n_{0i} parameters. In the binary formalism and the ternary B, $\zeta=1$ subcase these n_{0i} are the same while in the other cases the relations are different. We notice that even if at the macroscopic level $m_0=m_0^B=m_0^T$, at the microscopic level n_{0i}^B and n_{0i}^T can be different. There exist two linear relations between the n_{0i} and a third one which is quadratic. There is a common relation $2n_{01}=(m_0-n_{00}-yn_{02})/b$ while the two others can be different:

$$\text{Binary: } 2n_{01}=(n_{00}+yn_{02}+n_2(y-\bar{n}_1^2))/\bar{n}_1,\quad n_{01}^2=n_{00}n_{02},\tag{2.10a}$$
$$\text{Ternary class A: } 2n_{01}=(n_{00}+yn_{02})/\bar{n}_1,\quad n_{01}^2=n_{00}n_{02}+n_2(y-\bar{n}_1^2),\tag{2.10b}$$
$$\text{Ternary class B: } 2n_{01}=(n_{00}+yn_{02}+\zeta n_2(y-\bar{n}_1^2))/\bar{n}_1,\quad n_{01}^2=n_{00}n_{02}.\tag{2.10c}$$

Choosing any one n_{0i} with i fixed, we find a second degree algebraic equation with two solutions. This means, for instance, that for class B we have four subcases.

For the N_i being constructed we must check their positivity as well as the positivity of $1+\tau M$ which is trivially tested from m, m_0, and τ. The necessary and sufficient conditions for $N_i\geqslant 0$ are $n_i\geqslant 0$ and $n_i+n_{0i}\geqslant 0$. We must use the computer for the numerical resolution and as an illustration

we give some numerical examples. The reader can check the physical constraints: m_0, m_0+m, $1+\tau m_0$, $1+\tau(m+m_0)$, n_{0i}, $n_{0i}+n_i$ are non-negative quantities. In all cases we choose $m_0=1$ and discuss first the Broadwell, $b=2$ model for different m,c values. In Fig. 1 we plot only the total mass M^B, M^T. For $\tau>0$, as explained above, we must choose class B, $\zeta=1$ for which we know $R<3\sqrt{3}/8<1$ and $n_{0i}^B=n_{0i}^T$, $n_i^B=n_i^T$. In Fig. 1(a) we choose $m=1.1, c=-0.9$ and deduce $\tau=10$, $R=0.06$, as well as $n_{00}=0.29$, $n_{01}=0.16$, $n_{02}=0.085$, $n_0=0.94$, $n_1=0.05$, $n_2=-0.05$, $\gamma^B=0.52$, $\gamma^T=11.53$. For $\tau<0$, we begin with class A knowing only that $R>3\sqrt{3}/8$ and give examples with $R<1$ and $R>1$. In Fig. 1(b), starting with $m=-0.9$, $c=-0.9$ we find $\tau=-1$, $R=0.72<1$, $n_0=-0.769-0.043$, $n_2=0.04$, and for the n_{0i} and γ which are different, $\gamma^B=-0.43$, $\gamma^T=-0.77$, $n_{00}^T=0.76$, $n_{01}^T=0.046$, $n_{02}^T=0.045$, $n_{00}^B=0.773$, $n_{01}^B=0.056$, $n_{02}^B=0.04$. Moving on, in Fig. 1(c) starting with $m=-0.1$, $c=0.3$ we find an example $\tau=-1$, $R=6.5>1$ and $n_0=0.01$, $n_1=-0.022$, $n_2=-0.019$, $\gamma^B=0.05$, $\gamma^T=0.01$, $n_{00}^T=0.36$, $n_{01}^T=0.14$, $n_{02}^T=0.058$, $n_{00}^B=0.34$, $n_{01}^B=0.15$, $n_{02}^B=0.06$. Finally for the last $\tau<0$ subclass B, $\zeta=-1$ for which macroscopically both $R>1$ and $R<1$ are allowed, taking into account the microscopic positivity $N_i>0$, we have not found numerical examples with $R<1$ but only examples with $R>1$.

For the four-velocity planar Gatignol model with $b=1$ in (2.1), we give only one example of class A for which $\tau<0$ but $R<1$ in Fig. 1(d). Starting with $m_0=1$, $m=-0.95$, $c=-0.98$, we find $R=0.68$, $n_0=-0.92$, $n_1=-0.018$, $n_2=0.93$, $\gamma^B=-0.48$, $\gamma^T=-0.92$, $n_{00}^T=0.92$, $n_{01}^T=0.03$, $n_{02}^T=0.01$, $n_{00}^B=0.93$, $n_{01}^B=0.031$, $n_{02}^B=0.001$.

In order to show that the $M_u^T-M_u^B$ sign cannot be only τ dependent, it is sufficient to remark that for τ fixed, this sign depends on u. For instance, in Fig. 1(a) this sign is positive for $u\in(0,0.5)$ and negative outside; while in Fig. 1(b) $M_u^T-M_u^B$ positive for $u<0.7$ is negative for $u>0.7$.

C. Scaling property

In (2.1) we note the invariance $(N_i,M,u,\tau)\to(\phi N_i,\phi M,\phi^{-1}u,\phi^{-1}\tau)$, ϕ being a scaling parameter. We directly find this invariance from the microscopic relations (2.8a)–(2.8c). If the macroscopic quantities (m_0,m,c) become $(\phi m_0,\phi m,c)$ then $n_{0i}\to\phi n_{0i}$, $n_i\to\phi n_i$, $\tau\to\tau\phi^{-1}$, $\gamma\to\phi\gamma$, while the thickness ratio R proportional to $|\gamma^B/\gamma^T|$ is invariant. In consequence the same shock profiles (in particular Fig. 1) are represented by $\tau\to\phi^{-1}\tau$ if we scale the u axis by $u\phi^{-1}$ and the M total mass axis by $M\phi$. Doing that transformation, Fig. 1 represents arbitrarily small τ values as well as arbitrarily large ones. What remains invariant is the product τM. This transformation, while giving the possibility of small τ parameter values, does not modify the contribution of the ternary collisions into $1+\tau M$.

Can we have $|\tau M|$ small or at least less than 1? We recall that $\tau M\in(\tau m_0,\tau(m_0+m))$. For class A or subclass B, $\zeta=-1$, either $\tau m_0=-1$ or $\tau(m_0+m)=-1$ and the answer is no. The same result holds for subclass B, $\zeta=1$ and $\tau>0$ for which $m\geqslant m_0>0$ leads to $\tau(m_0+m)\geqslant 0$. The remaining case is for the subclass B, $\zeta=1$, $\tau<0$ for which $-m_0<m<0$ and $|\tau M|\in(m_0/(m_0-m),\ (m_0+m)/(m_0-m))$ can be less than 1 but at least $|\tau M|\geqslant\frac{1}{2}$.

III. (1+1)-DIMENSIONAL SOLUTIONS FOR THE MODELS WITH THREE INDEPENDENT N_i

We still study the model (1.5), b real being a parameter, explain the algebraic determination of the solutions (all details are in Appendix A), and finally, make explicit the results for the physical cases: $b=1,2$, $N_i\geqslant 0$, $1+M\geqslant 0$. We start with the ansatz (1.4) for the N_i, write the associated M,

$$
\begin{aligned}
&N_i=n_{0i}+\sum_{j=1,2}n_{ji}D_j^{-1/2},\\
&D_j=1+d_j\exp(\gamma_j x+\rho_j t),\quad d_j>0,\\
&M=m_0+\sum m_j D_j^{-1/2},\quad m_0=n_{00}+n_{02}+2bn_{01},\\
&m_j=n_{j0}+n_{j2}+2bn_{j1},
\end{aligned}
\tag{3.1}
$$

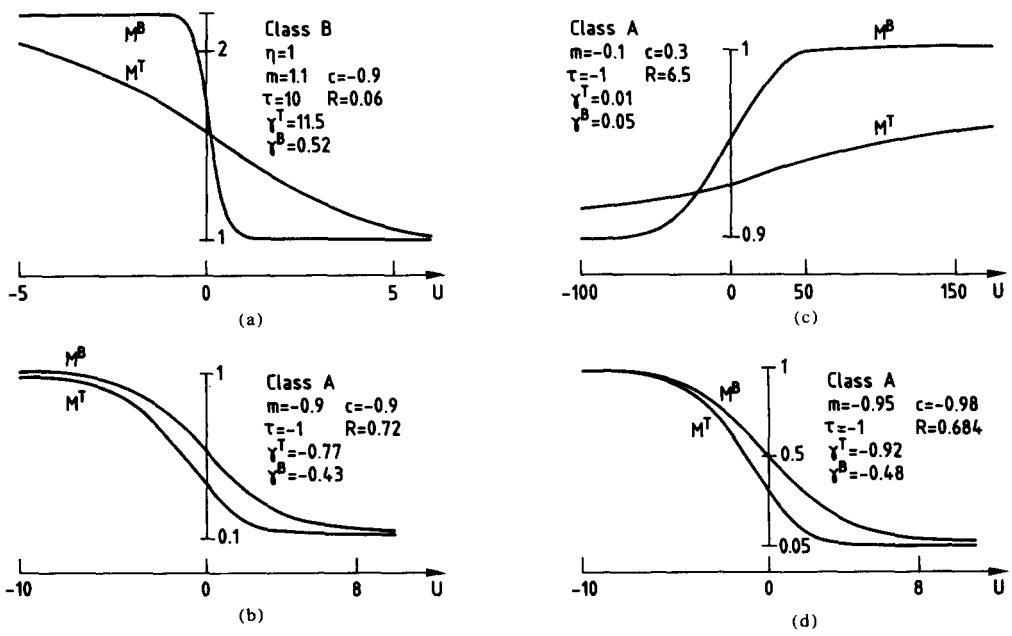


FIG. 1. Similarity shock waves: $m_0=1$. (a)–(c) Broadwell model ($b=2$). (d) Gatignol four-velocity model ($b=1$).

and substitute into (1.5). The collision term has a constant $(1+\tau m_0)(n_{01}^2 - n_{00}n_{02}) = 0$. Depending on whether the first or the second factor is zero, we should have the $1+1$ generalization of the classes A and B of Sec. II. It turns out that class B leads to impossibilities and there remains only the class A of solutions:

$$\tau = -m_0^{-1}. \tag{3.2}$$

If the two jth components are complex conjugate we define $n_i = n_{1i} = n_{2i}^*$, $\gamma = \gamma_1 = \gamma_2^*$ ($\gamma = \gamma_R + i\gamma_I = i\gamma_I$ for periodic solutions), $\rho = \rho_1 = \rho_2^*$, $m = m_1 = m_2^*$ and we rewrite (3.1),

$$N_i = n_{0i} + 2\,\mathrm{Re}\, n_i D^{-1/2}, \quad D = 1 + d\exp(\gamma x + \rho t). \tag{3.1'}$$

A. Relations

The ansatz (3.1) has 13 parameters and 12 relations, leaving one arbitrary parameter. It is in fact a scaling parameter with two opposite effective values ∓ 1. We explain the method for the determination of the parameters and note that once the n_{0i} are known, τ has a fixed value.

First, we have five relations among the six n_{ji},

$$m_1 = \zeta m_2, \quad \zeta = \mp 1, \quad bn_{j1}(n_{j0} + n_{j2}) = -2n_{j0}n_{j2},$$
$$n_{j1}^2 - n_{j0}n_{j2} = \zeta(n_{10}n_{22} + n_{20}n_{12} - 2n_{11}n_{21})$$
$$= n_{00}n_{02} - n_{01}^2, \quad j = 1,2. \tag{3.3}$$

Besides the last of relations (3.3) mixing the n_{ji} and the n_{0i}, there exist two other relations for the n_{0i}:

$$2n_{01}n_{j1} = n_{02}n_{j0} + n_{00}n_{j2}. \tag{3.4}$$

Second, the linear (3.4) and quadratic (3.3) relations give two possible solutions for the n_{0i} which will be distinguished by a parameter $\eta = \mp 1$. Third, once the n_{ji}, n_{0i} are known, there exist four relations giving the ρ_j, γ_j:

$$b\rho_j + \tau m_j(2n_{j1} + b(n_{j0} + n_{j2})) = 0,$$
$$\gamma_j(n_{j0} + n_{j2}) = \rho_j(n_{j2} - n_{j0}). \tag{3.5}$$

Finally both the scaling parameters ζ and η give eight possible algebraic solutions.

B. Determination of the solutions

Our strategy is the following. First, we solve the n_{ji} equations, leaving one scaling parameter n_{12}; second, we solve those for the n_{0i} (then all n_{ji}, n_{0i} are functions of n_{00}); finally, we solve those for the ρ_j and γ_j. The first step leads to discrete possible values for the two ratios $y_j = n_{j0}/n_{j2}$ and for $\bar{n}_{j1} = n_{j1}/n_{j2} = -2y_j/b(1+y_j)$. The relations (3.3) give two coupled polynomials in y_1, y_2 which can be written [(A12)] in terms of $P = y_1y_2$, $S = y_1 + y_2$:

$$A_{p2}P^2 + A_{p1}P + A_{p0} = 0, \quad p = 1,2, \tag{3.6}$$

with A_{pq} polynomials in S. Equation (3.6) has a simple solution:

$$P = 1, \quad S = S^{\mp} = (-3 \mp \sqrt{1 + 48b^{-2}})/2,$$
$$y_1 \text{ or } y_2 = y^{\mp} = (S \mp \sqrt{S^2 - 4}). \tag{3.7}$$

For $b \geqslant 1$, S_- gives real y_j while S_+ for $b > 1$ leads to complex conjugate y_j (in fact $\gamma = i\gamma_I, \rho = \rho_R$ or periodic nonpropagating solutions). Once the y_j are fixed in (3.7), then all n_{ji} are functions of n_{12} with two classes of solutions $n_{22} = y_1\zeta n_{12}$, $\zeta = \mp 1$, $n_{j0} = y_j n_{j2}$, $n_{j1} = \bar{n}_{j1}n_{j2}$. In the second step we find $n_{02} = n_{00}$, $n_{01} = -b(S+2)n_{00}/4$, and $(n_{12}/n_{00})^2$ is a known y_j-dependent number. We must distinguish between the two cases, n_{12} real or complex. For instance, $n_{12} = \eta n_{00} b(1+y_1)/4y_1$ with two possible solutions $\eta = \mp 1$ for n_{12} real. Finally, all n_{ji} and n_{0i} are functions of the scaling parameter n_{00} and of discrete values for the y_j, ζ, and η. From (3.5) we see that this property holds also for γ and ρ. Among these solutions, we must select the physically relevant ones, which can be found in Tables I and II.

TABLE I. Sums of two real similarity shock waves.

$N_0/n_{00} = 1 + b(1+y)(D_1^{-1/2} + y^{-1}D_2^{-1/2})/4$

$N_2 = N_0(y \to 1/y, D_j \to D_j) \quad n_{00} > 0 \quad y^{\mp} = y \quad y^+y^- = 1$

$y^+ + y^- = S \quad D_1 = 1 + d_1\exp(\gamma x + \rho t)$

$D_2 = 1 + d_2\exp(-\gamma x + \rho t) \quad \gamma(1+y) = \rho(1-y)$

Broadwell model: $b = 2 \quad S = S^- = -(3 + \sqrt{13})/2$

$2y^{\mp} = -(3+\sqrt{13})/2 \mp \sqrt{3(1+\sqrt{13})/2}$

$\frac{N_1}{n_{00}} = \frac{\sqrt{13}-1}{4} - \frac{1}{2}\sum D_j^{-1/2}$

$\frac{M}{n_{00}} = 1 + \sqrt{13} - (7+\sqrt{13})\frac{1}{4}\sum D_j^{-1/2} \quad \frac{16\rho}{n_{00}} = 7 + \sqrt{13}$

$\tau(1+\sqrt{13})/n_{00} = -1$

Four velocity planar model: $b = 1 \quad S = S^- = -5$

$2y^{\mp} = -5 \mp \sqrt{21} \quad 7\tau/n_{00} = -2$

$\frac{N_1}{n_{00}} = \frac{3}{4} - \frac{1}{2}\sum D_j^{-1/2} \quad \frac{M}{n_{00}} = \frac{7}{2} - \frac{7}{4}\sum D_j^{-1/2} \quad \frac{8\rho}{n_{00}} = 7$

C. Solution sums of two real similarity shock waves (Table I)

We note that $1 + \tau M$ proportional to a positive constant multiplied by $\Sigma D_j^{-1/2}$ is positive. For the $N_i \geqslant 0$, we begin with the Broadwell model $b = 2$ and look at $|x| \to \infty$. The limits for (N_0, N_2) are either $(3+y)/2 > 0$ or $(3y+1)/2 > 0$, while those for N_1 are $(\sqrt{13}-3)/4 > 0$ with $y = y^{\mp}$ and $n_{00} = 1$. Then for the positivity at fixed x values and $t = 0$, we can always manage the d_j constants in D_j.[3] We notice the invariance: if $y^{\mp} \leftrightarrow y^{\pm}$ and $d_1 \leftrightarrow d_2$ then $N_0 \leftrightarrow N_2$, $N_1 \leftrightarrow N_1$, $M \leftrightarrow M$. In Fig. 2 we present the solution for $y = y^+$. We note that for the total mass the shock limits are the same; there are no discontinuities when $|x| \to \infty$. This corresponds to an infinitely small shock.

TABLE II. Periodic solutions for the Broadwell model $b = 2$.

$S = S^+ = (-3 + \sqrt{13})/2 \quad y = y_R + iy_I \quad |y| = 1 \quad 2y_R = S$

$2y_I = \sqrt{3(\sqrt{13}-1)/2} \quad n_{00} < 0 \quad \tau(\sqrt{13}-1) = n_{00}$

$-N/n_{00} = -1 + \mathrm{Re}[(1+\sqrt{13})/4 + iy_I]D^{-1/2}$

$N_2 = N_0(y_I \to -y_I, D \to D) \quad -N_1/n_{00} = (1+\sqrt{13})/4 - \mathrm{Re}\, D^{-1/2}$

$-M/n_{00} = \sqrt{13} - 1 + (\sqrt{13}-7)\mathrm{Re}\, D^{-1/2}/2$

$D = 1 + d\exp(7 - \sqrt{13})(t - ix4y_I/(1+\sqrt{13}))(n_{00}/16)$

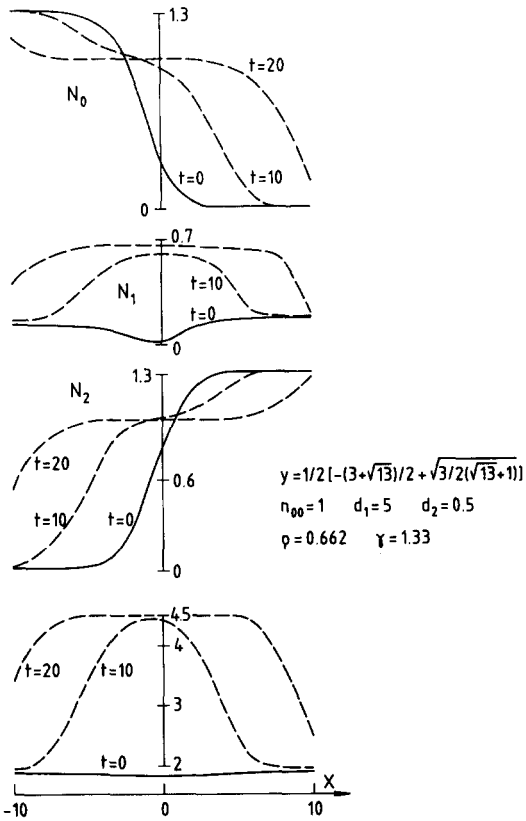


FIG. 2. Broadwell model.

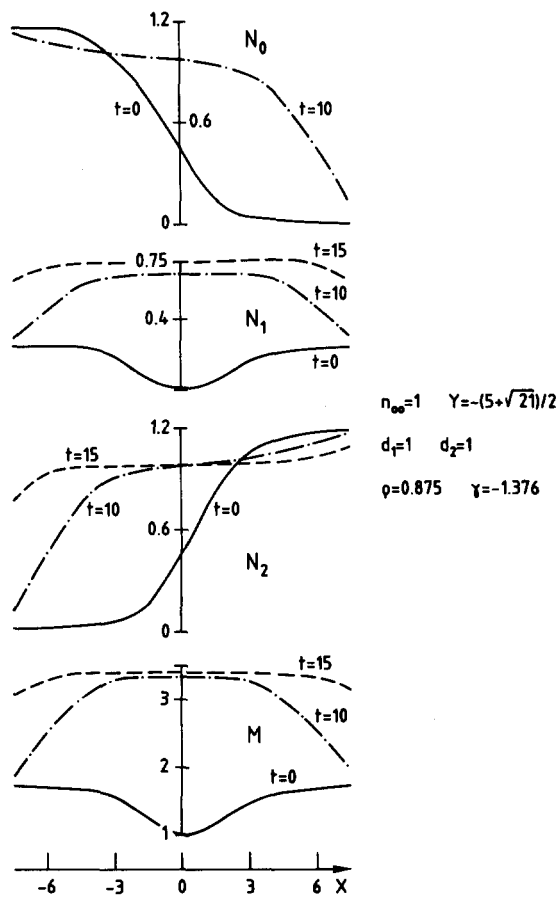


FIG. 3. Velocity planar model.

We go on with the planar four-velocity model with $b=1$ and we can check that the asymptotic x positivity for the N_i are satisfied. In Fig. 3 we present the exact solution for $y=y^+$ and remark that still the two shock limits are the same.

D. Periodic nonpropagating solution for the Broadwell model $b=2$ (Table II)

Choosing $n_{00}<0$, then $\rho_R=\rho<0$. We can check that $n_{0i}+2\,\mathrm{Re}\,n_i$ and $-2\,\mathrm{Re}\,m/m_0$, which are the $t\to\infty$ limits of, respectively, N_i and $1+\tau M$, are positive. Then we can always choose[3] the complex d parameter in D so that $N_i\geqslant 0$ and $1+\tau M\geqslant 0$ will be satisfied for all x values and all $t\geqslant 0$. In Fig. 4 we present the relaxation curves for this solution periodic in x, damped but nonpropagating in t.

IV. SIX-VELOCITY PLANAR MODEL WITH FOUR DIFFERENT N_i

We sketch some results for the model (1.7). With the nonlinear differential equation satisfied by the similarity M solutions being (Gatignol[2]) the same as (2.2) for three different N_i, the macroscopic results are the same as in Sec. II. If $\tau>0$ we find for the ratio of the two shock thickness $R=w^T/w^B<1$ while if $\tau<0$ we cannot draw that conclusion. At both macroscopic and microscopic levels we must distinguish the two classes $1+\tau m_0=0$ or $\neq 0$, with M^B, M^T, N_i^B, N_i^T of the type (2.7). Physically acceptable similarity exact shock-wave solutions exist.

In Appendix B we determine a class of solutions (see

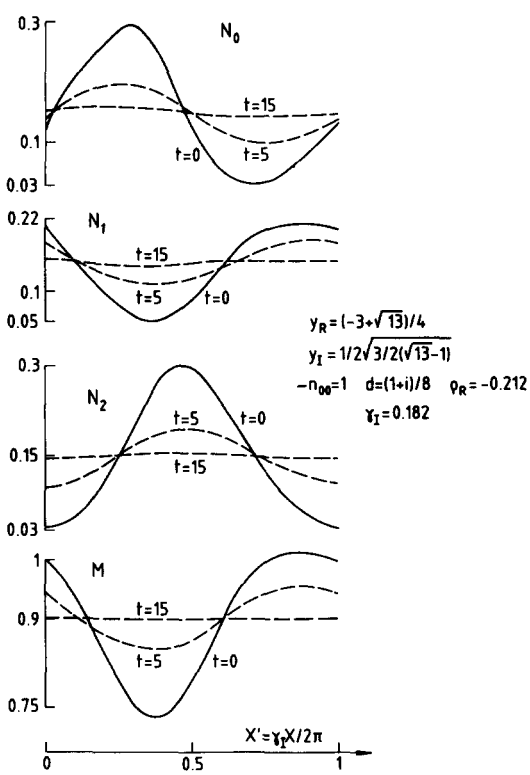


FIG. 4. Broadwell model.

TABLE III. Periodic solutions for the hexagonal six-velocity model.

$$n_{00} = -1 \quad -1.26 < n_{03} < -0.79 \quad y_R = (-17 + \sqrt{265})/12$$
$$y_I^2 = 1 - y_R^2 \quad z = (5 + 3y_R)/8$$
$$n_{3R}^2/(-4z^3) = n_{03}^2 + n_{00}^2 + n_{00}n_{03}(1 - 10z^2)/3z^2$$
$$2m_0 = -3(1 + y_R)(n_{00} + n_{03}) \quad m_0\rho_R = -(1 - y_R^2)(3n_{3R}/4z)^2$$
$$N_0 = n_{00} + 2n_{3R}\,\mathrm{Re}\{1 + iy_I/(+y_R)\}D^{-1/2}$$
$$N_1 = z(n_{00} - 3n_{03}) - z^{-1}n_{3R}\,\mathrm{Re}(1 + iy_I/\{2(1 + y_R)\})D^{-1/2}$$
$$N_3 = N_0(n_{00} \to n_{03}, y_I \to -y_I, D \to D)$$
$$N_2 = N_1(n_{00} \leftrightarrow n_{03}, y_I \to -y_I, D \to D)$$
$$M = m_0 - (3/2z)n_{3R}(1 - y_R)\mathrm{Re}\,D^{-1/2}$$
$$D = 1 + d\exp\rho_R(t - iy_I x/(1 + y_R))$$

Table III) periodic in the space variable and nonpropagating when the time is growing. As in Sec. III we find solutions only for $\tau m_0 = -1$ with $M = m_0 + 2\,\mathrm{Re}\,m/D$, $D = 1 + d\exp(\rho_R t + i\gamma_I x)$. These solutions depend on two arbitrary parameters n_{00}, n_{03} and for $n_{00} = -1$, $-1.26 < n_{03} < -0.79$ we find physically acceptable solutions ($N_i \geqslant 0$, $1 + \tau M \geqslant 0$) with $m_0 > 0$ and $\rho_R < 0$. For $t \to \infty$ we can check that the limits $n_{0i} + 2n_{iR}$, $-2m/m_0$ for both N_i and $1 + \tau M$ are positive. When the asymptotic $t \to \infty$ positivity is satisfied we can manage the complex d parameter in D such that $N_i \geqslant 0$, $1 + \tau M \geqslant 0$ for $t \geqslant 0$. We report in Fig. 5 a numerical example for $n_{00} = -1$, $n_{03} = -1.1$, $d = 0.1(1 + i)$, the other parameter values being obtained from Table III, for instance, $\rho_R = -0.2$, $\gamma_I = 0.22$.

V. CONCLUSION

This paper presents the first determination of exact solutions for discrete models when ternary collisions are included. We find fewer 1 + 1 exact solutions than for the binary collisions alone, the reason being that the presence of both quadratic and cubic nonlinearities require more relations to be satisfied while the number of parameters is the same.

For the present exact solutions, two classes occur depending on whether $\tau m_0 = -1$ or not. While the second class is close to the exact binary solutions (but the limit $\tau \to 0$ is singular), the first class is really new, being due to the presence of ternary collisions. It is clear that going from quadratic to cubic, quartic, etc., nonlinearities some solutions have links while other entirely new solutions are due to the introduction of higher nonlinearity.

ACKNOWLEDGMENT

I thank Professor Gatignol who suggested this work and Dr. Hermann for an interesting question.

APPENDIX A: SOLUTION SUM OF TWO SIMILARITY WAVES FOR THREE DIFFERENT N_i

We seek solutions

$$N_i = n_{0i} + \sum_{j=1,2} n_{ji} D_j^{-1/2}, \quad D_j = 1 + d_j \exp(\gamma_j x + \rho_j t), \tag{A1}$$
$$M = m_0 + \sum_{j=1,2} m_j D_j^{-1/2}, \quad m_0 = n_{00} + n_{02} + 2bn_{01},$$
$$m_j = n_{j0} + n_{j2} + 2bn_{j1},$$

of the nonlinear PDE

$$N_{0t} + N_{0x} = N_{2t} - N_{2x}$$
$$= -bN_{1t} = (1 + \tau M)(N_1^2 - N_0 N_2), \tag{A2}$$

where we assume $\tau = -m_0^{-1}$. We cover both cases where the two jth components in (A1) are real and where they are complex conjugate. In the last case we put $n_i = n_{1i} = n_{2i}^*$, $m = m_1 = m_2^*$, $\rho = \rho_1 = \rho_2^*$, $\gamma = \gamma_1 = \gamma_2^*$.

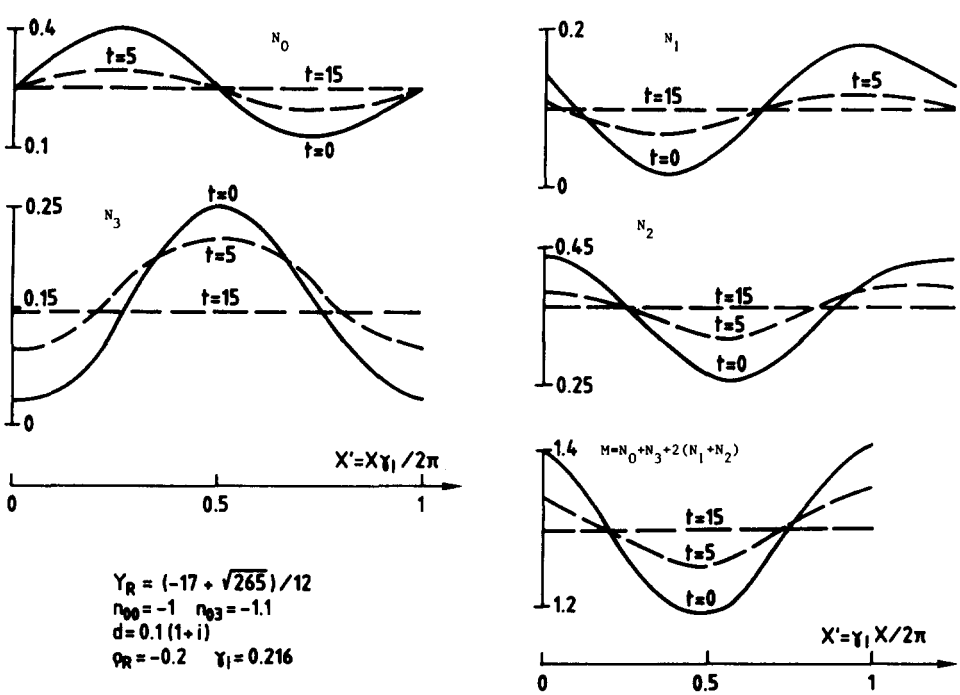


FIG. 5. Six-velocity planar model.

1. Relations

We substitute (A1) into (A2). The terms D_j^{-1}, $D_1^{-1/2}D_2^{-1}$, and $D_j^{-1}D_k^{-1/2}$ are present in the nonlinear part and not in the linear one:

$$m_j(2n_{01}n_{j1} - n_{02}n_{j0} - n_{00}n_{j2}) = 0\,, \tag{A3}$$

$$\sum_j m_k(2n_{01}n_{j1} - n_{02}n_{j0} - n_{00}n_{j2}) = 0, \quad j \neq k\,,$$

$$n_{j1}^2 - n_{j0}n_{j2} + m_j(2n_{11}n_{21} - n_{10}n_{22} - n_{20}n_{12})/m_k = 0\,. \tag{A4}$$

The first terms represent jth similarity component relations while the others are compatibility relations between them for a sum solution. The terms $D_j^{-1/2}$ and $D_j^{-3/2}$, which are other jth component relations, are present in both linear and nonlinear parts:

$$\begin{aligned} n_{j0}(\rho_j + \gamma_j) &= n_{j2}(\rho_j - \gamma_j) \\ &= -bn_{j1}\rho_j = 2\tau m_j(n_{j1}^2 - n_{j0}n_{j2}) \\ &= 2\tau m_j(n_{02}n_{00} - n_{01}^2)\,. \end{aligned} \tag{A5}$$

We have 13 relations; however, $m_j = 0$ for $j = 1$ or 2 is not possible; otherwise (A5) becoming zero for that component leads to an impossibility. Consequently

$$2n_{01}n_{j1} = n_{02}n_{j0} + n_{00}n_{j2} \tag{A3'}$$

in the first two (A3) relations while the third one is identically zero. We have 12 relations for 13 $n_{0i}, n_{ji}, \gamma_j, \rho_j$ parameters leaving only one arbitrary scaling parameter. Later on, we choose n_{00} to be the scaling parameter so that at the final stage all n_{0i}/n_{00}, n_{ji}/n_{00}, ρ_j/n_{00}, γ_j/n_{00} ratios will only depend on b, and for the physical $b = 1$ and $b = 2$ models, these ratios will be fixed numbers.

From the last (A5) relation we see that $n_{ji}^2 - n_{j0}n_{j2} = n_{02}n_{00} - n_{01}^2$ is independent of the two $j = 1,2$ values and we can rewrite (A4):

$$m_k(n_{02}n_{00} - n_{01}^2) + m_j(2n_{11}n_{21} - n_{10}n_{21} - n_{20}n_{12}) = 0\,, \quad k \neq j\,. \tag{A4'}$$

We cannot have $n_{02}n_{00} - n_{01}^2 = 0$ or all the (A5) relations equal to zero and we must have $m_1^2 = m_2^2$:

$$m_1 = \zeta m_2, \quad \zeta = \pm 1 \tag{A5'}$$

($m = \zeta m^*$ for complex conjugate components),

$$\begin{aligned} n_{j1}^2 - n_{j0}n_{j2} &= \zeta(n_{10}n_{22} + n_{20}n_{12} - 2n_{11}n_{21}) \\ &= n_{00}n_{02} - n_{01}^2, \quad j = 1,2\,. \end{aligned} \tag{A6}$$

The elimination of ρ_j, γ_j into (A5) leads to

$$bn_j(n_{j0} + n_{j2}) = -2n_{j0}n_{j2} \tag{A7}$$

and the determination of the ρ_j, γ_j from the n_{ji}

$$\begin{aligned} &b\rho_j + \tau m_j(2n_{j1} + b(n_{j0} + n_{j2})) = 0, \\ &\gamma_j(n_{j2} + n_{j0}) = \rho_j(n_{j2} - n_{j0})\,. \end{aligned} \tag{A8}$$

Our strategy for the determination of the parameters as functions of n_{00} is the following. First, in (A5)–(A7) we have five relations among the six n_{ji}, leaving one scaling parameter chosen to be n_{12} and two possible $\zeta = \pm 1$ values. Second, (A3') and (A6) give two linear and one quadratic relation for the n_{0i}, so that the ratio $(n_{12}/n_{00})^2$ is only b dependent, leaving two possible solutions $n_{12} = \eta n_{00}(\cdots)$ distinguished by two values $\eta = \pm 1$. Third, (A8) gives γ_j, ρ_j as n_{ji} (or n_{00}) dependent. Finally, the n_{00} scaling parameter can be either positive or negative, so that for b fixed we shall find eight different possible types of solutions with $\zeta = \pm 1, \eta = \pm 1, n_{00} = \pm 1$.

2. Determination of the n_{ji} as a function of n_{12}

We introduce intermediate parameters and use (A7):

$$y_j = n_{j0}/n_{j2}, \quad \bar{n}_{j1} = n_{j1}/n_{j2} = -2y_j/b(1 + y_j)\,. \tag{A9}$$

As we shall see the y_j are fixed numbers when b is fixed. From the definition of m_j and (A3') we find a first expression for the ratio $n_{12}/\zeta n_{22}$:

$$n_{12}/\zeta n_{22} = (1 + y_2 + 2b\bar{n}_{21})/(1 + y_1 + 2b\bar{n}_{11})\,. \tag{A10}$$

From (A4'), (A5'), and (A6) we obtain

$$\zeta(2n_{11}n_{21} - n_{10}n_{22} - n_{20}n_{12}) = n_{j1}^2 - n_{j0}n_{j2}, \quad j = 1,2,$$

and two other expressions for the above ratio:

$$\begin{aligned} n_{12}/\zeta n_{22} &= (y_1 + y_2 - 2\bar{n}_{11}\bar{n}_{21})/(n_{11}^{-2} - y_1) \\ &= (n_{21}^{-2} - y_2)/(y_1 + y_2 - 2\bar{n}_{11}\bar{n}_{21})\,. \end{aligned} \tag{A10'}$$

In (A10) and (A10') we have two independent relations for the y_j. Introducing the sum $S = y_1 + y_2$ and the product $P = y_1y_2$ we find

$$\begin{aligned} &(1 - y_j)^2 y_k P_2(y_k) + X(1 - y_k)^2 = 0\,, \quad k \neq j\,, \\ &P_2(y_k) = (1 + y_k)^2 - 4/by_k\,, \\ &X = S(P + 1) + S^2 - 8P/b^2\,. \end{aligned} \tag{A11}$$

Equations (A11) give two second-order polynomials in P with S-dependent polynomial coefficients:

$$\begin{aligned} &A_{2p}P^2 + A_{1p}P + A_{0p} = 0\,, \quad p = 1,2\,, \quad A_{21} = 1\,, \\ &A_{11} = S^2 - 4S - 6 + (24 - 8S)b^{-2}\,, \\ &A_{01} = S^3 + 1 - 4Sb^{-2}\,, \\ &A_{22} = -2S^2 + 6S + (20S - 24)b^{-2} \\ &\qquad + 4(1 - 6b^{-2})(1 + b^{-2})\,, \\ &A_{12} = 3S^3 - S^2 + 2S + 16S(S + 1)b^{-2}\,, \\ &A_{02} = -S^2(S + 1)^2\,. \end{aligned} \tag{A12}$$

We have found solutions for $P \neq 1$ with $(S,P) = (2.79,4.79)$, $(0.58,0.208)$, $(-1.79,0.208)$, $(-8.58,4.79)$, corresponding either to real or complex conjugate components. However, we have not found, for these values, solutions sharing both $N_i \geqslant 0$ and $1 + \tau M \geqslant 0$. In the following we assume $P = y_1y_2 = 1$ for which the two (A12) polynomials have a common factor:

$$\begin{aligned} &(S - 2)Q(S) = 0\,, \\ &(S^2 + 2S - 4 - 4b^{-2})Q(S) = 0\,, \\ &Q = S^2 + 3S + 2(1 - 6b^{-2})\,. \end{aligned} \tag{A13}$$

Here $Q = 0$ leads to two $S_\pm$ values for S and with $P = 1$ we determine the y_j:

$$\begin{aligned} &2S_\pm = -3 \mp B, \quad (y_1 \text{ or } y_2) = (S \mp \sqrt{S^2 - 4})/2, \\ &S = S_+ \text{ or } S_-, \quad B = \sqrt{1 + 48b^{-2}}\,. \end{aligned} \tag{A13'}$$

Let us consider b as a continuous parameter $b \geqslant 1$. Due to

$3(S^2-4)/2 = 8b^{-2} - B$ we see that the y_j are complex conjugates for $S = S_+$, $b \geqslant 1$ and real for $S = S_-$, $b > 1$. For $b = 2$ we have both complex conjugate and real y_j roots, while for $b = 1$ the acceptable y_j are real. Once the y_j are known numbers, we recover the original parameters $n_{j0} = y_j n_{j2}$, $n_{j1} = \bar{n}_{j1} n_{j2}$ as functions of two parameters n_{12} and n_{22}; however, if in (A10) we substitute for the $\bar{n}_{j1}$ their y_j dependence, we get a relation between n_{12} *and* n_{22}:

$$n_{12}/n_{22}\zeta = (1+y_1)(1-y_2)^2/(1+y_2)(1-y_1)^2 = 1/y_1 . \tag{A14}$$

Finally all the n_{ij} depend only on the scaling parameter n_{12}:

$$\zeta n_{21} = n_{11} = -2y_1 n_{12}/b(1+y_1), \quad n_{20} = \zeta n_{10}/y_1 = \zeta n_{12} . \tag{A15}$$

3. Determination of n_{0i}, n_{12} as functions of n_{00}

We rewrite (A3′) and (A6) in terms of the known intermediate parameters:

$$n_{00} + n_{02} y_j = 2n_{01}\bar{n}_{j1}, \quad n_{01}^2 - n_{00}n_{02} = n_{12}^2 (y_1 - n_{11}^{-2}) . \tag{A16}$$

Writing the first relation both for y_1 and y_2, using $y_1 y_2 = 1$, we get $(n_{00} - n_{02})(1 - y_j) = 0$. Then with $n_{00} = n_{02}$ in the lhs of this relation we find n_{01}. Finally the second relation determines n_{12}^2:

$$\begin{aligned} &n_{02} = n_{00}, \quad n_{01} = -b(S+2)n_{00}/4, \\ &(n_{12}/n_{00})^2 = (b(1+y_1)/4y_1)^2 , \end{aligned} \tag{A17}$$

where we have used the identities

$$\begin{aligned} &(1+y_j)^2 = (2+S)y_j, \\ &(b(S+2)/4)^2 - 1 = (b^2(S+2) - 4)/16 . \end{aligned}$$

The last of relations (A17) is the crucial one because either n_{12}, y_1 are complex or they are real and we must distinguish between these two cases.

4. Sums of two real similarity solutions ($b \geqslant 1$)

In this case, y_j, n_{ji} are real, we choose $S = S_-$, and the last real relation (A17) gives two solutions

$$n_{12} = \eta n_{00} b(1+y_1)/4y_1, \quad \eta = \mp 1 . \tag{A18}$$

With (A15), (A17), and (A18), all n_{ji}, n_{0i} have n_{00} as a scaling factor. Similarly m_0, m_i, which are linear combinations of these parameters, factorize n_{00} also and we write the four types of solutions in closed form:

$$\begin{aligned} &N_0/n_{00} = 1 + b(1+y)(D_1^{-1/2} + y^{-1} D_2^{-1/2}\zeta)\eta/4, \\ &N_2/n_{00} = 1 + b(1+y)(y^{-1} D_1^{-1/2} + D_2^{-1/2}\zeta)\eta/4 , \\ &N_1/n_{00} = b(B-1)/8 - \eta(D_1^{-1/2} + \zeta D_2^{-1/2})/2, \\ &4\tau/n_{00} = -8 - b^2(B-1) < 0 , \\ &M/n_{00} = 2 + b^2(-1+B)/4 - \eta b(7+B) \\ &\qquad \times (D_1^{-1/2} + \zeta D_2^{-1/2})/8 , \end{aligned} \tag{A19}$$

where y is any one of the two y_j defined in (A13′) for $S = S_-$ and B is given in (A13′). It remains to find γ_j, ρ_j, which in (A8) are deduced from the n_{ji}, m_j of (A19),

$$\begin{aligned} &\rho_1 = \rho_2 = \eta(7+B)n_{00}/16, \\ &\gamma_1 = -\gamma_2 = \rho_1(1-y)/(1+y) . \end{aligned} \tag{A20}$$

For physical solutions we must have $1 + \tau M > 0$ with

$$1 + \tau M = b(7+B)\eta(D_1^{-1/2} + D_2^{-1/2}\zeta)/2(8 + b^2(B-1)) . \tag{A20′}$$

Since the sign is given by $\eta(D_1^{-1/2} + D_2^{-1/2}\zeta)$, looking at the limits $|x| \to \infty$ we see that necessarily $\zeta = \eta = 1$. Coming back to (A20) with $\eta = 1$, $n_{00} > 0$, we see that $\rho_j > 0$ and the limits $N_i = n_{0i}$ are positive when $t \to \infty$. We do not discuss the positivity conditions anymore and for the physical case $b = 1$ and 2 quoted in Table I and discussed in Sec. III C, we require $n_{00} > 0$, $\eta = \zeta = 1$ (and appropriate values for the d_j in D_j). Finally we notice that while N_1/n_{00} and M/n_{00} have fixed numbers for coefficients, on the contrary N_0/n_{00} and N_2/n_{00} depend on the choice of the two y numbers: either y_1 or y_2. However, due to $y_1 y_2 = 1$, we notice the invariance: if $y \leftrightarrow 1/y$, then $D_1 \leftrightarrow D_2$, $\gamma_1 \leftrightarrow \gamma_2$, $N_0 \leftrightarrow N_0$, $N_2 \leftrightarrow N_2$ if we choose $d_1 = d_2$ in D_j.

5. Sums of two complex conjugate similarity solutions ($b > 1$)

For the intermediate parameters we have $y_2 = y_1^*$, $\bar{n}_{2i} = \bar{n}_{1i}^*$. We define $n_i = n_{iR} + i n_{iI} = n_{1i}$, $m = m_R + i m_I$, $y = y_R + i y_I = y_1$ $(y_R = S_+/2)$, $\rho = \rho_R + i\rho_I$, $\gamma = \gamma_R + i\gamma_I$, while the other parameters remain real. Recalling $P = 1 = |y|^2$, we choose $S = S_+$ in (A13′) and (A5′) becomes $m_1 = m = \zeta m^*$. We restrict our study to $\zeta = 1$, which means $m = m_R$, $m_I = 0$. We rewrite the expressions (A13)–(A17), where the n_i are now functions of n_{2R}, the n_{0i} functions of n_{00}, and a quadratic relation $(n_{2R}/n_{00})^2$ still allows us to express the solutions in terms of the scaling parameter n_{00}:

$$\begin{aligned} &n_2 = y^* n_1^*, \quad n_0^* = n_2, \quad n_1^* = n_1 = -2y n_2/b(1+y), \\ &n_{0R} = n_{2R}, \quad n_{1R} = -2n_{2R}/b(1+y_R) , \quad n_{1I} = 0 , \\ &n_{0I} = -n_{2I} = y_I n_{2R}/(1+y_R), \quad y_I^2 + y_R^2 = 1, \\ &n_{00} = n_{02}, \quad n_{01} = -b(1+y_R)n_{00}/2, \\ &n_2^2 = (n_{00} b(1+y)/4y)^2, \quad n_{2R} = \eta n_{00}(1+y_R)b/4 . \end{aligned} \tag{A21}$$

We still obtain two possible solutions with $\eta = \pm 1$. We notice that $m_I = n_{1I} = n_{0I} + n_{2I} = n_{0R} - n_{2R} = 0$ and, inserting these results into relations (A8) for the determination of ρ, γ, we find

$$\begin{aligned} &b\rho + 2\tau m_R(n_{1R} + b n_{2R}), \quad \rho_I = 0, \\ &\gamma = i\rho_R n_{2I}/n_{2R}, \quad \gamma_R = 0 . \end{aligned} \tag{A8′}$$

It follows that the solutions are periodic $\gamma_R = 0$ in the space variable, but nonpropagating $\rho_I = 0$ with the time and they are written with the substitution $16 n_{2R} = \eta b(1+B) n_{00}$:

$$\begin{aligned} &N_0/n_{00} = 1 + \eta\,\mathrm{Re}((1+B)/8 + iy_I/2)bD^{-1/2}, \\ &N_2/n_{00} = 1 + \eta\,\mathrm{Re}((1+B)/8 - iy_I/2)bD^{-1/2} , \\ &N_1/n_{00} = -b(1+B)/8 - \eta\,\mathrm{Re}\,D^{-1/2}, \\ &M/n_{00} = 2 - b^2(1+B)/4 + \eta b(-7+B)/4\,\mathrm{Re}\,D^{-1/2}, \\ &B = \sqrt{1 + 48b^{-2}} , \\ &\rho_R/n_{00} = (B-7)(-4 + b + bB)/8(8 - b^2 - b^2 B), \\ &\gamma_I = -4\rho_R y_I/(1+B) , \\ &D = 1 + (d_R + i d_I)\exp(\rho_R t + i\gamma_I x) . \end{aligned} \tag{A22}$$

For the physical $b=2$ value, $1+\tau M=\mathrm{Re}\,D^{-1/2}(-\eta)\times(7-\sqrt{13})/2(\sqrt{13}-1)$ having the $-\eta$ sign, we choose $\eta=-1$. Always for $b=2$, let us choose $n_{00}>0$; then $\rho_R=(7-\sqrt{13})n_{00}/16>0$, $\lim D^{-1/2}=0$, $\lim N_1=n_{01}=-(1+\sqrt{13})n_{00}/4<0$ when $t\to\infty$, and the positivity property is violated. It follows that we must choose $\eta=-1$, $n_{00}<0$ for the Broadwell $b=2$ model for which $\tau=1/n_{00}(\sqrt{13}-1)<0$. We still notice in (A21) that N_1/n_{00} and M/n_{00} have fixed number coefficients while N_0/n_{00} and N_2/n_{00} depend on the choice of y_I or y. However, when $y\leftrightarrow y^*$, we find $D\leftrightarrow D^*$, $N_0\leftrightarrow N_0$, $N_2\leftrightarrow N_2$, if we choose $d=d_R$, $d_I=0$.

6. Specular reflection at a wall $x=0$

It is worth noticing that these solutions have a physical meaning for specular reflection at a wall $x=0$. The specular reflection condition for the densities N_0,N_2 associated with the two opposite velocities perpendicular to the wall is

$$N_0(x=0,t)-N_2(x=0,t)=0\,. \tag{A23}$$

First, for the sums of two real similarity solutions (A19), let us choose $d_1=d_2=d$ in D_j; then we find

$$D_1(x=0,t)=D_2(x=0,t)=1+d\exp(\rho t)$$

and (A23) is satisfied. Second, for the periodic (A22) solutions, let us choose $d=d_R$, $d_I=0$ in D; then $D(x=0,t)=1+d_R\exp(\rho_R t)$ is real and we find

$$N_0(x=0,t)-N_2(x=0,t)=(D(x=0,t))^{-1/2}\,\mathrm{Re}(n_0+n_0^*-n_2-n_2^*)=0\,.$$

APPENDIX B: SOLUTION SUM OF TWO COMPLEX CONJUGATE SIMILARITY WAVES FOR FOUR DIFFERENT N_i

We seek solutions

$$\begin{aligned}&N_i=n_{0i}+2\,\mathrm{Re}\,n_i D^{-1/2},\\ &\quad D=1+d\exp(\gamma x+\rho t),\\ &M=m_0+2\,\mathrm{Re}\,mD^{-1/2},\\ &\quad m_0=n_{00}+n_{03}+2(n_{01}+n_{02}),\\ &\quad m=n_0+n_3+2(n_1+n_2),\end{aligned} \tag{B1}$$

n_{0i} real, n_i,γ,ρ complex, of the nonlinear PDE

$$\begin{aligned}N_{0t}+N_{0x}=N_{3t}-N_{3x}&=-2N_{1t}-N_{1x}\\ &=-2N_{2t}+N_{2x}=(1+\tau M)(N_1N_2-N_0N_3)\,,\end{aligned} \tag{B2}$$

where we assume $\tau=-m_0^{-1}$. We seek periodic solutions in space, $\gamma=\gamma_R+i\gamma_I$, $\gamma_R=0$, and nonpropagating in time, $\rho=\rho_R+i\rho_I$, $\rho_I=0$.

1. Relations

We subsitute (B1) into (B2). We assume $m\neq0$ (in fact, $m=0$ leads to impossibilities). The terms proportional to D^{-1}, $|D|^{-1}$, $D^{-1/2}|D|^{-1}$ present only in the collision term give the relations

$$\begin{aligned}&n_{01}n_2+n_{02}n_1=n_{00}n_3-n_{03}n_0,\\ &n_1n_2^*+n_2n_1^*-n_0n_3^*-n_3n_0^*+(n_1n_2-n_0n_3)m/m^*=0\,,\end{aligned} \tag{B3}$$

while $D^{-1/2}$,$D^{-3/2}$ present everywhere lead to

$$\begin{aligned}n_0(\gamma+\rho)=n_3(\rho-\gamma)&=-n_1(2\rho+\gamma)\\ &=n_2(-2\rho+\gamma)\\ &=2\tau m(n_1n_2-n_0n_3)\\ &=2\tau m(n_{00}n_{03}-n_{01}n_{02})\,.\end{aligned} \tag{B4}$$

We have two arbitrary parameters among the n_{0i},n_i,γ,ρ and we choose (n_{00},n_{03}). From the last (B3)–(B4) relations we find m/m^* real and we choose $m=m_R$ real. Similar to Appendix A we write the relations in three different groups. First,

$$\begin{aligned}&m=m^*,\quad -2n_0n_3=n_1(n_0+3n_3)=n_2(n_3+3n_0),\\ &n_1n_2-n_0n_3=n_0n_3^*+n_3n_0^*-n_1n_2^*-n_2n_1^*\\ &\qquad=n_{00}n_{03}-n_{01}n_{02}\,,\end{aligned} \tag{B5}$$

allowing the determination of the n_i from a scaling parameter n_{3R} (we define $n_i=n_{iR}+in_{iI}$). Second, the first (B3) and the last (B5) relations give two linear and one quadratic relation between the n_{oi}, leading to a quadratic equation for n_{3R}. Then n_{3R} as well as all other n_i are functions of the arbitrary n_{00},n_{03}. Finally, from (B4) we obtain ρ,γ from the n_i,n_{0i},

$$\begin{aligned}&2\rho=m(n_1+n_2+2(n_0+n_3))/m_0,\\ &\gamma(n_0+n_3)+\rho(n_0-n_3)=0\,.\end{aligned} \tag{B6}$$

Let us define $y=n_0/n_3$; then from the last relation we get $\gamma_R=0,\rho_I=0$ iff $|y|=1$.

2. n_i functions of n_{3R}

We introduce other intermediate parameters $\bar n_1=n_1/n_3=-2y/(3+y)$, $\bar n_2=n_2/n_3=-2y/(1+3y)$. Equation (B5) gives four indentities:

$$\begin{aligned}&n_3/n_3^*=A^*/A=B/C=C^*/B,\\ &\quad A=1+y+2(\bar n_1+\bar n_2),\\ &\quad B=y+y^*-\bar n_1\bar n_2^*-\bar n_2\bar n_1^*\,,\end{aligned} \tag{B7}$$

and $C=\bar n_1\bar n_2-y$. We have two relations among y and y^*. Seeking solutions such that $|y|=1$, $S=y+y^*$, $S^2<4$, we obtain

$$\begin{aligned}&A^*/A=1/y,\\ &B=3(S+2)(3S^2+14S-8)/(3S+10)^2,\\ &C=-3y(S+2)/(3S+10)\,.\end{aligned} \tag{B7'}$$

Consequently we find $3S^2+17S+2=0$, $6S^{\mp}=-17\mp\sqrt{265}$ with $(S^-)^2>4$, which is excluded. We put $y=y_R+iy_I$ with $y_R^2+y_I^2=1$, $2y_R=S$ and the acceptable solution is for S^+ or

$$y_R=(-17+\sqrt{265})/12=-0.0601<0. \tag{B8}$$

The first (B7) relation gives $n_3y=n_3^*$ and, recalling the definitions of y, $\bar n_1$, $\bar n_2$,

$$\begin{aligned}&n_0=n_3^*,\quad n_1=n_2^*,\quad n_{3I}=-y_In_{3R}/(1+y_R),\\ &n_{1I}=-y_In_{3R}/(1+y_R)4z,\quad n_{1R}=-n_{3R}/2z,\\ &m=m_R=3(y_R-1)n_{3R}/(4z),\quad z=(5+3y_R)/8>0,\end{aligned} \tag{B9}$$

where, for simplicity, we have introduced a new numerical parameter z. All the n_i are expressed as functions of n_{3R}.

3. Relation between n_{3R} and n_{00}, n_{03}

From the two linear relations (B3) and the quadratic relations (B5) we find

$$\begin{aligned} &n_{01} = z(n_{00} - 3n_{03}), \quad n_{02} = z(n_{03} - 3n_{00}), \\ &m_0 = (-3/2)(1 + y_R)(n_{00} + n_{03}), \\ &(n_{03}/n_{00})^2 + (n_{03}/n_{00})(1 - 10z^2)/3z^2 + 1 \\ &\quad = -(n_{3R}/2zn_{00})^2/z . \end{aligned} \tag{B10}$$

Here n_{3R} solutions exist algebraically only if the lhs of the last of relations (B10) is negative or $0.53 < n_{03}/n_{00} < 1.884$. Notice that we have two determinations, $n_{3R} > 0$ and $n_{3R} < 0$.

4. Determination of ρ, γ

We find from (B6)

$$\begin{aligned} &m_0 \rho_R = -(3n_{3R}/4z)^2(1 - y_R^2) < 0, \\ &\gamma_I = -\rho_R y_I/(1 + y_R) . \end{aligned} \tag{B11}$$

5. Physical solutions

We have both the analytical results $m_R n_{3R} < 0$, $m_0(n_{00} + n_{03}) < 0$, $m_0 \rho_R < 0$ and the two physical constraints: when $t \to \infty$, $\lim N_i$ equals either $n_{0i} > 0$ (if $\rho_R > 0$) or $n_{0i} + 2n_{iR} > 0$ (if $\rho_R < 0$) and $1 + \tau M = -(m/m_0) 2 \operatorname{Re} D^{-1/2} > 0$ or $m_0 m < 0$. The only possible choice is $n_{00} + n_{03} < 0$.

Starting with $n_{00} < 0$, $n_{03} < 0$, it follows from the analytical results that $m_0 > 0, \rho_R < 0, m_R < 0$ if we choose the solution $n_{3R} > 0$. Then we find $1 + \tau M \geqslant 0$ and it remains to check the condition $n_{0i} + 2n_{iR} > 0$. Choosing $n_{00} = -1$, numerically we have found physically acceptable solutions for $-1.26 < n_{03} < -0.79$.

[1] J. E. Broadwell, Phys. Fluids **7**, 1243 (1964).

[2] R. Gatignol, *Lecture Notes in Physics*, Vol. 36 (Springer, Berlin, 1975); Colloque CNRS Paris, 1974; S. Harris, Phys. Fluids **9**, 1328 (1966); McKean, Commun. Pure Appl. Math. **28**, 435 (1975); H. Cabannes, Lect. Notes Berkeley, 1980; R. Illner, Math. Methods Appl. Sci. **1**, 187 (1979); T. W. Ruijgrok and T. T. Wu, Physica A **113**, 401 (1982). See S. Wolfram, Rev. Mod. Phys. **55**, 601 (1983); D'Humieres, Lallemand, and Frisch, Eur. Phy. Lett. **2**, 291 (1986).

[3] H. Cornille, J. Phys. A **20**, 1973 (1987); J. Math. Phys. **28**, 157 (1987); J. Stat. Phys. **48**, 789 (1987); C. R. Acad. Sci. **304**, 1091, 1987).

[4] A. Bobylev, Math. Cong. Warsaw, 1983; J. Wick, Math. Methods Appl. Sci. **6**, 515 (1984).

[5] H. Cabannes and D. M. Tiem, C. R. Acad. Sci. **304**, 29 (1987); J. Stat. Phys. (to be published).

[6] T. Platkowski, Bull. Pol. Acad. Sci. **32**, 247 (1984); Mech. Res. Comm. **11**, 201 (1984).

Theory for correlation functions of fermion quantum fluids

Suresh K. Sinha
Department of Physics, Langat Singh College, Bihar University, Muzaffarpur-842001, India

(Received 14 October 1986; accepted for publication 3 February 1988)

A theory for the pair correlation function of a fermion fluid is developed in terms of a function $S^{(m)}$ defined as a certain sum over graphs. The theory is further extended to derive expressions for the radial distribution function and for the four distinct direct correlation functions. These four direct correlation functions are combined together leading to an integral equation for fermion quantum fluids, which is much simpler and is expressed in a computationally convenient form.

I. INTRODUCTION

In recent years, much progress has been made in the theory of the ground state fermion system.[1-3] The most successful approach has been the Feenberg–Jastrow variational method, in which a trial many-body wave function is chosen as[4,5]

$$\psi_F = \psi_B \Phi , \tag{1.1}$$

where

$$\psi_B = \prod_{i<j} \exp\left[\frac{1}{2} u_2(i,j)\right] = \prod_{i<j} [1 + F_{ij}]^{1/2} , \tag{1.2}$$

is a bosonic wave function, and Φ is an antisymmetric function describing a state of the N-particle fermion system in the absence of interaction. For a normal Fermi liquid, Φ is taken to be a Slater determinant constructed from products of plane wave orbital and spin functions,[3,6]

$$\Phi = a\left[\prod_{i=1}^{N} \exp(i\mathbf{k}_i \cdot \mathbf{r}_i)\xi_i(m_i)\right], \tag{1.3}$$

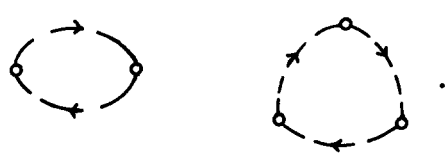
where a is the antisymmetrizer and the ξ_i are spin functions.

Equation (1.1) has been used to develop a cluster expansion of the distribution function in terms of diagrams containing F- and f-bonds, where f is defined by

$$f(k_F r) = 3[\sin(k_F r) - (k_F r)\cos(k_F r)]/(k_F r)^3 , \tag{1.4}$$

where $k_F = (6\pi^2\rho/s)^{1/3}$ is the Fermi momentum at density ρ and the factor s gives the degeneracy of the system ($s = 2$ for the paramagnetic state of ^{3}He or neutron matter, $s = 4$ for nuclear matter).

The graphical theory[7-9] for the correlation functions of fermion fluids described by the function of the form of Eq. (1.1) has been developed. The cluster expansion for the correlation functions is expressed in terms of diagrams involving F- and f-bonds. In this case, a diagram is a collection of circles (vertices or points) and F- and f-bonds connecting some pairs of circles. The f-bonds are directional bonds and always connect circles in such a way that they form a closed loop or ring. For example, f-loops with two and three circles may be constructed by connecting these circles with two and three f-bonds, respectively:



Each loop connecting P circles is associated with a factor $(-s)^{1-P}$. There are two types of circles: black and white. The white circles are labeled, but the black circles are unlabeled. The value of a diagram is defined in terms of F- and f-functions and an integration over the position; which can be assigned to each black circle.

Most of our theory for calculating distribution functions of fermion fluids is, however, obtained under different approximations, by neglecting different sets of diagrams. Of these, the Fermi hypernetted chain (FHNC) approximation has been the most widely used.[8,9] By partial summation in the FHNC method, one tries to take into account the antisymmetry correctly, but this method is not expected to give good results, at least not at high densities.[10] Another method for calculating the distribution function is the Fermi–Born–Green–Yvon (FBGY) approximation.[11] Recently Campbell *et al.*[12] have used the FBGY method to calculate the ground state properties of the fermion quantum liquid and found that this approximation also provides good results only at low densities.

Recently, we[7] have developed a method, which we have called a "modified" FHNC method for calculating the distribution function of the fermion system, in which the series is expressed in terms of the boson distribution functions and the permutation expansions of the antisymmetry are partially summed. This approach differs from the FHNC method in the selection of the subseries of diagrams to be summed in the permutation expansion for the antisymmetry, and in expressing the series in terms of the boson distribution functions.

At high density, a large set of diagrams should be included in the series to give a reasonally accurate description of the fermion system.

In this paper, we develop a theory that includes a large set of allowed diagrams and is computationally convenient. Much of the diagrammatic analysis presented here is based on our previous work[7] on the Fermi system.

In Sec. II, we present the theoretical background for the cluster series of correlation functions in terms of F- and f-bonds. Section III is devoted to reducing the expansion in terms of the pair correlation function h_B (generated by the F-bond) and a new function $S^{(m)}$ defined by Eq. (3.2). The series is divided into five classes of diagrams. The five classes of diagrams are summed and expressed in compact form in Sec. IV. Final expressions for the radial distribution function

 0022-2488/88/071678-09$02.50

and direct correlation functions are reported there under different approximations. Section V is concerned with developing an integral equation for the correlation function of the fermion fluid. Concluding remarks are given in Sec. VI.

II. THEORETICAL BACKGROUND

The spin-averaged pair distribution function for the fermion system is defined as[7]

$$\rho^2 g(1,2) = \Xi^{-1} \sum_{N\geq 2} \frac{1}{(N-2)!} \times \int \cdots \int \prod_{i=1}^{N} z(i)\, W_N(1,2,\ldots,N) \prod_{i=3}^{N} d\mathbf{r}_i , \quad (2.1)$$

where

$$\Xi = \sum_{N\geq 0} \frac{1}{N!} \int \cdots \int \prod_{i=1}^{N} z(i)\, W_N(1,2,\ldots,N) \prod_{i=1}^{N} d\mathbf{r}_i \quad (2.2)$$

is a generating functional for the distribution functions of the fermion system and W_N is the spin-averaged Slater sum

$$W_N = \frac{1}{A} \sum_{\sigma} |\Psi_F|^2 . \quad (2.3)$$

Here Σ_σ indicates the summation over all spin states and the constant A is the norm of the wave function Ψ_F.

With the help of Eqs. (2.1)–(2.3) we obtain the cluster expansion of the pair correlation function $h(1,2)$ [where $h(1,2) = g(1,2) - 1$] in terms of composite diagrams containing F and f bonds. We adopt the standard terminology and notation of Refs. 13 and 14. These diagrams have the following specific features.[7]

(i) The exchange bond f is a directional bond.

(ii) The f-bonds form a closed loop or ring. Each loop with p circles is multiplied by a factor $(-s)^{1-p}$.

(iii) Each circle is attached to at most one f-loop.

(iv) A black circle attached to an f-bond, must be attached to at least one F-bond.

An allowed diagram is a diagram that does not violate these restrictions, along with the usual restrictions adopted for simple classical fluids[13,14] or boson quantum fluids. In terms of these diagrams, we can obtain the cluster expansions of the pair correlation function $h(1,2)$ and the direct correlation function $c(1,2)$ for a Fermi fluid. In the absence of the exchange bond, the problem becomes analogous to one discussed by Stell[14] for classical fluids. We follow the method of Stell[14] for deriving a cluster expansion for the Fermi fluid (in the presence of the exchange bond). Thus the cluster series for the pair correlation function $h(1,2)$ and direct correlation function $c(1,2)$ are given by the following:

$h(1,2)$ = [sum of all distinct connected composite diagrams (CCD) with two white one-circles labeled 1 and 2, respectively, some or no black ρ-circles, some or no F-bonds, some or no f-bonds forming a closed loop, at least one F-bond and/or f-loop and no articulation circles], (2.4)

$c(1,2)$ = [the sum of all nodeless diagrams in h]. (2.5)

We have adopted the diagrammatic notation of Refs. 13 and 14. Thus an articulation circle is a circle in a connected diagram such that its removal divides the diagram into two or more components of which at least one component contains no white circles. The first few diagrams in the cluster series for the correlation functions $h(1,2)$ and $c(1,2)$ are shown in Figs. 1 and 2, respectively.

Equations (2.4) and (2.5) thus give a graph theoretic recipe for the pair correlation function and direct correlation function.

III. TOPOLOGICAL REDUCTION

In this section we use the topological reduction technique and eliminate F-bonds from Eq. (2.4) in favor of h_B-bonds [where $h_B(r)$ is the pair correlation function generated by F-bonds only]. Thus we obtain

$$h(1,2) = h_B(1,2) + H(1,2), \quad (3.1)$$

where $H(1,2)$ is the sum of all allowed CCD with two white one-circles labeled 1 and 2, respectively, any number of black ρ-circles, at least one f-loop, some or no h_B-bonds, no articulation circles, no articulation f-loop, and no articulation h_B-bonds.

The expansion series of $H(1,2)$ includes (a) diagrams with a single loop of f-bonds connecting all the circles associated with a diagram and (b) diagrams with one or more loops of f-bonds, connecting only some of the circles of a diagram. It is clear from the definition of $H(1,2)$ that the circles associated with an f-loop may be connected by some or no h_B-bond, such that each black circle must be conected with at least one h_B-bond. Further an f-loop may be connected with another f-loop or free circle by at least one h_B-bond. By free circle we mean a circle not associated with a loop.

We first consider a set of diagrams of type (a) with two

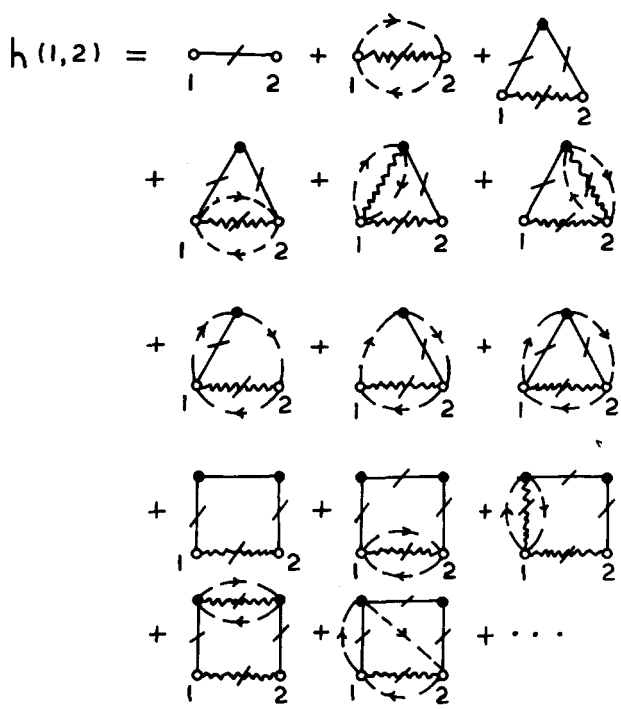


FIG. 1. The first few diagrams of $h(1,2)$ for a fermion fluid. The dotted line represents the F-bond and the dashed line with an arrow denotes the f-bond.

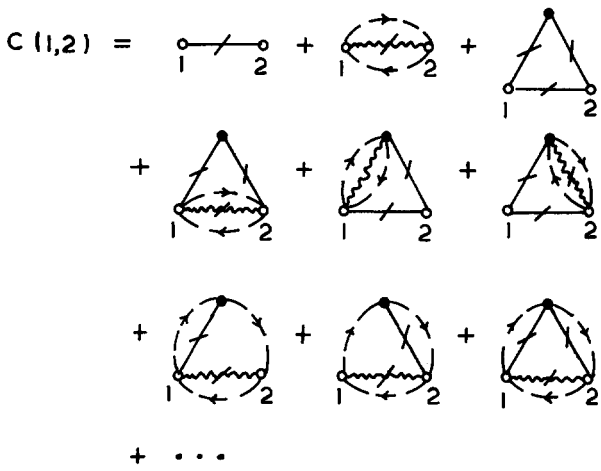


FIG. 2. The first few diagrams of the direct correlation function $c(1,2)$ for a fermion fluid. The symbols are the same as in Fig. 1.

white one-circles appearing in the expansion of $H(1,2)$. We denote this set by a function $S^{(2)}(1,2)$. The first few diagrams of $S^{(2)}(1,2)$ are demonstrated in Fig. 3, where h_B- and $(1+h_B)$-bonds are represented by solid and curly lines, respectively. Similarly we get a function $S^{(m)}(1,2,...,m)$ $(m>2)$ with m white one-circles defined as

$S^{(m)}(1,2,...,m)$ = [sum of all allowed CCD with m white one-circles labeled 1,2,...,m, respectively, some or no black ρ-circles, some f-bonds forming a closed loop connecting all circles, some or no h_B-bonds connecting circles in all possible ways such that each black circle is connected with at least one h_B-bond, no articulation circles and no articulation h_B-bonds]. (3.2)

In terms of the function $S^{(m)}$, one can obtain an expression for $H(1,2)$ in a compact form. We first evaluate the function $S^{(2)}(1,2)$ appearing in Fig. 3. Each diagram of $S^{(2)}(1,2)$ may be considered as the product of two chain diagrams at the white circles. Considering all the diagrams, the function $S^{(2)}$ can be expressed in terms of a new function $L(1,2)$, which is defined by

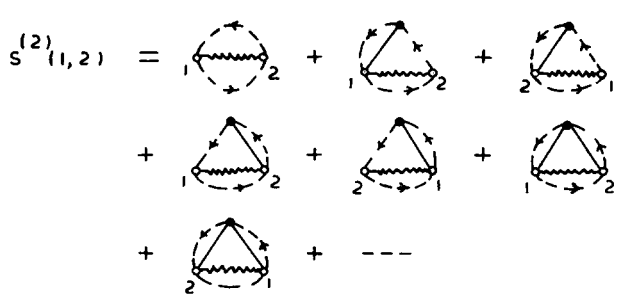


FIG. 3. The first few diagrams of $S^{(2)}(1,2)$. The solid and curly lines represent the h_B- and g_B-bonds, respectively, and the dashed line with an arrow denotes the f-bond.

$L(1,2)$ = [sum of all chain diagrams with two white one-circles consisting of only one path of f-bonds, h_B-bonds connecting in all possible ways such that each black circle is connected with at least one h_B-bond, each black circle has a vertex function $(-\rho/s)$]. (3.3)

The first few diagrams of $L(1,2)$ are shown in Fig. 4. The function L can be expressed in a compact form in terms of the $\mathscr{C}^{(n)}$-bond, which is defined in our previoud paper,[7] where a method of summing these chain diagrams has been discussed in detail. In the present paper, we only mention the final result. If $\mathscr{C}^{(0)}$ is the sum of the simple chains of f- and (fh_B)-bonds, then $\mathscr{C}^{(0)}$ can be expressed as[7]

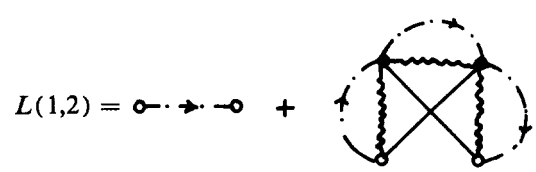
$$\mathscr{C}^{(0)}(i,j)=\frac{1}{(2\pi)^3(-\rho/s)}\int d\mathbf{k}\, e^{i\mathbf{k}\cdot\mathbf{r}_{ij}} \times\left[\frac{\tilde{f}(1+\tilde{\beta})+\tilde{\beta}^2(1+\tilde{f})}{1-\tilde{\beta}(1+\tilde{f})}\right], \tag{3.4}$$

where $\tilde{f}$ and $\tilde{\beta}$ are the Fourier transforms of $f(k_F r)$ and $f(k_F r)\,h(r)$, given by

$$\tilde{f}=\frac{-\rho}{s}\int dr\, e^{-ik\cdot r} f(k_f r)\,, \tag{3.5}$$

$$\tilde{\beta}=\frac{-\rho}{s}\int dr\, e^{-ik\cdot r} f(k_f r) h_B(r)\,. \tag{3.6}$$

We can sum the complex chains by replacing the f-bond by the $\mathscr{C}^{(0)}$-bond in Eq. (3.4). We can repeat the process of summing the chain diagrams except that the basic link in each chain is not f but $\mathscr{C}^{(n-1)}$ to obtain the new link $\mathscr{C}^{(n)}$. Thus $\mathscr{C}^{(n)}$ is expressed by Eq. (3.4) using $\mathscr{C}^{(n-1)}$ in place of f. The L can be expressed as



+ [more complex elementary diagrams with more than two black circles]. (3.7)

Here O—·—·→·—·O represents the $\mathscr{C}^{(n)}$-bond. In terms of the function L, $S^{(2)}$ can be expressed as

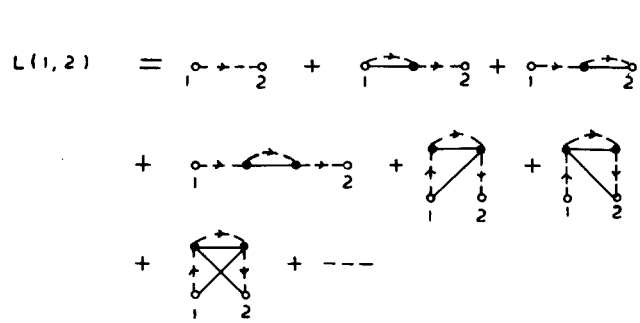
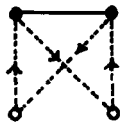


FIG. 4. Diagrammatic representation of $L(1,2)$. The symbols are the same as in Fig. 3.

$$S^{(2)}(1,2) = g_B(1,2)\xi(1,2)\,, \tag{3.8}$$

where

$$\xi(1,2) = L^2(1,2) + \quad + \cdots\,. \tag{3.9}$$



Here O—···→···—O denotes the L-bond. Similarly one can obtain an expression for $S^{(m)}$ $(m>2)$ as

$$S^{(m)}(1,2,...m) = g_B(1,2,...,m)L(1,2)L(2,3)\cdots L(m,1)\,. \tag{3.10}$$

By including the elementary diagrams in the expressions of L and $S^{(2)}$, one can obtain an exact expression for $S^{(2)}$. However, one can neglect the elementary diagrams to obtain $S^{(2)}$ approximately as

$$S^{(2)}(1,2) \simeq g_B(1,2)\gamma^{(n)}(1,2)$$
$$= g_B(1,2)[\,-(1/s)\mathscr{C}^{(n)^2}(1,2)]\,. \tag{3.11}$$

This approximation for $S^{(2)}$ is expected to provide a reasonably accurate expression for $h(1,2)$ of the Fermi system.

In terms of $S^{(m)}$ $(m\geqslant 2)$, Eq. (2.4) can be expressed as

$h(1,2) = h_B(1,2)$ + [sum of all allowed CCD with two white one-circles labeled 1 and 2, respectively, any number of black ρ-circles, at least one $S^{(m)}$-bond connecting any m circles $(m\geqslant 2)$ such that each black circle attached to an $S^{(m)}$-bond is also attached to at least one h_B-bond, some or no h_B-bonds, no articulation circle, no articulation S-bonds, and no articulation h_B-bonds]. (3.12)

In this description of $h(1,2)$, the pair of circles is singly connected either by an h_B-bond or an $S^{(m)}$-bond. The first few diagrams of $h(1,2)$ are shown in Fig. 5, where a solid line represents the h_B-bond, a helical line denotes the $S^{(2)}$-bond, and a helical triangle denotes the $S^{(3)}$-bond. In Fig. 5, the first diagram contains no S-bond, the last two diagrams contain an $S^{(3)}$-bond, while the rest contain an $S^{(2)}$-bond. The second diagram of Fig. 5 contains no h_B-bond.

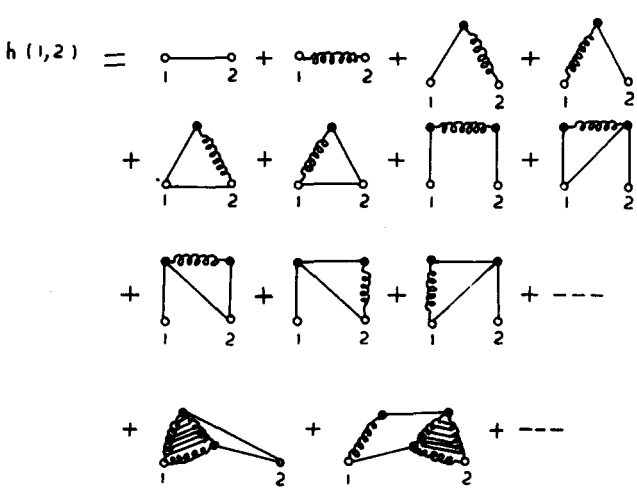


FIG. 5. The first few diagrams of $h(1,2)$. The solid line represents h_B-bonds, the helical line denotes the $S^{(2)}$-bond, and the helical triangle represents the $S^{(3)}$-bond.

Similarly the direct correlation function c can be written as

$$c(1,2) = c_B(1,2) + [\text{sum of all nodeless diagrams in } h]. \tag{3.13}$$

Equation (3.12) can be written as

$$h(1,2) = c(1,2) + [\text{sum of diagrams with one or more model circles}]. \tag{3.14}$$

Using the relation

$$g(1,2) = 1 + h(1,2), \tag{3.15}$$

we can write the pair distribution function $g(1,2)$ for the fermion system as

$$g(1,2) = g_B(1,2)[1 + D(1,2)], \tag{3.16}$$

where

$$D(1,2) = [\text{sum of all } h_B(1,2) \text{ irreducible diagrams in } h], \tag{3.17}$$

where an $h_B(1,2)$ irreducible diagram means there is no direct h_B-bond between the white circles 1 and 2.

In order to obtain an expression for D, we write it as a sum of five classes of diagrams,

$$D = D_d + D_0 + D_l + D_r + D_b\,, \tag{3.18}$$

where D_d is the subset of diagrams in D for which white circles are linked directly with the $S^{(2)}$-bond, D_0 is the subset of diagrams in D for which no white circle is the $S^{(m)}$-circle, D_l is the subset of diagrams in D for which the left white circle is the $S^{(m)}$-circle, D_r is the subset of diagrams in D with the right circle being a $S^{(m)}$-circle, and D_b is the subset of diagrams in D for which both white circles are $S^{(m)}$-circles, but with no direct $S^{(2)}$-bond between white circles. The five classes of diagrams are shown in Fig. 6. A simple difference between D_d and D_b is that the former has a direct link of the $S^{(2)}$-bond, but the latter has no such direct link.

Similarly, the direct correlation function can be expressed as a sum of five classes of diagrams

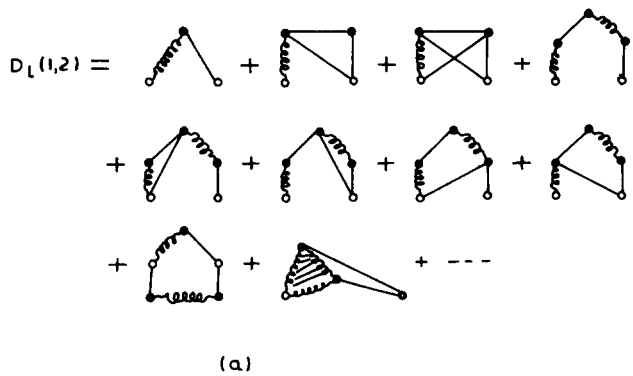
$$c = c_d + c_0 + c_l + c_r + c_b\,, \tag{3.19}$$

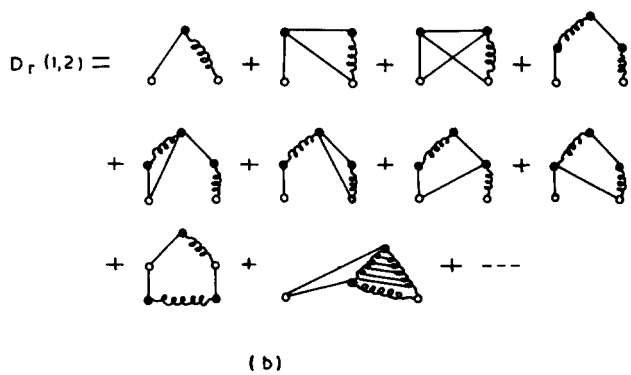
whre c_d, c_0, c_l, c_r, and c_b are the sets of nodeless diagrams in h corresponding to D_d, D_0, D_l, D_r, and D_b, respectively.

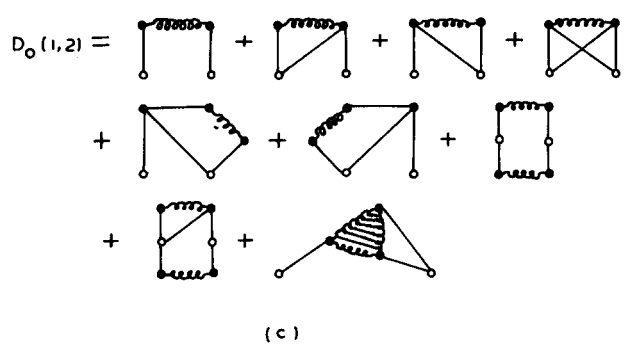
IV. THE SUMMATION OF SERIES

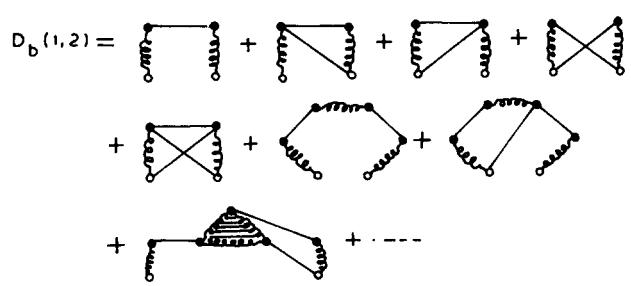
We first consider the contribution of simple chain diagrams appearing in Eq. (3.18), namely those diagrams of D_0, D_l, D_r, and D_b, which consist of only one path of lines joining the white circles 1 and 2. The diagrams of these types consist of alternate h_B- and $S^{(2)}$-bonds, which may be constructed by repeatedly inserting a "propagator" of Fig. 7(e), in place of black points. The few diagrams of series are shown in Figs. 7(a)–7(d). In the previous paper,[7] we have discussed the Fourier transform approach to summing these series. These diagrams can alternatively be summed to give

$$P_l(1,2) = Q_l(1,2) + \rho\int Q_l(1,3)P_l(3,2)d\mathbf{r}_3\,, \tag{4.1}$$









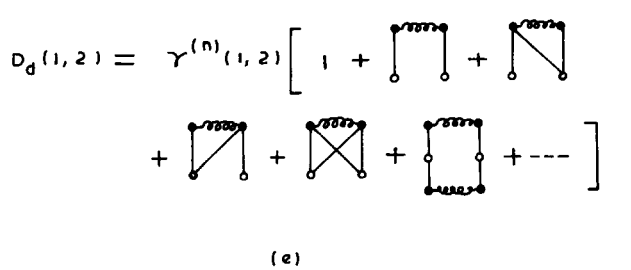


FIG. 6. The first few diagrams of (a) $D_l(1,2)$, (b) $D_r(1,2)$, (c) $D_0(1,2)$, (d) $D_b(1,2)$, and (e) $D_d(1,2)$. The symbols are the same as in Fig. 5.

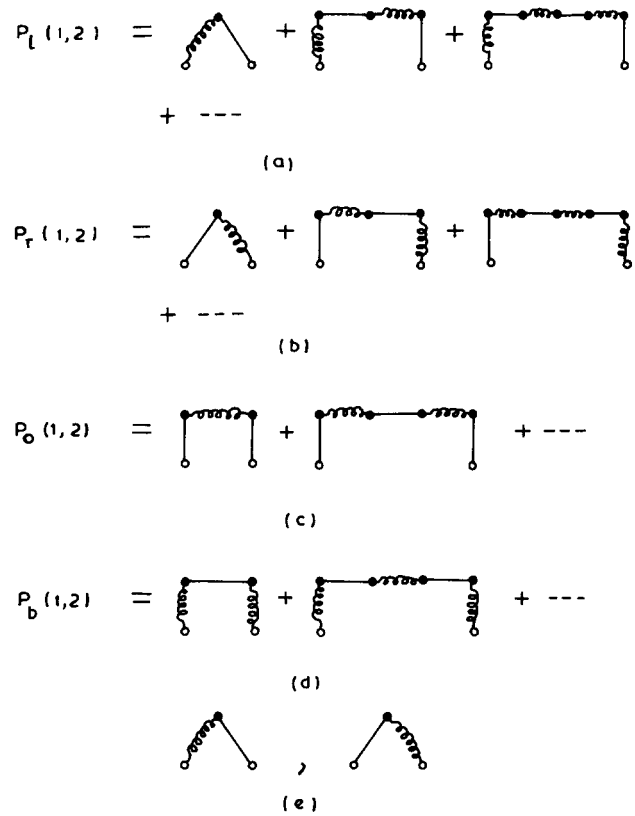


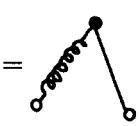
FIG. 7. Simple chains of h_B- and $S^{(2)}$-bonds: (a) $P_l(1,2)$, (b) $P_r(1,2)$, (c) $P_0(1,2)$, (d) $P_b(1,2)$, and (e) propagator.

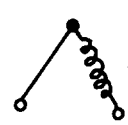
$$P_r(1,2) = Q_r(1,2) + \rho \int Q_r(1,3) P_r(3,2) d\mathbf{r}_3 , \tag{4.2}$$

$$P_0(1,2) = \rho \int h_B(1,3) P_l(3,2) d\mathbf{r}_3 , \tag{4.3}$$

$$P_b(1,2) = \rho \int P_l(1,3) S^{(2)}(3,2) d\mathbf{r}_3 , \tag{4.4}$$

where

$Q_l(1,2) =$  , (4.5a)

$Q_\gamma(1,2) =$  . (4.5b)

Numerically, $Q_l = Q_r$ and $P_l = P_r$. For a given h_B and $S^{(2)}$, we can solve these linear integral equations to obtain P_l, P_r, P_0, and P_b.

The expansion of D can now be given in terms of h_B-, S-, and P-bonds. This prescription includes a chain diagram (of P-bonds), nodal diagrams, and elementary diagrams.

We now define a new set of chain sums by including the set of nodal diagrams with one nodal point and one P-bond, as shown in Fig. 8, in Eqs. (4.1)–(4.4). These new chain sums are given by

$$G_l(1,2) = P_l(1,2) + A_l(1,2), \tag{4.6}$$

$$G_0(1,2) = P_0(1,2) + A_0(1,2), \tag{4.7}$$

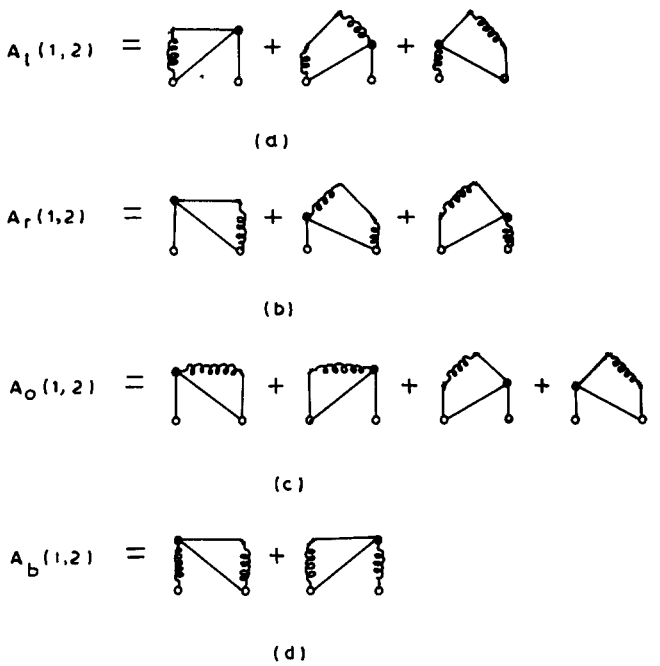


FIG. 8. Diagrams of (a) $A_l(1,2)$, (b) $A_r(1,2)$, (c) $A_0(1,2)$, and (d) $A_b(1,2)$.

$$G_b(1,2) = P_b(1,2) + A_b(1,2)\,, \tag{4.8}$$

where

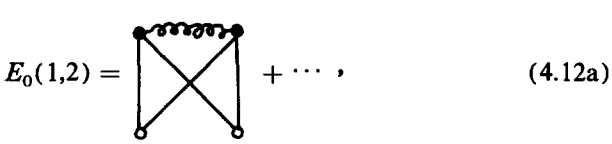
$$A_l(1,2) = \rho\int [P_l(1,3) + P_b(1,3)]h_B(1,3)h_B(3,2)d\mathbf{r}_3 + \rho\int S^{(2)}(1,3)P_0(3,2)h_B(3,2)d\mathbf{r}_3\,, \tag{4.9}$$

$$A_0(1,2) = 2\rho\int [P_r(1,3) + P_0(1,3)]h_B(1,3)h_B(3,2)d\mathbf{r}_3\,, \tag{4.10}$$

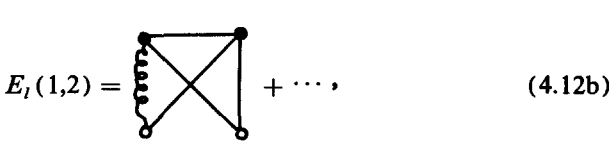
$$A_b(1,2) = 2\rho\int S^{(2)}(1,3)P_r(3,2)h_B(3,2)d\mathbf{r}_3\,. \tag{4.11}$$

Here $G_r = G_l$.

Let δ represent the remaining nodal diagrams with two or more P-bonds, and/or two or more nodal points. For convenience, we may call these diagrams "complex" nodal diagrams, while the diagrams of A may be called "simple" nodal diagrams. Some complex nodal diagrams are shown in Fig. 9. The elementary diagrams appearing in the expansion of D are given by



(4.12a)



(4.12b)

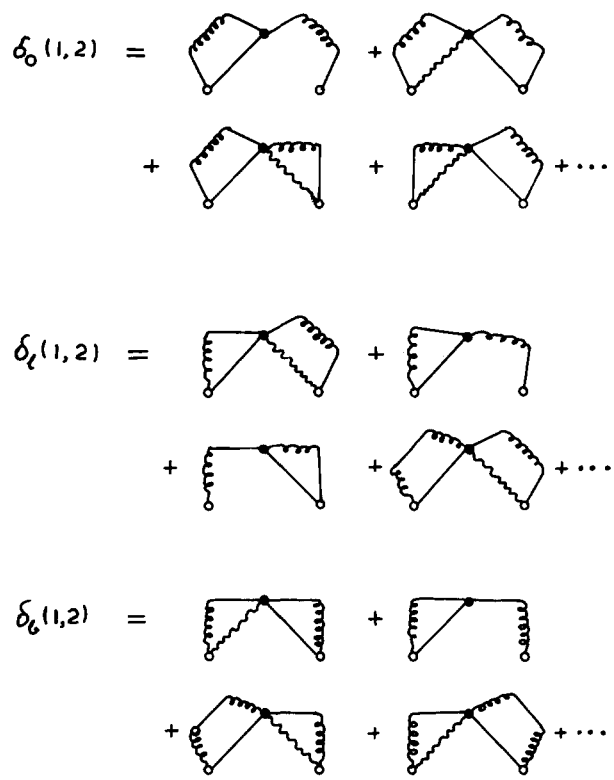
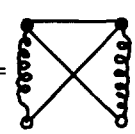


FIG. 9. Diagrams of (a) $\delta_l(1,2)$, (b), $\delta_0(1,2)$, and (d) $\delta_b(1,2)$.



(4.12c)

Now the diagrammatic expansion of D can be expressed in terms of $\gamma^{(n)}$, G, δ, and E. This prescription consists of summing these diagrams and their products at the white circles. Introducing a new function B defined by

$$B_0 = G_0 + \delta_0 + E_0, \tag{4.13a}$$

$$B_l = G_l + \delta_l + E_l\,, \tag{4.13b}$$

$$B_r = G_r + \delta_r + E_r\,, \tag{4.13c}$$

$$B_b = G_b + \delta_b + E_b\,. \tag{4.13d}$$

We now construct the "composite" diagrams by connecting the two white circles 1 and 2 by two or more B-bonds. Then we can write the expansion of D in the compact form as

$$D_d(1,2) = \gamma^{(n)}(1,2) + [\text{a set of composite diagrams with two white one-circles, labeled 1 and 2, one } \gamma^{(n)}\text{-bond and one or more } B_0\text{-bonds connecting the two white circles}], \tag{4.14a}$$

$$D_0(1,2) = B_0(1,2) + [\text{a set of composite diagrams with two white one-circles labeled 1 and 2, two or more } B_0\text{-bonds connecting the two white circles}], \tag{4.14b}$$

$D_l(1,2) = B_l(1,2) +$ [a set of composite diagrams with two white one-circles labeled 1 and 2, one B_l-bond and one or more B_0-bonds connecting two white circles], (4.14c)

$D_b(1,2) = B_b(1,2) +$ [a set of composite diagrams with two white one-circles labeled 1 and 2, one B_b-bond and one or more B_0-bond, or one B_l, one B_r -, and any number of B_0-bonds connecting the two white circles]. (4.14d)

Finally using Lemma 4 of Mortia and Hiroike,[13] the expansions of D_d, D_0, D_l, D_r, and D_b can be expressed in compact form. Thus

$$D_d = \gamma^{(n)} + \gamma^{(n)}[\exp(G_0 + \delta_0 + E_0) - 1], \quad (4.15a)$$

$$D_0 = [G_0 + \delta_0 + E_0] + [\exp(G_0 + \delta_0 + E_0) - 1 - (G_0 + \delta_0 + E_0)], \quad (4.15b)$$

$$D_l = [G_l + \delta_l + E_l] + [G_l + \delta_l + E_l] \times [\exp(G_0 + \delta_0 + E_0) - 1], \quad (4.15c)$$

$$D_r = D_l, \quad (4.15d)$$

$$D_b = [(G_b + \delta_b + E_b) + (G_l + \delta_l + E_l)(G_r + \delta_r + E_r)] \times [(G_b + \delta_b + E_b) + (G_l + \delta_l + E_l)(G_r + \delta_r + E_r)] \times [\exp(G_0 + \delta_0 + E_0) - 1], \quad (4.15e)$$

where δ_0, δ_l, and δ_b are sets of the complex model diagrams. The first few diagrams of δ_0, δ_l, and δ_b can be evaluated to give

$$\delta_0 = \rho[T_0 * T_0 + T_0 * (G_0 + G_l) + (G_0 + G_r) * T_0], \quad (4.16a)$$

$$\delta = \rho[S^{(2)} * (T_0 - h_B P_0) + 2G_l * T_l + (T_l + T_b) * (G_0 + T_0)], \quad (4.16b)$$

$$\delta_b = \rho[T_l * T_r + 2T_b * (G_r + T_r) + 2(G_l + G_b) * T_r + 2S^{(2)} * (T_r - h_B P_r)], \quad (4.16c)$$

where

$$T_0 = g_B[\exp G_0 - 1] - G_0, \quad (4.17a)$$

$$T_l = g_B G_l \exp G_0 - G_l, \quad (4.17b)$$

$$T_B = g_b[G_b + G_l G_r] \exp G_0 - G_b. \quad (4.17c)$$

Here $*$ denotes a spatial convolution, i.e.,

$$(A * B)(\mathbf{r}_1, \mathbf{r}_2) = \int A(\mathbf{r}_1, \mathbf{r}_3) B(\mathbf{r}_3, \mathbf{r}_2) d\mathbf{r}_3. \quad (4.18)$$

With the help of Eq. (4.15), D can be expressed as

$$D = \exp[t_0 + E_0][1 + \gamma^{(n)} + (t_l + E_l) + (t_r + E_r) + (t_b + E_b) + (t_l + E_l)(t_r + E_r)] - 1, \quad (4.19)$$

where

$$t = G + \delta. \quad (4.20)$$

Then the final results for g is

$$g = g_B \exp[t_0 + E_0][1 + \gamma^{(m)} + (t_l + E_l) + (t_r + E_r) + (t_b + E_b) + (t_l + E_l)(t_r + E_r)], \quad (4.21)$$

and for direct correlation functions

$$c_d = g_B \gamma^{(n)} \exp[t_0 + E_0], \quad (4.22a)$$

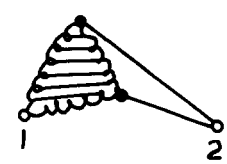
$$c_0 = g_B \exp[t_0 + E_0] - 1 - [h_B - c_B + t_0], \quad (4.22b)$$

$$c_l = g_B[t_l + E_l] \exp[t_0 + E_0] - t_l, \quad (4.22c)$$

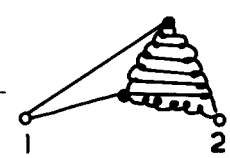
$$c_r = g_B[t_r + E_r] \exp[t_0 + E_0] - t_r, \quad (4.22d)$$

$$c_b = g_B[t_b + E_b + (t_l + E_l)(t_r + E_r)] \times \exp[t_0 + E_0] - t_b. \quad (4.22e)$$

In deriving Eqs. (4.21) and (4.22), we have not considered the diagrams with $S^{(m)}$-bonds ($m \geqslant 3$) such as the last diagrams in Fig. 5(a)–5(d). It is, however, possible to include some of these diagrams in the summation by simply replacing Q [defined in Eq. (4.5)] by Q^e in the expressions of P_l and P_r, respectively, where the Q^e are defined as

$$Q_l^e(1,2) = Q_l(1,2) +$$ 

, (4.23)

$$Q_\gamma^e(1,2) = Q_r(1,2) +$$ 

. (4.24)

Expressions of P thus include a set of diagrams with $S^{(3)}$-bonds. Equations (4.21) and (4.22) thus give formally exact expressions for the correlation functions of the fermion system. However, it is important to emphasize that different approximations can be obtained from Eqs. (4.21) and (4.22) by neglecting a different set of diagrams.

If we set all $E = 0$, this set of equations sums all diagrams of the HNC type, which we may term a modified FHNC approximation. Thus under this approximation the pair distribution function (PDF) is given by

$$g_{\text{FHNC}} = g_B \exp[t_0][1 + \gamma^{(n)} + t_l + t_r + t_b + t_l t_r] \quad (4.25)$$

and the direct correlation functions are expressed as

$$c_d = g_B \gamma^{(n)} \exp t_0, \quad (4.26a)$$

$$c_0 = g_B \exp t_0 - 1 - [h_B - c_B + t_0], \quad (4.26b)$$

$$c_l = g_B t_l \exp t_0 - t_l, \quad (4.26c)$$

$$c_r = c_l, \quad (4.26d)$$

$$c_b = g_B[t_b + t_l t_r] \exp t_0 - t_b. \quad (4.26e)$$

It is interesting to note that this approximation is a close analog to the FHNC approximation of Fantoni and Rosati (FR).[9] However, it differs from the FR-FHNC in the sense

that it is expressed in terms of h_B. When compared with our previous expressions[7] for g, we find that the present result includes δ and A, which are not considered previously.

The corresponding Percus–Yevick (PY) analog is obtianed by further linearizing t. Thus we get

$$c_d = S^{(2)} \equiv g_B \gamma^{(n)}, \tag{4.27a}$$

$$c_0 = c_B + h_B t_0, \tag{4.27b}$$

$$c_l = h_B t_l, \tag{4.27c}$$

$$c_r = c_l, \tag{4.27d}$$

$$c_b = h_B t_b . \tag{4.27e}$$

This may be termed as a modified Fermi–Percus–Yevick (FPY) approximation.

Another approximation can be obtained by setting all $\delta = 0$. Thus we obtain an expression for g as

$$g(r) = g_B(r)\exp[G^0(r)]\,[1 + \gamma^{(n)}(r) + G_l(r) + G_r(r) + G_b(r) + G_l(r)G_r(r)]\,, \tag{4.28}$$

and for the direct correlation function as

$$c_d = g_B \gamma^{(n)} \exp G_0, \tag{4.29a}$$

$$c_0 = g_B \exp G_0 - 1 - [h_B - c_B + G_0], \tag{4.29b}$$

$$c_l = g_B G_l \exp G_0 - G_l, \tag{4.29c}$$

$$c_r = c_l, \tag{4.29d}$$

$$c_b = g_b\,[G_b + G_l G_r]\exp G_0 - G_b. \tag{4.29e}$$

Equations (4.28) and (4.29) give approximate expressions for correlation functions. They are obtained by neglecting highly connected diagrams, namely δ and E, whose contributions are expected to be small. It is interesting to note that this approximation is in a close analog to the Exp approximation considered by Andersen and Chandler[15] for a simple atomic fluid. We may call this an F-EXP approximation.

Retaining the term of δ given in Eq. (4.16), we obtain an expression for g;

$$g(r) = g_{\text{F-EXP}}(r)\exp\delta_0(r)\,[1 + \{\delta_l(r) + \delta_r(r) + \delta_b(r) + G_l(r)\delta_r(r) + \delta_l(r)G_r(r) + \delta_l(r)\delta_r(r)\} \times \{1 + \gamma^{(n)} + G_l(r) + G_r(r) + G_b(r) + G_l(r)G_r(r)\}^{-1}]\,. \tag{4.30}$$

This is an improvement over the F-EXP approximation. This approximation corresponds to the RHNC approximation of Stell and Weis[16] for simple atomic fluids.

V. INTEGRAL EQUATIONS

We first take into consideration only the simple linear chains and write the pair correlation function $h(r)$ of the Fermi system as

$$h(r) = h_B(r) + S^{(2)}(r) + P_0(r) + P_l(r) + P_r(r) + P_b(r), \tag{5.1}$$

where $r = |\mathbf{r}_1 - \mathbf{r}_2|$. The Fourier transform of it yields

$$h(\mathbf{k}) = h_B(\mathbf{k}) + S^{(2)}(\mathbf{k}) + P_0(\mathbf{k}) + P_l(\mathbf{k}) + P_r(\mathbf{k}) + P_b(\mathbf{k}). \tag{5.2}$$

Using the Fourier transforms of Eqs. (4.1)–(4.4) we get an expression for $h(\mathbf{k})$ in the matrix form

$$H^0 = h - S^{(2)} = \omega c_B \omega/(1 - \rho\omega c_B) \tag{5.3}$$

or

$$H^0 = \omega c_B[\omega + \rho h\,], \tag{5.4}$$

where ω is the matrix with elements

$$\omega(\mathbf{k}) = l + \rho S^{(2)}(\mathbf{k}) \tag{5.5}$$

and c_B is the matrix with the element

$$c_B(\mathbf{k}) = \int c_B(r)\exp[-i\mathbf{k}\cdot\mathbf{r}]d\mathbf{r}. \tag{5.6}$$

Here $c_B(r)$ is the direct correlation function of boson system generated by F-bonds only. From Eq. (5.4), one gets

$$\chi = 1 + \rho h \equiv \omega + \rho H^0 = [1 - \rho\omega c_B\,]^{-1}\omega. \tag{5.7}$$

In Eq. (5.7), the contribution of diagrams involving $S^{(2)}$-bonds to the direct correlation function c are neglected. In order to improve the result, their contribution should be taken into account. One can improve the result by replacing the direct correlation function c_B by c. Thus Eq. (5.7) takes the form

$$\chi = \omega + \rho H^0 = [1 - \rho\omega c]^{-1}\omega. \tag{5.8}$$

The bears a resemblance to the reference interaction site model (RISM) integral equation[17] for the site–site correlation function for the molecular fluid, where the exchange bond $S^{(2)}$ is replaced by the intramolecular bond. However, this approximation includes some unallowed diagrams, which do not appear in the usual expansion. In this sense, Eq. (5.8) is not a proper integral for the Fermi system.

When four direct correlation functions, discussed in the previous section [Eq. (4.22)], are used, the cluster series for the pair correlation function is expressed in the form

$$\chi = \omega + \rho H_d^0 = (1 - \rho\Omega c_0)^{-1}\Omega, \tag{5.9}$$

where

$$\Omega = (1 - \rho c_l)^{-1}(\omega + \rho c_b)(1 - \rho c_r)^{-1} = \omega + \Delta\Omega, \tag{5.10}$$

and

$$H_d^0 = h - c_d. \tag{5.11}$$

Here c_d is the Fourier transform of $c_d(r)$ given by Eq. (4.29a). Equation (5.9) is the principal result, which is identical to the proper integral equaiton of Chandler *et al.*[18] for the site–site correlation function of the molecular fluid. No unallowed diagrams are included in Eqs. (5.9) and (5.10). With the expressions for c_0, c_l, c_r, c_b, and c_d given by Eq. (4.29), one can evaluate χ and h easily.

VI. CONCLUDING REMARKS

We have developed a method that provides two routes for calculating the correlation functions for the fermion fluids. One is in terms of D obtained under different approximations and the other is the integral equation in terms of the four distinct direct correlation functions. The integral equation is much simpler and given in the computationally convenient form.

Using cluster expansion and topological reduction, we

have obtained systematically a set of computationally convenient approximations for the correlation functions of the fermion system. These approximations are expressed in terms of $S^{(2)}$, which represents a partial summation of the exchange expansion. A set of diagrams involving $S^{(m)}$ ($m \gtrsim 3$), which have a very compact structure, are not taken into consideration in these approximations. They are, however, highly connected diagrams, whose contributions are expected to be small. The approximations developed here are expected to yield the correlation functions of fermion fluids with the same accuracy as their counterparts have been found to yield in the case of classical fluids. The application of this theory to some fermion systems will be a subject of future publication.

ACKNOWLEDGMENT

I am grateful to Professor Y. Singh for many stimulating discussions and comments on the subject.

[1]C. E. Campbell, in *Progress in Liquid Physics*, edited by C. A. Croxton (Wiley, New York, 1978).
[2]B. D. Day, Rev. Mod. Phys. **50**, 495 (1978).
[3]E. Feenberg, *Theory of Quantum Fluids* (Academic, New York, 1969).
[4]D. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **16**, 3081 (1977).
[5]A. Bijl, Physica **1**, 869 (1940); R. Jastrow, Phys. Rev. **98**, 1479 (1955).
[6]M. D. Miller, Phys. Rev. B **14**, 3937 (1976).
[7]S. K. Sinha and Y. Singh, J. Math. Phys. **24**, 2504 (1983).
[8]E. Krotscheck and M. L. Ristig, Phys. Lett. A **48**, 17 (1974); Nucl. Phys. A **242**, 387 (1975).
[9]S. Fantoni and S. Rosati, Nuovo Cimento A **10**, 145 (1972); **20**, 179 (1974); **25**, 593 (1975); Lett. Nuovo Cimento **10**, 545 (1974).
[10]J. G. Zabolitzky, Phys. Lett. B **64**, 233 (1976); Phys. Rev. A **16**, 1258 (1977).
[11]E. Krotscheck, Z. Phys. B **32**, 395 (1979).
[12]C. E. Campbell, K. E. Kurten, and E. Krotscheck, Phys. Rev. B **25**, 1633 (1982).
[13]T. Morita and K. Hiroike, Prog. Theor. Phys. **25**, 537 (1961).
[14]G. Stell, in *The Equilibrium Theory of Classical Fluids*, edited by H. L. Frisch and J. L. Lebowitz (Benjamin, New York, 1964).
[15]H. C. Andersen and D. Chandler, J. Chem. Phys. **57**, 1918 (1972).
[16]G. Stell and J. J. Weis, Phys. Rev. A **21**, 645 (1980).
[17]D. Chandler, Mol. Phys. **31**, 1213 (1976).
[18]D. Chandler, C. G. Joslin, and J. M. Deutch, Mol. Phys. **47**, 871 (1982).

Solutions of Euclidean σ models on noncompact Grassmann manifolds

J.-P. Antoine and B. Piette[a)]
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

(Received 5 August 1987; accepted for publication 24 February 1988)

A method for constructing explicit solutions of classical, two-dimensional, Euclidean σ models on Grassmann manifolds of noncompact type, and their supersymmetric extensions, is presented. It generalizes the "holomorphic" method developed by Din, Zakrzewski, and Sasaki in the compact case, by taking properly into account the geometry of indefinite metric spaces. The solutions obtained are local, in the sense that they may become singular on certain curves in the plane, and the signature of the solution may change across such a curve.

I. INTRODUCTION AND DEFINITION OF THE MODEL

There are many reasons for considering two-dimensional σ models on noncompact manifolds. In the Minkowskian regime, they have arisen in a number of physical situations, such as the Gross–Neveu model,[1] relativistic string theory,[2] gravitation[3] (black holes, Ernst equations), supergravity models,[4] and gauge equivalence with various nonlinear dynamical systems.[5] In quantum field theory, Euclidean σ models on Grassmann manifolds have been used quite often as a laboratory for gauge and string theories, but almost always in the compact case.[6] On the other hand, Euclidean noncompact σ models, even supersymmetric ones, are used in solid state physics, for describing (Anderson) localization properties in disordered conductors[7] and superconductors.[8] Finally the mathematical interest is obvious. A classical solution of a σ model is nothing but a harmonic map into some (pseudo-)Riemannian manifold, and such objects have great geometrical significance. Grassmannian models, in particular, have received some attention recently,[9] but in the compact case only.

Since the object of a σ model is a massless field constrained to live on a manifold M, but otherwise free, the geometry of M is of central importance. In a previous paper,[10] we have clarified this issue for models on Grassmann manifolds, both of compact and noncompact type. Then we have discussed the implications of our results for the Bäcklund transformation method of Harnad *et al.*,[11] and we have also shown the validity of that method for Euclidean models, in addition to the usual Minkowskian ones.

Now, for Euclidean models, the most popular method of solution is the so-called "holomorphic" method, introduced by Borchers and Garber[12] and developed systematically by Din and Zakrzewski[13] and by Sasaki,[14] but only for the compact case (except for some brief remarks in the last paper). In the present paper we will show that the method (denoted DZS in the sequel) works for noncompact models as well, provided one takes properly into account the geometry of indefinite metric spaces. The same is true for the supersymmetric extensions of those models considered by Din *et al.*[15] and Fujii *et al.*[16,17] Indeed the key point in the DZS method consists in constructing an orthonormal basis for a subspace of $\mathbb{C}^N$ out of a family of homomorphic vectors $f \equiv f(x+it)$, using for instance the familiar Gram–Schmidt procedure. Now for a noncompact model, the metric on $\mathbb{C}^N$ is indefinite, and the construction breaks down in general. The best one can do is to construct an orthogonal family, but some of the vectors may have zero norm. An additional condition is needed to exclude this behavior, namely, that the subspace be nondegenerate.

To be more specific, let $V = \mathbb{C}^{p+q}$ (we may take $\mathbf{R}^{p+q}$ as well), equipped with the inner product given by the metric $t = I_{pq} = \mathrm{diag}[\mathbf{1}_p, -\mathbf{1}_q]$:

$$\langle f|g\rangle = f^\dagger t\, g = \sum_{\alpha=1}^{p} \overline{f_\alpha} g_\alpha - \sum_{\beta=p+1}^{p+q} \overline{f_\beta} g_\beta . \qquad (1.1)$$

This metric is indefinite, but nondegenerate, in the sense that no vector is orthogonal to the whole space: $V \cap V^\perp = \{0\}$. The geometry of such a space is described for instance in the monographs of Bognar[18] or Deheuvels.[19] The crucial property of a subspace $W \subset V$ is the equivalence between the following three conditions:

(i) W is nondegenerate: $W \cap W^\perp = \{0\}$;
(ii) W is orthocomplemented: $V = W \oplus W^\perp$;
(iii) W is the range of an orthogonal projection P: $W = PV$, where P is a $(p+q)\times(p+q)$ matrix satisfying the relations $P^2 = P$ and $\langle Pf|g\rangle = \langle f|Pg\rangle$.

This last aspect is the key one for Grassmannian σ models, since the field of such a model is precisely a projection $P = P(x)$, $x \in \mathbf{R}^2$, as we discussed at length in Ref. 10.

Before proceeding, we have to introduce some notation. Let $F = \{f_1, \ldots, f_m\}$ be a family of m vectors in $\mathbb{C}^N$. We will denote by the same symbol F the $N \times m$ matrix $F = [f_1, \ldots f_m]$, whose columns are the vectors f_j, and by $\langle F\rangle$ the vector subspace of $\mathbb{C}^N$ generated by F. If P is an orthogonal projection in $\mathbb{C}^N$ (as defined above), its action on the subspace $\langle F\rangle$ is given by

$$P\langle F\rangle = \langle PF\rangle \equiv \langle Pf_1, \ldots, Pf_m\rangle . \qquad (1.2)$$

Finally, if the subspace $\langle G\rangle$ is nondegenerate, we denote the corresponding projection by P_G or $P[G]$.

For the sake of completeness we recall the main definitions of (complex) Grassmannian σ models. The target manifolds are the following symmetric spaces:

[a)] Chercheur IISN (Belgium).

 0022-2488/88/071687-11$02.50

(i) compact type:

$$G_{pq}(\mathbb{C}) = \mathrm{SU}(p+q)/\mathrm{S}(\mathrm{U}(p)\times\mathrm{U}(q))\,; \tag{1.3a}$$

(ii) noncompact type:

$$G_{pi;qj}(\mathbb{C}) = \mathrm{SU}(p,q)/\mathrm{S}(\mathrm{U}(p-i,q-j)\times\mathrm{U}(i,j)),$$

$$0\leqslant i\leqslant p,\ 0\leqslant j\leqslant q\,. \tag{1.3b}$$

The field $Z = Z(x)$, $x\in\mathbb{R}^2$, is a $(p+q)\times r$ matrix verifying the relations

$$Z^{\dagger}t\,Z = t_Z\,, \tag{1.4}$$

where $r = q$, $t = \mathbb{1}_{p+q}$, and $t_Z = \mathbb{1}_q$ in the compact case, and $t = I_{pq}$, $t_Z = I_{ij}$, with $i+j=r$, in the noncompact one. The Lagrangian reads

$$\mathscr{L} = \tfrac{1}{2}\,\mathrm{Tr}[\,(D_\mu Z)^{\dagger}t\,(D^\mu Z)t_Z\,]\,,\quad \mu = 0,1\,, \tag{1.5}$$

the covariant derivative being defined as

$$\begin{aligned} D_\mu Z &= \partial_\mu Z - Zt_Z Z^{\dagger}t\,\partial_\mu Z \\ &= (1-P)\partial_\mu Z\,, \end{aligned} \tag{1.6}$$

where $P = Zt_Z Z^{\dagger}t$ is the projection on the r-dimensional subspace spanned by the columns of Z. The Euler–Lagrange equations derived from the Lagrangian (1.5) are

$$D_\mu D^\mu Z - Zt_Z(D_\mu Z)^{\dagger}t\,(D^\mu Z) = 0 \tag{1.7}$$

or, in terms of the projection-valued field $P = P(x)$,

$$[\partial_\mu\partial^\mu P, P\,] = 0\,. \tag{1.8}$$

Equations (1.5)–(1.8) are general, but from now on we consider only the Euclidean metric on $\mathbb{R}^2$. Then, going over to complex coordinates $x_{\pm} = (x \pm iy)/2$, Eqs. (1.7) and (1.8) become, respectively,

$$D_+D_-Z + Zt_Z(D_-Z)^{\dagger}t\,(D_-Z) = 0\,, \tag{1.9}$$

$$[\partial_+\partial_- P, P\,] = 0\,. \tag{1.10}$$

The self-dual solutions are those fields that verify the simplified equation

$$D_-Z = 0 \tag{1.11}$$

or, equivalently,

$$\partial_- P\cdot P = 0\,. \tag{1.12}$$

The paper is organized as follows. In Sec. II, we describe the general method of solution, valid for all complex or real Grassmannian models, compact or not. It generalizes the constructions of Din and Zakrzewski[13] and Sasaki.[14] The key technical ingredient is a recursive process for constructing an orthonormal basis in an indefinite metric space; this process, which is described in detail in the Appendix, is a proper substitute for the Gram–Schmidt procedure. We also consider the particular case of the symplectic models; these have no self-dual solutions, since the target manifold is not Kählerian. In Sec. III we discuss somewhat further the properties of our solutions. As mentioned above the solution of a Grassmannian model is a projection-valued field, of a certain rank and signature. The rank of the solution may be specified *a priori*; it is simply the number of linearly independent vectors one starts from. But a novel feature appears for noncompact models: the signature of the solution is not known beforehand, it can only be computed after the construction is completed. Worse yet, the signature of the projection $P(x)$ may vary across some lines of the x_+ plane, on which the solution becomes singular! In the same section, we indicate how to evaluate the action of the solutions (not all solutions have finite action!).

In Sec. IV, finally, we extend to the noncompact case the supersymmetric versions of the models introduced originally by Din *et al.*[15] and Fujii *et al.*[16,17]; more precisely, we examine successively the case of a fermion in the field of a bosonic background, and then the fully coupled supersymmetric equation.

II. SOLUTIONS OF THE NONCOMPACT GRASSMANNIAN σ MODELS

In this section we will construct explicit solutions of the Euclidean σ model on the complex Grassmann manifold $G_{pi;qj}(\mathbb{C}) = \mathrm{SU}(p,q)/\mathrm{S}(\mathrm{U}(p-i,q-j)\times\mathrm{U}(i,j))$. The basic field of the model is a projection on a subspace of dimension $(i+j)$ and signature (i,j) of $\mathbb{C}^{p+q}$. The method of construction is an extension of the one used by Din and Zakrzewski[13] and by Sasaki[14] for the compact case.

A. Self-dual solutions

As usual the simplest solutions are the self-dual ones, and actually they will be needed later on for constructing more general ones. So we start with the self-dual case.

For any r, $1\leqslant r\leqslant p+q$, choose a set of r linearly independent holomorphic vectors $F = \{f_1,\ldots,f_r\}$, $f_m\in\mathbb{C}^{p+q}$, $\partial_- f_m = 0$ $(m = 1,\ldots,r)$, and define the corresponding $(p+q)\times r$ matrix $F = [\,f_1,\ldots,f_r\,]$. Assume that the $(r\times r)$ matrix $t_F = F^{\dagger}t\,F$ is nonsingular. Then $P_F = Ft_F{}^{-1}F^{\dagger}t$ is the projection on the subspace $\langle F\rangle$ spanned by the columns of F. Following the arguments of Macfarlane,[20] one shows readily that P_F verifies the self-dual equation $\partial_- P_F\cdot P_F = 0$. We emphasize that, whereas rank $P_F = \dim\langle F\rangle = r$ has been chosen *a priori*, its signature is *not*: to obtain it, one must diagonalize the matrix t_F, which is not so easy in general. So we meet here the basic ambiguity of noncompact Grassmannian models: the method yields a solution of the model on $G_{pi;qj}$, for *some* pair (i,j) with $i+j=r$, but otherwise unknown. We will come back to this point in Sec. III. It is interesting to notice that the same ambiguity was found already[10] in the Bäcklund transformation method.

B. More general solutions

In the compact case, quite general, not necessarily self-dual, solutions have been obtained by Din and Zakrzewski[13] and by Sasaki.[14] For noncompact models, their methods apply, too, up to a point, because a given set of independent vectors need not always be orthonormalizable. The relevant information and algorithms are described in the Appendix, with the main result summarized in Sec. I above.

Again the starting point is a set of linearly independent holomorphic vectors $F=\{f_1,...,f_l\}$, $\partial_- f_m=0$ for $m=1,...,l$. We emphasize that the set F is completely arbitrary; for instance, a vector f_j might very well be the derivative of another one, say f_k: $f_j=\partial_+ f_k$. Now we partition F in an arbitrary way into subsets $F_1,F_2,...,F_k$, some of which may be empty (k is arbitrary). As usual each F_n may be considered as a $(p+q)\times r_n$ matrix, where r_n is the number of vectors in F_n. Then we extend our set of vectors by adding derivatives of the f_m's, through the following recursive process:

$$\begin{aligned} &H^1=F_1\,,\\ &H^2=\{\partial_+ H^1,F_2\}\,,\\ &H^n=\{\partial_+ H^{n-1},F_n\}\quad (2\leqslant n\leqslant k)\,. \end{aligned} \tag{2.1}$$

In those relations, if H^m is the subset $\{f_1^{(m)},f_2^{(m)},...\}$, $\partial_+ H^m$ denotes the set $\{\partial_+ f_1^{(m)},\partial_+ f_2^{(m)},...\}$, in the same order. Furthermore, it is understood that, at each step of the recursion, every vector linearly dependent on its predecessors is dropped. Then taking together all the blocks $H^1,...,H^k$, we obtain our basic set of holomorphic vectors,

$$H=\{H^1,H^2,...,H^k\}\,, \tag{2.2}$$

where we may assume without loss of generality that $\langle H\rangle=\mathbb{C}^{p+q}$. Next, we select out of H a certain number of *connected* subsets $E_1,E_2,...,E_s$, with $E_n=\{H^{k_1},...,H^{k_n}\}$, and denote by $G_1,G_2,...,G_{s+1}$ the blocks of H^m's left out ("gaps"),

$$H=\{G_1,E_1,G_2,E_2,...,G_s,E_s,G_{s+1}\}\,, \tag{2.3}$$

where all blocks E_n,G_n are nonempty, except possibly G_1 and G_{s+1}. Finally, we consider the following subsets of H:

$$E=\{E_1,E_2,...,E_s\}\,, \tag{2.4a}$$

$$H(E_n)=\{G_1,E_1,G_2,E_2,...,E_n\},\quad 1\leqslant n\leqslant s\,, \tag{2.4b}$$

$$H(G_n)=\{G_1,E_1,G_2,E_2,...,E_{n-1},G_n\},\quad 1\leqslant n\leqslant s+1\,. \tag{2.4c}$$

The following relations are immediate:

$$H(G_n)\subset H(E_n)\subset H(G_{n+1})\,, \tag{2.5a}$$

$$\partial_+ G_n\subseteq H(E_n)\,, \tag{2.5b}$$

$$\partial_+ E_n\subseteq H(G_{n+1})\,. \tag{2.5c}$$

We say that the partition (2.3) of H is *completely nondegenerate* if all subspaces $\langle H(E_n)\rangle,\langle H(G_n)\rangle$, for $n=1,...,s$, are nondegenerate. In addition, any set of holomorphic vectors E constructed by the process above will be called a *DZ set* (for Din and Zakrzewski[13]), and a completely nondegenerate one if the partition of H is. This definition is motivated by the following property.

Lemma 2.1: Any completely nondegenerate partition $H=\{G_1,E_1,G_2,...,E_s,G_{s+1}\}$ generates an orthonormal basis $\{{}^0G_1,{}^0E_1,{}^0G_2,...,{}^0E_s,{}^0G_{s+1}\}$ of $\mathbb{C}^{p+q}$.

Proof: By assumption, for each $n=1,...,s$, the orthocomplement of $\langle H(G_n)\rangle$ in $\langle H(E_n)\rangle$ is a nondegenerate subspace, namely, $\langle\widetilde{E}_n\rangle$, where $\widetilde{E}_n=(1-P[H(G_n)])E_n$ and $P[H(G_n)]$ is the projection on $\langle H(G_n)\rangle$,

$$\langle H(E_n)\rangle=\langle H(G_n)\rangle\oplus\langle\widetilde{E}_n\rangle\,. \tag{2.6a}$$

Similarly, for each $n=1,...,s+1$,

$$\langle H(G_n)\rangle=\langle H(E_{n-1})\rangle\oplus\langle\widetilde{G}_n\rangle\,, \tag{2.6b}$$

with $\widetilde{G}_n=(1-P[H(E_{n-1})])G_n$. Applying systematically this procedure to the given partition of H, starting, e.g., from the left, we obtain a new parametrization of $\mathbb{C}^{p+q}$,

$$\widetilde{H}=\{\widetilde{G}_1,\widetilde{E}_1,\widetilde{G}_2,...,\widetilde{G}_s,\widetilde{E}_s,\widetilde{G}_{s+1}\}\,, \tag{2.7}$$

where the subspaces $\langle\widetilde{E}_m\rangle,\langle\widetilde{G}_n\rangle$ are mutually orthogonal and have orthonormal bases ${}^0E_m,{}^0G_n$, respectively. Putting all those bases together, we obtain the required orthonormal basis of $\mathbb{C}^{p+q}$, and in particular for the sets $\langle H(E_n)\rangle$ and $\langle H(G_n)\rangle$. ■

The construction described in the proof of Lemma 2.1 is a block-by-block version of the Gram–Schmidt procedure, valid for any completely nondegenerate partition of H. The bases ${}^0E_m,{}^0G_n$ verify the following relations that will be used in the sequel [compare with (2.5)]:

$$P[H(G_{n+1})]\partial_+ {}^0E_n=\partial_+ {}^0E_n\,, \tag{2.8a}$$

$$P[H(E_n)]\partial_+ {}^0G_n=\partial_+ {}^0G_n\,. \tag{2.8b}$$

An explicit construction of the bases ${}^0E_m$, ${}^0G_n$ is described in the Appendix. We emphasize that, whereas the vectors of E_m,G_n are holomorphic by construction, those of $\widetilde{E}_m,\widetilde{G}_n$ and ${}^0E_m$, ${}^0G_n$ are not, in general.

For each $n=1,2,...,s$, we define the following subsets of $\widetilde{H}$, which also span nondegenerate subspaces:

$$\widetilde{E}^{(n)}=\{\widetilde{E}_1,\widetilde{E}_2,...,\widetilde{E}_n\}\,, \tag{2.9a}$$

$$\widetilde{G}^{(n)}=\{\widetilde{G}_1,\widetilde{G}_2,...,\widetilde{G}_n\}\,. \tag{2.9b}$$

We also call P_n (resp. Q_n) the projections on $\langle\widetilde{E}^{(n)}\rangle$ (resp. $\langle\widetilde{G}^{(n)}\rangle$), and denote by $Z_n\equiv{}^0E^{(n)}=\{{}^0E_1,{}^0E_2,...,{}^0E_n\}$ (resp. $W_n\equiv{}^0G^{(n)}=\{{}^0G_1,{}^0G_2,...,{}^0G_n\}$) the corresponding orthonormal bases.

Proposition 2.2: Let $E=\{E_1,...,E_s\}$ a completely nondegenerate DZ set of dimension r, $P=P_s$ the projection on the corresponding subspace. Then P is a solution of the Euclidean σ model on a Grassmann manifold $G_{pi;qj}(\mathbb{C})$, where $i+j=r$.

Proof: We follow Refs. 13 and 14. The projection $P_s+Q_s=P[H(E_s)]$ is a self-dual solution of rank r (since it consists of r consecutive vectors, with no gaps). Hence

$$\partial_-(P_s+Q_s)\cdot(P_s+Q_s)=0\,. \tag{2.10}$$

From (2.8b), with $n=s$, we get

$$(P_s+Q_s)\partial_+ Q_s=\partial_+ Q_s \tag{2.11}$$

and, using the analyticity of the starting vectors $\{f_m\}$,

$$(P_s+Q_s)\partial_- Q_s=\partial_- Q_s\,. \tag{2.12}$$

Taking conjugates, we get $[\partial_\pm Q_s,P_s+Q_s]=0$. From this and the relation $P_sQ_s=0$, (2.10) becomes

$$\partial_- P_s \cdot P_s + Q_s \cdot \partial_- Q_s = 0 . \tag{2.13}$$

Taking conjugates again and deriving, one obtains finally

$$[\partial_+ \partial_- P_s, P_s] = [\partial_+ \partial_- Q_s, Q_s] , \tag{2.14}$$

which means that P_s is a solution if Q_s is a solution. Now we iterate the process: $P_{s-1} + Q_s = P[H(G_s)]$ is also a self-dual solution. By the same argument, we obtain

$$[\partial_+ \partial_- Q_s, Q_s] = [\partial_+ \partial_- P_{s-1}, P_{s-1}] , \tag{2.15}$$

i.e., Q_s is a solution if P_{s-1} is one. Proceeding in the same way, we end up with a last set, namely, $\{E_1\}$ if G_1 is empty, $\{G_1\}$ otherwise, for which the corresponding projection is a self-dual solution. Hence $P \equiv P_s$ is a solution, and it has rank r by construction. ■

Exactly as in the case of self-dual solutions, the rank of P, i.e., dim E, has been specified *a priori*, but the construction gives no information whatsoever on the signature, i.e., we know that P is a solution of some model, but we do not know *a priori* which one ! We come back to this point in Sec. III below.

In view of Proposition 2.2, completely nondegenerate DZ sets are the key ingredients for solving the noncompact models, and in fact they are easy to construct with help of Proposition A3. Let $F = \{f_1,...,f_l\}$ be any set of linearly independent holomorphic neutral vectors: $\partial_- f_m = 0$ and $\langle f_m | f_m \rangle = 0$ for $m = 1,...,l$. Then every derivative $\partial_+{}^k f_m$ is also a holomorphic neutral vector; thus the set H generated by F is composed of neutral vectors, and so is every DZ set E contained in H. To get a completely nondegenerate DZ set E, it is therefore sufficient to verify the nonorthogonality asumption of Proposition A3; clearly this condition will be satisfied in general, especially if dim E is large.

For the sake of clarity, it is useful to extend to our noncompact model the terminology introduced by Din–Zakrzewski[13] and Sasaki.[14] Let again $H = \{G_1,E_1,G_2,..., E_s,G_{s+1}\}$ the set generated from $\{f_1,...,f_l\}$, P the solution corresponding to $E = \{E_1,...,E_s\}$. Then, that solution P is called (i) *generic* if $s = 1$ (i.e., $H = \{G_1,E_1,G_2\}$) and $l = r$; (ii) *degenerate* if $s = 1$ and $l < r$; (iii) *self-dual* if $s = 1$ and G_1 is empty: $H = \{E_1,G_2\}$; (iv) *anti-self-dual* if $s = 1$ and G_2 is empty: $H = \{G_1,E_1\}$; and (v) *reducible* in all other cases; indeed[14,21] a solution corresponding to $\{...,E_1,G_2,E_2,...\}$ is the direct sum of two solutions corresponding to E_1 and E_2, respectively.

Before closing this subsection, let us be more specific about the solutions. Let L stand for $\tilde{E}_n \tilde{G}_n, \tilde{Z}_n, \tilde{W}_n, \tilde{M}_n \equiv \{\tilde{G}_1,\tilde{E}_1,\tilde{G}_2,...,\tilde{E}_{n-1},\tilde{G}_n\}$ or $N_n \equiv \{\tilde{G}_1,\tilde{E}_1,\tilde{G}_2,...,\tilde{E}_n\}$. Then the following relations determine, respectively, an orthogonal basis of $\langle L \rangle$ and the corresponding orthonormal one,

$$\check{L} = L U_L , \tag{2.16a}$$

$${}^0L = L\,|t_L|^{-1/2} U_L , \tag{2.16b}$$

where $t_L \equiv L^\dagger t L$, $|t_L| \equiv (t_L^\dagger t_L)^{1/2}$, and U_L is a unitary matrix diagonalizing t_L,

$$\check{t}_L \equiv (t_L)_{\text{diag}} = U_L^\dagger t_L U_L = \check{L}^\dagger t \check{L} . \tag{2.17}$$

We will use these various bases in Sec. III below.

C. The Σ_m and Σ_1 transformations

For the $\mathbb{C}P^n$ model, Din and Zakrzewski[13] have found all the finite action solutions. For compact Grassmannian models, Sasaki[14] has proposed a different method. The idea is to start from an arbitrary solution and to construct a new one, orthogonal to the first, using the so-called Σ_m transformation. However, it is still not known whether the combined DZS methods yield all finite action solutions. Those have been constructed recently by Wood,[9] but it is not clear whether his solutions are all different from the DZS solutions. We show here that Sasaki's Σ_m transformation may be generalized and extended to the noncompact Grassmannian models, without, however, yielding a complete characterization of all finite action solutions.

Let $X^0 \in G_{pi;qj}(\mathbb{C})$ be a solution of rank $r = i + j$ and signature (i,j). Its covariant derivative, which is given as (see Ref. 10, §5)

$$D_+{}^0 X^0 = (1 - P_{X^0}) \partial_+ X^0 , \tag{2.18}$$

is a matrix of, say, rank m and signature (k,l). Sasaki assumes the rank of $D_+{}^0 X^0$ to be maximal, but it is enough to suppose that $D_+{}^0 X^0$ is nondegenerate, i.e., $k + l = m$. Without loss of generality, we may write $D_+{}^0 X^0$ under the form $D_+{}^0 X^0 = [Y; YA]$, where Y is of order $(p+q) \times m$ and of maximal rank, and A is of order $m \times (p+q-m)$. Since $D_+{}^0 X^0$ is assumed to be nondegenerate, it follows that $t_Y = Y^\dagger t Y$ is invertible. Let U_Y be a unitary matrix that diagonalizes t_Y and define $X^1 = Y |t_Y|^{-1/2} U_Y$, where $|t_Y| = (t_Y^\dagger t_Y)^{1/2}$. Then X^1 is an orthonormal basis of the subspace $\langle D_+{}^0 X^0 \rangle$ and, exactly as in Ref. 14, one shows that

$$\begin{aligned} D_-{}^1 X^1 &\equiv (1 - P_{X^1}) \partial_- X^1 \\ &= - X^0 t_{X^0} B\, |t_Y|^{1/2} U, \quad \text{where } B = [\mathbf{1}; A]^{\text{tr}} , \end{aligned} \tag{2.19}$$

and that X^1 is a solution of the model on $G_{pk;ql}(\mathbb{C})$. Notice that U is irrelevant in the compact case, since it is then a mere gauge transformation. If $D_+{}^0 X^0$ is of maximal rank, we recover Sasaki's result.

The Σ_m transformation just described is not sufficient to yield all solutions. This prompted Sasaki[14] to introduce another one, called the Σ_1 transformation, that we now extend to the noncompact case as well.

Let $Z = [z_1,...,z_r]$ $(r = i + j)$ be a solution of the $G_{pi;qj}(\mathbb{C})$ model. We assume Z to be in the *triangular gauge*, i.e.,

$$\langle z_k | z_m \rangle = 0, \quad \text{for } k > m . \tag{2.20}$$

If $D_+ z_1 \equiv (1 - P_Z)\, \partial_+ z_1$ is not neutral, we can define the vector

$$z_{r+1} = D_+ z_1 \cdot |D_+ z_1|^{-1} . \tag{2.21}$$

Then, as in Sasaki's paper,[14] *if* $Z' = [z_2,...,z_r,z_{r+1}]$ is in the triangular gauge, one can show that it is a solution of the $G_{pi';qj'}(\mathbb{C})$ model $(i' + j' = i + j)$, where i' and j' depend on the signature of z_1 and z_r. The proof, and all the remarks of Ref. 14, go over from the compact case to the noncompact one.

D. Solutions of the symplectic models

In the case of the symplectic models, with target spaces $G_{pi;qj}(\mathbb{H}) = \mathrm{Sp}(p,q)/\mathrm{Sp}(p-i,q-j)\times\mathrm{Sp}(i,j)$, the projection P must satisfy the additional constraint (see Ref. 10 for the notation)

$$P = -t\bar{J}\bar{P}Jt\,. \tag{2.22}$$

Then the self-dual equation $\partial_- P\cdot P = 0$ implies, by (2.22), the relation $\partial_+ P\cdot P = 0$ and therefore $\partial_+ P = P\cdot\partial_+ P$. Thus we get $\partial_- P = \partial_+ P = 0$, i.e., the only self-dual solutions of the symplectic models are constant. This is, of course, well-known, since the symplectic Grassmann manifold is not Kählerian.

In order to get nontrivial solutions, we may proceed as follows. Let $F = \{f_1,\ldots,f_l\}$ be a family of holomorphic vectors in $\mathbb{C}^{2N}$, $N = p+q$, and H be an extended set, in the sense of Eq. (2.2), of dimension at most N (so here we take $\langle H\rangle = \mathbb{C}^N$). Applying to this H the procedure of Sec. II B, we construct a solution of the $G_{2p,i;2q,j}$ model, for some i,j, namely a projection T of rank $(i+j)$ and signature (i,j) in $\mathbb{C}^{2N}$. Assume now that $\langle H\rangle$ and its conjugate $\langle\hat{H}\rangle$, where $\hat{H} = -Jt\bar{H}$, verify the following orthogonality conditions:

$$\hat{H}^{\dagger}t\,H = 0, \tag{2.23a}$$

$$\hat{H}^{\dagger}t\,\partial_{\pm}H = 0 \tag{2.23b}$$

(notice that $t_H = t_{\hat{H}}$, i.e., the two subspaces have the same dimension and signature). Then $\hat{T} = -Jt\,\bar{T}t\,J$ is another solution of the same model, orthogonal to T. It follows that $\partial_- T\cdot\hat{T} = \partial_-\hat{T}\cdot T = 0$, and therefore $P = T+\hat{T}$ satisfies condition (2.22). This means that P, a projection of rank $2(i+j)$ and signature $(2i,2j)$, is a solution of the symplectic model.

A practical way of implementing this construction is to start from a set F with special symmetry properties. For that purpose it is convenient to revert to the standard parametrization[11] of the $G_{pi;qj}(\mathbb{H})$ model, in which both t and J are independent of i and j:

$$t = K_{pq} = \mathrm{diag}[\mathbb{1}_p,-\mathbb{1}_q,\mathbb{1}_p,-\mathbb{1}_q]\,,\quad J = J_{p+q}\,,$$

$$I = \mathrm{diag}[\mathbb{1}_{p-i},-\mathbb{1}_i,\mathbb{1}_{q-j},-\mathbb{1}_j,\mathbb{1}_{p-i},-\mathbb{1}_i,\mathbb{1}_{q-j},-\mathbb{1}_j]\,.$$

In this parametrization, only the indices k and $N+k$ ($k = 1,\ldots,N$) are coupled by the matrix J; we call them *dual*. Let now the matrix $F\equiv[F_{am}]$ ($1\leqslant a\leqslant 2N$, $1\leqslant m\leqslant 1$) verify the following condition.

(S) For each row of F, one has either (i) it is identical to its dual row, or (ii) one of them consists of zeros only.

It is easy to check that the set H generated from such an F indeed verifies Eqs. (2.23), and thus yields solutions of the symplectic models.

We give some explicit examples. Let $p = 2$, $q = 1$, so that $J = J_3$, $t = \mathrm{diag}[\mathbb{1}_2,-1,\mathbb{1}_2,-1]$. Take for F the single vector $f^T(x_+) = (u,0,v,u,0,0)$, where $u = u(x_+)$ and $v = v(x_+)$ are two arbitrary holomorphic functions; this vector obviously verifies condition (S). Then the matrix $H = [f,\partial_+ f]$ is indeed orthogonal to $\hat{H} = -Jt\,\bar{H}$. Selecting $G_1 = \varnothing$, $E_1 = \{f\}$, $G_2 = \{\partial_+ f\}$, we obtain $T = \langle f|f\rangle^{-1}\cdot ff^{\dagger}t$, $\hat{T} = \langle\hat{f}|\hat{f}\rangle^{-1}\cdot\hat{f}\hat{f}^{\dagger}t$, where $\hat{f} = -Jt\bar{f}$, $\langle f|f\rangle = \langle\hat{f}|\hat{f}\rangle = 2|u|^2-|v|^2$, and $\langle f|\hat{f}\rangle = 0$. The resulting solution is the 6×6 matrix,

$$P = (2|u|^2-|v|^2)^{-1}\cdot\{ff^{\dagger}t+\hat{f}\hat{f}^{\dagger}t\}\,. \tag{2.24}$$

This solution is clearly singular whenever $2|u|^2 = |v|^2$, its rank is 2, and its signature is (2,0) or (0,2) depending on the sign of $\langle f|f\rangle = 2|u|^2-|v|^2$: if $\langle f|f\rangle > 0$, $(i,j) = (1,0)$ and P is a solution of the $\mathrm{Sp}(2,1)/\mathrm{Sp}(1,1)\times\mathrm{Sp}(1)$ model; and if $\langle f|f\rangle < 0$, $(i,j) = (0,1)$ and P is a solution of the $\mathrm{Sp}(2,1)/\mathrm{Sp}(2)\times\mathrm{Sp}(1)$ model. The generic form of this example is given by the vector $f^T = (u,v,w,u',v',w')$, where $u' = u$ or 0, $v' = v$ or 0, $w' = w$ or 0. It can easily be generalized to arbitrary p and q, with the following result: if $\langle f|f\rangle > 0$, $(i,j) = (1,0)$ and P is a solution of the $\mathrm{Sp}(p,q)/\mathrm{Sp}(p-1,q)\times\mathrm{Sp}(1)$ model; and if $\langle f|f\rangle < 0$, $(i,j) = (0,1)$ and P is a solution of the $\mathrm{Sp}(p,q)/\mathrm{Sp}(p,q-1)\times\mathrm{Sp}(1)$ model. Another class of examples is obtained by taking for F two nonorthogonal neutral vectors and applying the procedure described in the Appendix. For instance, in the case $p = q = 1$, one may start from the vectors $f^T = (u,u,0,0)$ and $g^T = (v,ve^{i\phi},0,0)$, where ϕ is a nonzero constant phase; then, indeed, $\langle f|f\rangle = \langle g|g\rangle = 0$ and $\langle f|g\rangle = uv(1-e^{i\phi})$, so that the assumption of Proposition A3 is satisfied. We leave the rest as an exercise to the reader.

III. PROPERTIES OF THE SOLUTIONS

As we have mentioned several times already, the solutions of the noncompact Grassmannian models constructed in Sec. II have a prescribed rank but their signature can only be ascertained *a posteriori*, through a calculation that involves the diagonalization of a potentially large matrix. In other words, we do not know to which model they belong! Actually, starting from a given set of holomorphic vectors, one may choose different partitions $H = \{G_1,E_1,\ldots,G_{s+1}\}$, which lead to solutions of different models. An explicit example is given in the Appendix, for rank 1 solutions.

In fact the situation is worse. Indeed, everything is a function of x_+, i.e., this is a *local* theory, as already pointed out by Din and Zakrzewski[13] and by Sasaki.[14] As we have seen in Sec. II, an easy way to get a well-defined solution is to construct a completely nondegenerate DZ set from a family of neutral vectors $F = \{f_1,\ldots,f_l\}$, provided that the nonorthogonality condition of Proposition A3 is satisfied. But it may happen that this is possible only *locally*, i.e., in some region $\Gamma\subset\mathbb{C}$. Then one gets a local solution that has no global extension. Furthermore, even if the solution is globally defined, the norm $\langle g|g\rangle = g^{\dagger}tg$ of some basis vector g from ${}^0E$ may vanish, necessarily along certain lines in the x_+ plane; there the solution becomes singular. Moreover the norm $\langle g|g\rangle$ may change sign across such a line, and then the signature of the solution P will change across that line. The only reasonable interpretation of this pathology seems to be to consider the element of $\mathrm{U}(p,q)$ corresponding to P, namely $Q = I_{p+q-i-j,i+j}(\mathbb{1}-2P)$, as a singular solution of the principal model, which reduces locally to solutions of some Grassmannian models on $G_{pi;qj}$.

Let us give an example. Take $G = \mathrm{U}(2,1)$, $t = I_{21}$, and

consider the rank 1 self-dual solutions generated from the vector $f^T(x_+) = (1, x_+^2, x_+\sqrt{2a})$, $a > 0$. The norm of the vector is $\langle f|f\rangle = |x_+|^4 - 2a|x_+|^2 + 1$. So there are three cases.

(i) $a < 1$: $\langle f|f\rangle > 0$ for all x_+; we get a global solution of positive signature, i.e., a solution of the model on $G_{21;10} = \mathrm{U}(2,1)/\mathrm{U}(1,1)\times\mathrm{U}(1)$, with finite action (see below).

(ii) $a = 1$: $\langle f|f\rangle = 0$ for $|x_+| = 1$, positive otherwise; we get a solution of the same model, but now singular on $|x_+| = 1$.

(iii) $a > 1$: writing $a = \cosh\lambda$, $\lambda > 0$, we get the following picture: (a) $\langle f|f\rangle > 0$ for $|x_+| < e^{-\lambda}$, $|x_+| > e^{\lambda}$: solution of the $G_{21;10}$ model; (b) $\langle f|f\rangle = 0$ for $|x_+| = e^{\pm\lambda}$: two lines of singularity; and (c) $\langle f|f\rangle < 0$ for $e^{-\lambda} < |x_+| < e^{\lambda}$: solution of the model on $G_{20;11} = \mathrm{U}(2,1)/\mathrm{U}(2)\times\mathrm{U}(1)$.

As indicated above, one should rather consider this as a solution of the principal U(2,1) model, which reduces locally to solutions of the two Grassmannian models, depending on the value of the parameter a.

Whatever point of view is adopted, one is interested mostly in finite action solutions; hence it is necessary to indicate how the action of our solutions can be computed. It is important to notice that the action $\mathcal{S}(Z) \equiv \int d^2\mathbf{x}\, \mathcal{L}(Z)$ corresponding to the Lagrangian (1.5) is in general *not* bounded below. However, when the Grassmann manifold is a *Riemannian* space, the action does have a definite sign: it is positive definite for the compact model (1.3a) and negative definite for the noncompact model (1.3b) with $i = 0, j = q$, that is, the $\mathrm{SU}(p,q)/\mathrm{S}(\mathrm{U}(p)\times\mathrm{U}(q))$ model. In the latter case, $|Z|^2$ is positive, hence $|\partial_\mu Z|^2$ must be negative, since they are orthogonal to each other, which proves our assertion. Thus the Feynman quantization procedure can be performed in this case, too; it suffices to change sign in the Lagrangian density (1.5). For pseudo-Riemannian manifolds, however, only the classical models seem viable.

In the compact case, the action is usually[13] computed in terms of the topological charge of the solution. But here we have to be more careful. Indeed, the σ model on a manifold M possesses a topological charge only if the second homotopy group $\pi_2(M)$ is nontrivial and discrete. When M is a symmetric space $M = G/H$, the group $\pi_2(M)$ may be computed from the (exact) homotopy sequence[22,23] of the principal bundle $G \to M$. If M is simply connected, the latter takes the following form: $0 \to \pi_2(M) \to \pi_1(H) \to \pi_1(G) \to 0$. Let now M be the noncompact Grassmann manifold $G_{pi;qj}(\mathbb{C}) = \mathrm{SU}(p,q)/\mathrm{S}(\mathrm{U}(p-i,q-j)\times\mathrm{U}(i,j))$. First, it is pseudo-Hermitian, hence simply connected.[24] Next we have $\pi_1(G) = \mathbb{Z}$. Thus the result depends only on $\pi_1(H)$. A straightforward analysis shows that three cases may arise, depending on the number of noncompact factors in $H = \mathrm{S}(\mathrm{U}(p-i,q-j)\times\mathrm{U}(i,j))$: (i) two non-compact factors: $\pi_1(H) = \mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}$, $\pi_2(M) = \mathbb{Z}\times\mathbb{Z}$; (ii) one non-compact factor: $\pi_1(H) = \mathbb{Z}\times\mathbb{Z}$, $\pi_2(M) = Z$; and (iii) both factors compact: $\pi_1(H) = Z$, $\pi_2(M) = 0$. Thus in cases (i) and (ii), which are pseudo-Riemannian, a topological charge may be defined. But in case (iii), once again the Riemannian one, it is not possible to define a topological charge for the corresponding model. Nevertheless, for computing the action of a solution Z of the $G_{pi;qj}$ model, it is useful to introduce the following quantity:

$$q(Z) = 2\,\mathrm{Tr}\left[(D_+Z)^\dagger t\,(D_+Z)t_Z - (D_-Z)^\dagger t\,(D_-Z)t_Z\right] \tag{3.1a}$$

$$= 2\,\mathrm{Tr}\left[(\partial_+Z)^\dagger t\,(\partial_+Z)t_Z - (\partial_-Z)^\dagger t\,(\partial_-Z)t_Z\right]. \tag{3.1b}$$

Although it is not of topological origin, $q(Z)$ can still be called a *charge*, since it takes only integer values, as will be shown below. It is interesting to note that the Riemannian case (iii) is the most frequent one in the applications; it shows up, for instance, in supergravity[4] and in solid state physics.[7,8] The reason is, presumably, that the corresponding Grassmannian is a Kähler manifold; and indeed, these play an increasingly important role in physical applications.[25]

Using the relation $D_\pm Z = (\partial_\pm P)Z$, with $P \equiv P_Z = Zt_Z Z^\dagger t$, we get

$$\mathrm{Tr}\left[(D_\pm Z)^\dagger t\,(D_\pm Z)t_Z\right] = \mathrm{Tr}\left[P\cdot\partial_+P\cdot\partial_\pm P\right]. \tag{3.2}$$

Let now $E = \{E_1,\ldots,E_s\}$ be a completely nondegenerate DZ set and Z_n, W_n the subsets defined in Sec. II, with projections P_n, Q_n, respectively. As in the proof of Proposition 2.2, one has, for each $n = 1,\ldots,s$,

$$\partial_-P_n\cdot P_n = -Q_n\cdot\partial_-Q_n, \quad \partial_+P_n\cdot Q_n = -\partial_+Q_n\cdot Q_n, \tag{3.3}$$

and therefore

$$\mathrm{Tr}\left[P_n\cdot\partial_+P_n\cdot\partial_-P_n\right] = \mathrm{Tr}\left[Q_n\cdot\partial_-Q_n\cdot\partial_+Q_n\right] = \mathrm{Tr}\left[P_{n-1}\cdot\partial_+P_{n-1}\cdot\partial_-P_{n-1}\right]. \tag{3.4}$$

Using (3.4), we may express the Lagrangian density of the solution Z_n in terms of the charge densities (3.1),

$$\begin{aligned}\mathcal{L}(Z_n) &= 2\,\mathrm{Tr}\left[(D_+Z_n)^\dagger t\,(D_+Z_n)\,t_{Z_n} + (D_-Z_n)^\dagger t\,(D_-Z_n)t_{Z_n}\right] \\ &= q(Z_n) + 2q(W_n) + 2q(Z_{n-1}) + 2q(W_{n-1}) + \cdots + 2q(W_1)\end{aligned} \tag{3.5}$$

[here we have assumed $G_1 \neq \varnothing$; otherwise the last term is $2q(Z_1)$]. Since the subspaces $\langle\tilde{E}_m\rangle, \langle\tilde{G}_k\rangle$ are mutually orthogonal, the charges are completely additive and we get

$$\mathcal{L}(Z_n) = \sum_{m=1}^{n}\left[(2n+1-2m)q({}^0E_m) + (2n+2-2m)q({}^0G_m)\right]. \tag{3.6}$$

To get the charge and the action, it remains to integrate $q({}^0E_m)$ and $q({}^0G_m)$ over the whole complex plane. In the compact case,[13] the Gram–Schmidt procedure is available. Here it is not in general (see the remarks below, however), so we have to make a detour.

From the definitions (2.16) and (2.17), we have

$$q({}^0E_m) = q({}^0N_m) - q({}^0M_m)\,, \tag{3.7a}$$

$$q({}^0G_m) = q({}^0M_m) - q({}^0N_{m-1})\,. \tag{3.7b}$$

Let g_k be the k th vector of $\check{M}_m$ and $e_k = g_k \cdot |\langle g_k | g_k \rangle|^{-1/2}$. Then a direct calculation shows[13] that [we write $\mathrm{U}(L)$ for U_L and $t(L)$ for t_L]

$$\begin{aligned}(\partial_+ e_k)^\dagger t\,(\partial_+ e_k) t_{e_k} - (\partial_- e_k)^\dagger t\,(\partial_- e_k) t_{e_k} \\ = \partial_+\partial_- \ln(g_k^\dagger t\, g_k t_{e_k}) + [(\partial_+ U(M_m))^\dagger \partial_+ U(M_m) \\ - (\partial_- U(M_m))^\dagger \partial_- U(M_m)]_{kk}\,.\end{aligned} \tag{3.8}$$

Using the cyclicity of the trace, one gets from there

$$q({}^0M_m) = \sum_k \partial_+\partial_- \ln(g_k^\dagger t\, g_k t_{e_k}) \tag{3.9a}$$

$$= \mathrm{Tr}[\partial_+\partial_- \ln \check{t}\,(M_m)]\,. \tag{3.9b}$$

Similarly

$$q({}^0N_m) = \mathrm{Tr}[\partial_+\partial_- \ln \check{t}(N_m)] \tag{3.9c}$$

and therefore, by (3.7),

$$q({}^0E_m) = \mathrm{Tr}[\partial_+\partial_- \ln \check{t}(E_m)]\,, \tag{3.10a}$$

$$q({}^0G_m) = \mathrm{Tr}[\partial_+\partial_- \ln \check{t}(G_m)]\,. \tag{3.10b}$$

The expression (3.10) for the charge density makes sense if ${}^0E_m$ is a global solution, for then the elements of the diagonal matrix $\check{t}(E_m)$ have no zero in the x_+ plane. In this case, the charge may be computed as in the compact case,

$$\mathcal{Q}(g_k) \equiv \int_{\mathbb{R}^2} dx_+\, dx_-\, \partial_+\partial_- \ln(g_k^\dagger t g_k t_{e_k}) = 2\pi\beta_k\,, \tag{3.11}$$

where β_k is defined by the asymptotic behavior of the matrix element,

$$g_k^\dagger t\, g_k \to |\mathbf{x}|^\beta k, \quad \text{for } |\mathbf{x}| \to \infty, \tag{3.12}$$

and the full charge is obtained by additivity from (3.10a) and (3.6).

Now it may happen that the Gram–Schmidt procedure *is* applicable for orthonormalizing H; a necessary and sufficient condition[19] is that the principal minors of $H^\dagger t H$ do not vanish. In this case one can also use the wedge product formalism.[12] Let $\{h_1, h_2, \ldots, h_N\}$ denote the vectors (columns) of H and define

$$\begin{aligned} h^{(m)} &= h_1 \wedge h_2 \wedge \cdots \wedge h_m \quad (1 \leqslant m \leqslant N)\,, \\ g^{(m)}{}_b &= \langle h^{(m-1)\dagger} \cdot h^{(m)} \rangle_b \\ &= [h^{(m-1)\dagger}]^{a_1 \cdots a_{n-1}} [h^{(m)}]_{a_1 \cdots a_{n-1} b}\,, \end{aligned}$$

where the contractions are performed with the metric t. Then the vectors $g^{(m)}$ form an orthonormal basis of $\langle H \rangle = \mathbb{C}^N$, equivalent to the one obtained by the Gram–Schmidt procedure and

$$\langle g^{(m)} | g^{(m)} \rangle = \langle h^{(m)} | h^{(m)} \rangle / m \langle h^{(m-1)} | h^{(m-1)} \rangle\,. \tag{3.13}$$

So, if $|\langle h^{(m)} | h^{(m)} \rangle| \to |\mathbf{x}|^{\alpha(m)}$, modulo an overall factor, when $|\mathbf{x}| \to \infty$, then, by (3.11), one has, up to constants,

$$\mathcal{Q}(g^{(m)}) = 2\pi[\alpha(m) - \alpha(m-1)]\,. \tag{3.14}$$

Then, inserting (3.14) into (3.5) or (3.6), we get the action

$$\begin{aligned} \mathcal{S}(Z_n) &\equiv \int d^2\mathbf{x}\, \mathcal{L}(Z_n) \\ &= 2\pi \sum_{m=1}^{n} [\alpha(u_m) + \alpha(v_m)]\,, \end{aligned} \tag{3.15}$$

where $g^{(u_m)}, g^{(v_m)}$ are the last columns of ${}^0E_m, {}^0G_m$, respectively. So whenever Gram–Schmidt is applicable, the action is given directly in terms of the starting vectors $\{f_m\}$. If it is not the case, one has to use (3.11) and the vectors $\{g_k\}$, i.e., evaluate first the columns of $\check{M}_m$.

In the compact case, the action is finite if and only if the starting vectors $\{f_m\}$ are polynomials in x_+. In the noncompact case, the condition is sufficient, if the solution is globally defined [as for the U(2,1) solution given above, in the case $a < 1$], but not necessary! Take, for instance, the $G_{p1,q0}$ model and start from the single vector,

$$g^T(x_+) = \{f(x_+), s^T(x_+); f(x_+), r^T(x_+)\}\,, \tag{3.16}$$

with $t \equiv I_{pq}$; s and r are vectors of length $p-1$ and $q-1$, respectively, with polynomial components and f an arbitrary holomorphic function. Then $\langle g|g\rangle = \langle s|s\rangle - \langle r|r\rangle$ is a polynomial in x_+, x_-. If $\langle g|g\rangle \neq 0$ (take, e.g., $r \equiv 0$), then the self-dual solution corresponding to the vector g is globally defined and has finite action, yet g is not necessarily polynomial. The same argument applies to non-self-dual solutions, as one can show easily.

IV. SUPERSYMMETRIC MODELS

Like many classical models, Grassmannian σ models admit supersymmetric extensions. In the compact case, a detailed description was given by Din *et al.*[15] and by Fujii *et al.*[16,17] Here again the analysis extends to the noncompact models in a straightforward way; hence our treatment will be very sketchy, following Refs. 16 and 17 almost verbatim.

Let Z be a solution of the bosonic $G_{pi;qj}(\mathbb{C})$ model, with projection $P_Z = Z t_Z Z^\dagger t$. The corresponding superfield reads, as usual,

$$\phi(x,\theta) = Z(x) + i\theta_\alpha \chi^\alpha(x) + (i/2)\theta\gamma_5\theta F(x)\,, \tag{4.1}$$

where the Grassmann variable $\theta = [\theta_1, \theta_2]^T$ is a real two-component spinor, $\chi(x)$ a two-component (anticommuting) spinor field, and F an auxiliary scalar field; γ_1, γ_2, and $\gamma_5 = \gamma_1\gamma_2$ are the familiar 2×2 γ matrices. In terms of the supercovariant derivative,

$$\nabla = \mathscr{D} - A\,, \tag{4.2a}$$

$$\mathscr{D} = \frac{\partial}{\partial\theta} + i\gamma_\mu\, \partial^{\mu\theta}\,, \tag{4.2b}$$

where A is a gauge superfield, the action of the model is

$$\mathcal{S} = \int d^2x\, d\theta^T \gamma_5\, d\theta\, \mathrm{Tr}[(\nabla\phi)^\dagger \gamma_5 t\, (\nabla\phi) t_Z]\,, \tag{4.3}$$

and it yields the equation of motion for A,

$$A = t_Z \phi^\dagger t \mathcal{D}\phi . \tag{4.4}$$

The action (4.3) may be written more explicitly as

$$\begin{aligned}\mathcal{S} = \int d^2\mathbf{x}\, \mathrm{Tr}\big[& (D_\mu Z)^\dagger t\,(D^\mu Z) t_Z - i\Psi^\dagger t \gamma_\mu D^\mu \Psi t_Z \\ & + \tfrac{1}{4}\{(\Psi^\dagger t \Psi t_Z)^2 + (\Psi^\dagger \gamma_5 \Psi t_Z)^2 \\ & - (\Psi^\dagger t \gamma_\mu \Psi t_Z)(\Psi^\dagger t \gamma^\mu \Psi t_Z)\} \\ & - (G^\dagger + \rho t_Z \gamma_5 \Psi^\dagger) t (G - \Psi \gamma_5 t_Z \rho) t_Z \big] ,\end{aligned} \tag{4.5}$$

where we have put

$$\Psi = (1 - P_Z)\chi , \tag{4.6a}$$

$$G = (1 - P_Z)F , \tag{4.6b}$$

$$\rho = (i/2)(\chi^\dagger t Z - Z^\dagger t \chi) . \tag{4.6c}$$

Actually the equation of motion for G implies the vanishing of the last term of (4.5), which we will drop henceforth.

A. The linearized fermion equation

Following Ref. 16 we treat first the linearized fermion equation in the fixed bosonic background Z, obtained by dropping the four-fermion interaction in (4.5), namely,

$$(1 - P_Z)\gamma_\mu D^\mu \Psi = 0, \quad Z^\dagger t \Psi = 0 , \tag{4.7}$$

where $Z_n = \{{}^0E_1, \ldots, {}^0E_n\}$ is a solution of the bosonic model with $G_1 \neq \varnothing$, thus non-self-dual (Sec. II). If we write

$$\Psi = \begin{bmatrix} 1 \\ -i \end{bmatrix} \Psi_{(+)} + \begin{bmatrix} 1 \\ i \end{bmatrix} \Psi_{(-)} , \tag{4.8}$$

the equation of motion (4.7) decouples:

$$(1 - P_Z) D_\pm \Psi_{(\pm)} = 0 , \quad \chi^\dagger t \Psi_{(\pm)} = 0 . \tag{4.9}$$

Then, exactly as in Ref. 16, the solutions of Eqs. (4.9) are easily found to be

$$\Psi_{(+)} = P[G_1]\,[\Gamma^1_{(+)} a_1^\dagger; \Gamma^2_{(+)} a_2^\dagger; \ldots \Gamma^n_{(+)} a_n^\dagger] , \tag{4.10a}$$

$$\begin{aligned}\Psi_{(-)} = & (1 - P[H(E_n)]) \\ & \times [\Gamma^1_{(-)} a_1^{-1} t\,({}^0E_1); \cdots \Gamma^2_{(-)} a_n^{-1} t\,({}^0E_n)] ,\end{aligned} \tag{4.10b}$$

where

$$a_j = ({}^0E_j)^\dagger t H(E_n), \quad j = 1,2,\ldots,n , \tag{4.11}$$

and $\Gamma^j_{(\pm)}$ are *arbitrary* (anti-)holomorphic Grassmann-valued matrices

$$\partial_+ \Gamma^j_{(+)} = 0, \quad \partial_- \Gamma^j_{(-)} = 0 . \tag{4.12}$$

B. The general equation

Going back to the full action (4.5), we rewrite it, as in Ref. 17, using the $x_\pm$ coordinates:

$$\begin{aligned}\mathcal{S} = \int d^2\mathbf{x}\, \mathrm{Tr}\big[& 2(D_+ Z)^\dagger t (D_+ Z) t_Z \\ & + 2(D_- Z)^\dagger t\,(D_- Z) t_Z \\ & - 4i(\Psi^\dagger_{(+)} t D_- \Psi_{(-)} + \Psi^\dagger_{(-)} t D_+ \Psi_{(+)}) t_Z \\ & + 4\Psi^\dagger_{(+)} t \Psi_{(+)} t_Z \Psi^\dagger_{(-)} t \Psi_{(-)} t_Z \\ & - 4\Psi^\dagger_{(+)} t \Psi_{(-)} t_Z \Psi^\dagger_{(-)} t \Psi_{(+)} t_Z \big] ,\end{aligned} \tag{4.13}$$

with the constraint term

$$\begin{aligned}\mathcal{S}_c = \int d^2\mathbf{x}\, \mathrm{Tr}\big[& \lambda(Z^\dagger t Z t_Z - 1) + \mu Z^\dagger t \Psi_{(+)} \\ & + \Psi^\dagger_{(+)} t Z \mu^\dagger + \nu Z^\dagger t \Psi_{(-)} + \Psi^\dagger_{(-)} t Z \nu^\dagger \big] ,\end{aligned}$$

where λ, μ, ν are matrix Lagrange multipliers. Minimizing $\mathcal{S} + \mathcal{S}_c$, one gets the Euler–Lagrange equations

$$\begin{aligned} & D_+ D_- Z + Z t_Z (D_- Z)^\dagger t\,(D_- Z) - i\{(D_+ Z) t_Z \Psi^\dagger_{(-)} t \Psi_{(+)} + (D_- Z) t_Z \Psi^\dagger_{(+)} t \Psi_{(-)}\} \\ & \quad + i\Psi_{(+)} t_Z \Psi^\dagger_{(-)} t \partial_+ Z + i \Psi_{(-)} t_Z \Psi^\dagger_{(+)} t\, \partial_- Z = 0 , \\ & (1 - P) D_- \Psi_{(-)} + i\{\Psi_{(+)} t_Z \Psi^\dagger_{(-)} t \Psi_{(-)} - \Psi_{(-)} t_Z \Psi^\dagger_{(-)} t \Psi_{(+)}\} = 0 , \\ & (1 - P) D_+ \Psi_{(+)} + i\{\Psi_{(-)} t_Z \Psi^\dagger_{(+)} t \Psi_{(+)} - \Psi_{(+)} t_Z \Psi^\dagger_{(+)} t \Psi_{(-)}\} = 0 .\end{aligned} \tag{4.14}$$

In the compact case, Fujii *et al.*[17] have given explicit solutions of the supersymmetric equations of motion, under the assumption that the fermion field be treated as a commuting (c-number) field. The same restriction was imposed already by Din *et al.*[15] for the $\mathbb{C}P^n$ model. Here we will extend the method of Ref. 17 to the general case and exhibit an explicit solution to Eqs. (4.14), with the same proviso that $\Psi_{(\pm)}$ are commuting fields. As in Sec. II we start with a set of holomorphic vectors and build the set $H = \{H^1, \ldots, H^k\}$ using the procedure (2.1). Then introduce a further partition of the even-numbered blocks,

$$H^{2j} = \{F^{(-1)}_j, F^{(0)}_j\} , \tag{4.15}$$

where the only restriction is that the number of vectors in $F^{(-1)}_j$ must be smaller or equal to that in $F^{(0)}_j$. Upon defining $F^{(1)}_j \equiv \partial_+ F^{(-1)}_j$, the odd-numbered blocks take the form

$$H^{2j+1} = \{F^{(1)}_j, G_{2j+1}\} . \tag{4.16}$$

Thus we get for the basic set of holomorphic vectors ($j = 1,2,...,n$),

$$H = \{G_1; F_1^{(-1)}, F_1^{(0)}, F_1^{(1)}; G_3; F_2^{(-1)}, ..., G_{2j-1}; F_j^{(-1)}, F_j^{(0)}, F_j^{(1)}; G_{2j+1}; ...\} . \quad (4.17)$$

Assuming that the partition (4.17) is completely nondegenerate, we orthogonalize it,

$$\begin{aligned} Y_j^{(-1)} &= (1 - P[H(G_{2j-1})])F_j^{(-1)} , \\ Y_j^{(0)} &= (1 - P[H(F_j^{(-1)})])F_j^{(0)} , \\ Y_j^{(1)} &= (1 - P[H(F_j^{(0)})])F_j^{(1)} , \end{aligned} \quad (4.18)$$

and normalize the resulting blocks,

$$X_j^{(k)} = Y_j^{(k)} |t(Y_j^{(k)})|^{-1/2} U(Y_j^{(k)}), \quad k = -1,0,1 . \quad (4.19)$$

Then it is easy to show, exactly as in Ref. 17, that the following quantities are solutions of the general supersymmetric equations (4.14):

$$\Psi_{(\pm)} = [\Psi_{(\pm)}^1; \Psi_{(\pm)}^2; ..., \Psi_{(\pm)}^n] ,$$

$$Z = [Z^1; Z^2; ..., Z^n] , \quad (4.20)$$

where

$$\Psi_{(+)}^j = [\mathbb{O}_j; c_{+j} X_j^{(-1)} (a_j^{(-1)\dagger})^{-1}] a_j^{(0)\dagger} ,$$

$$\Psi_{(-)}^j = [\mathbb{O}_j c_{-j} X_j^{(1)}) t(X_j^{(1)}) a_j^{(1)}] (a_j^{(0)})^{-1} t(X_j^{(0)}) ,$$

$$Z^j = X_j^{(0)} , \quad (4.21)$$

and

$$a_j^{(k)} = X_j^{(k)\dagger} t F_j^{(k)}, \quad k = -1,0,1 . \quad (4.22)$$

In Eq. (4.21) $\mathbb{O}_j$ denotes the $(p+q) \times m_j$ null matrix, with $m_j = \dim\langle X_j^{(0)}\rangle - \dim\langle X_j^{(-1)}\rangle$, and $c_{\pm j}$ are arbitrary constants subject to the constraint

$$\overline{c_{+j}} \cdot c_{-j} = i . \quad (4.23)$$

The proof that these are indeed solutions follows exactly the one of Ref. 17 and we shall not repeat it. Our solutions are more general, in that they allow arbitrary degenerate solutions for the bosonic sector. The crucial property that enables us to obtain this generalization is the following relation:

$$(Y_i^{(k)})^\dagger t(\partial_+ Y_j^{(m)}) = 0, \quad \text{for } i \neq j, \quad k,m = -1 \text{ or } 0 , \quad (4.24)$$

which results in a factorization of the problem.

ACKNOWLEDGMENTS

It is a pleasure to thank W. J. Zakrzewski, D. Lambert, and L. Woodward for many stimulating discussions and suggestions about the contents of this paper.

APPENDIX: ORTHOGONAL BASES IN INDEFINITE INNER PRODUCT SPACES

The treatment of noncompact Grassmannian σ models requires some facts about the existence and construction of orthogonal bases in a complex vector space with an indefinite metric. Part of those results are, of course, standard and may be found, for instance, in the monographs of Bognar[18] and Deheuvels.[19] Some statements, however, seem to be new, especially Proposition A3 below.

Throughout this Appendix we consider the space $V \equiv \mathbb{C}^N$ with a nondegenerate inner product. For any two vectors $f, g \in \mathbb{C}^N$, we denote their inner product as

$$\langle f|g\rangle = f^\dagger t g , \quad (A1)$$

where t is a Hermitian $N \times N$ matrix, and the inner product is assumed to be *indefinite*, so that V contains *neutral* vectors ($\langle f|f\rangle = 0$), and *nondegenerate*, i.e., t is invertible or, equivalently, $V \cap V^\perp = \{0\}$. A family of vectors $\{e_j\} \in V$ is called *orthogonal* if $\langle e_i|e_j\rangle = 0$ for $i \neq j$, and *orthonormal* if $|\langle e_i|e_j\rangle| = \delta_{ij}$.

Now we turn to subspaces of V. Let $\{f_1,...,f_r\}$ be a set of r linearly independent vectors, $F = [f_1,...,f_r]$ the corresponding $N \times r$ matrix, and $W \equiv \langle F\rangle$ the r-dimensional subspace it generates. Then the following results are standard.[18,19]

Proposition A1: The subspace $W = \langle F\rangle$ always has an orthogonal basis X. Furthermore the following three conditions are equivalent: (i) W has an orthonormal basis; (ii) W is nondegenerate: $W \cap W^\perp = \{0\}$; and (iii) the $r \times r$ matrix $t_F = F^\dagger t F$ is invertible. ■

The construction of those bases runs as follows. Since t_F is Hermitian, it may be diagonalized by a unitary matrix U: $U^\dagger t_F U = \check{t}_F$, with $\check{t}_F$ diagonal. Then $X = FU$ is an orthogonal basis of $\langle F\rangle$. Some vectors of X may have zero norm, however. This is excluded iff $\check{t}_F$, hence t_F, is invertible. Then the orthonormal basis associated with X is

$$Z = F|t_F|^{-1/2} U , \quad (A2)$$

where $|t_F| = (t_F^\dagger t_F)^{1/2}$. One has indeed, by direct verification,

$$Z^\dagger t Z = \check{t}_F \cdot |\check{t}_F|^{-1} = \mathrm{sgn}(\check{t}_F) , \quad (A3)$$

that is, a diagonal matrix whose entries equal ± 1. By a suitable permutation of rows and columns, (i.e., by choosing a different U), one may write

$$Z^\dagger t Z = I_{ij} = \mathrm{diag}[1_i, -1_j] \quad (i + j = r) ,$$

where (i,j) is the signature of the r-dimensional subspace $\langle F\rangle$.

An *orthogonal projection* in $V = \mathbb{C}^N[t]$ is a symmetric, idempotent operator P, i.e., a $N \times N$ matrix such that

$$P^\dagger t = tP, \quad P^2 = P . \quad (A4)$$

Then, since V is nondegenerate and finite dimensional, we have the following.[18,19]

Proposition A2: Given a subspace $W = \langle F\rangle$, the following three conditions are equivalent: (i) W is nondegenerate: $W \cap W^\perp = \{0\}$; (ii) W is the range of an orthogonal projection; and (iii) W is orthocomplemented: $V = W \oplus W^\perp$. ■

In that case, the projection P on W is given by $P = F t_F^{-1} F^\dagger t$, with $t_F = F^\dagger t F$. Its rank is $r = \dim\langle F\rangle$ and the signature is (i,j) with $i + j = r$ as above. Then, of course, $W^\perp = (1 - P)V$ is also nondegenerate. For $r = 1$, in particular, the projection on the one-dimensional subspace spanned by the vector f is

$$P[f] = \langle f|f\rangle^{-1}\cdot ff^{\dagger}t. \tag{A5}$$

Now we describe the explicit construction of orthogonal bases. The trouble with formula (A2) is that it requires diagonalizing the matrix t_F, and this becomes cumbersome when r is not small. In practice one has to proceed differently and use a step-by-step approach. In the positive definite case, the Gram–Schmidt procedure is all one needs and it is always available. In the indefinite case, however, the method is applicable only if all principal minors of t_F are nonvanishing,[19] which is rather exceptional. Thus a different algorithm is required. The one we present below is based on the following property, which is rather surprising, and apparently new.

Proposition A3: Let $F = \{f_1,...,f_k\}$ be a set of linearly independent *neutral* vectors, none of which is orthogonal to all the other ones. Then the subspace $\langle F\rangle$ spanned by F is nondegenerate: $\langle F\rangle \cap \langle F\rangle^{\perp} = \{0\}$.

Proof: Choose j such that $\langle f_1|f_j\rangle \neq 0$ We may assume that $\langle f_1|f_j\rangle$ is not purely imaginary, otherwise we multiply f_1 by a phase. Define $h_1 = f_1 + f_j$. Then $\langle h_1|h_1\rangle \neq 0$ and $P_1 \equiv P[h_1]$ is well-defined by (A5). Defining the vectors $\tilde{f}_m = (1 - P_1)f_m$, $m = 2,3,...,k$, we get a new set $F_1 = \{h_1, \tilde{f}_2,..., \tilde{f}_k\}$, where each $\tilde{f}_m$ is orthogonal to h_1. The vector $\tilde{f}_m$ need not be neutral, but if it is, it cannot be orthogonal to all the other ones. Suppose the contrary: $\langle \tilde{f}_i|\tilde{f}_i\rangle = 0$ and $\langle \tilde{f}_i|\tilde{f}_m\rangle = 0$, for each $m = 2,...,k$. Since f_i is neutral, the first condition implies $\langle f_i|P_1 f_i\rangle = 0$, hence $\langle f_i|h_1\rangle = 0$, and therefore, by the second condition, $\langle f_i|f_n\rangle = 0$, for every $n = 1,2,...,k$, which contradicts the assumption of the proposition. Now we may proceed, according to the following alternative.

(A) If every $\tilde{f}_m$, $m = 2,...,k$, is neutral, the assumption of the proposition is satisfied and we may repeat the argument above.

(B) If, say, $\tilde{f}_l$ is not neutral, define $h_2 = \tilde{f}_l$, $P_2 \equiv P[h_2]$, and $\hat{f}_m = (1 - P_2)\tilde{f}_m$, $m = 2,...,k$, $m \neq l$.

In either case, after relabeling the vectors $\hat{f}_m$'s, we get a new set $F_2 = \{h_1,h_2,g_3,...,g_k\}$, where h_1 and h_2 are non-neutral and orthogonal to each other, each g_m is orthogonal to h_1 and h_2, and no neutral g_m is orthogonal to all vectors of F_2. Iterating the process, using either (A) or (B) above, we end up with an orthogonal basis $\{h_1,...,h_k\}$, with $\langle h_m|h_m\rangle \neq 0$ for every m; in other words, $\langle F\rangle$ is nondegenerate, with orthonormal basis $\{e_1,...,e_k\}$, where $e_m = h_m|\langle h_m|h_m\rangle|^{-1/2}$.

Remark: Instead of $\tilde{f}_j = (1 - P_1)f_j$, one may also use the following vector, which is proportional to $\tilde{f}_j$, but easier to compute: $\check{f}_j = f_1 - \langle f_j|f_1\rangle\langle f_1|f_j\rangle^{-1}f_j$.

With the help of Proposition A3, we are finally ready to construct an orthogonal basis for an arbitrary subspace. It suffices obviously to consider a set of linearly independent vectors $F = \{e_1,...,e_k,g\}$, where $E \equiv \{e_1,...,e_k\}$ is an orthogonal family. Denote by I, (resp. J) the subsets of E consisting of neutral (resp. non-neutral) vectors. Then the subspace $\langle J\rangle$ is nondegenerate, with projection P, and the vector $\tilde{g} = (1 - P)g$ is orthogonal to J. Define further

$$I' = \{e_i \in I,\ \langle e_i|\tilde{g}\rangle \neq 0\}, \quad I'' = \{e_i \in I,\ \langle e_i|\tilde{g}\rangle = 0\}.$$

If $\tilde{g}$ is neutral, we may apply Proposition A3 to the set $I' \cup \{\tilde{g}\}$, and get an orthonormal basis H. If $\tilde{g}$ is not neutral, we do the same, using branch (B) of the alternative. In either case, $X = J \cup H \cup I''$ is an orthogonal basis of $\langle F\rangle$, with neutral part I''. Clearly $\langle F\rangle$ is nondegenerate iff $I'' = \varnothing$.

Using this construction repeatedly, we obtain a step-by-step algorithm that is a genuine substitute for the Gram–Schmidt procedure, this one valid for an *arbitrary* subspace.

We give two examples of the orthogonalization procedure.

Example 1: We take first $p = 2$, $q = 1$, $t = I_{21}$ and start with the single holomorphic vector $f^T(x_+) = (1, x_+^2, x_+\sqrt{2})$. Then our basic set is $H = \{f, \partial_+ f, \partial_+^2 f\}$. This is a case where the Gram–Schmidt procedure works. From H we obtain an orthogonal family $\{h_1 = f, h_2, h_3\}$, with square lengths,

$$\langle h_1|h_1\rangle = (1 - |x_+|^2)^2,$$
$$\langle h_2|h_2\rangle = -2,$$
$$\langle h_3|h_3\rangle = 4(1 - |x_+|^2)^{-2}.$$

The corresponding projections are

$$P_m = \langle h_m|h_m\rangle^{-1}\cdot h_m h_m^{\dagger}t = \langle h_m|h_m\rangle^{-1}\cdot(|h_m\rangle\langle h_m|).$$

From the single vector f, we may get three different solutions of rank 1.

(i) With $G_1 = \varnothing$, $E_1 = \{f\}$, $G_2 = \{\partial_+ f, \partial_+^2 f\}$, we obtain the *self-dual* solution P_1, which is singular on the circle $|x_+| = 1$; the signature is positive everywhere, i.e., it is a solution of the model on $G_{21,10} = \mathrm{U}(2,1)/\mathrm{U}(1,1)\times\mathrm{U}(1)$.

(ii) With $G_1 = \{f\}$, $E_1 = \{\partial_+ f\}$, $G_2 = \{\partial_+^2 f\}$, we obtain the *generic* solution P_2; this one is regular everywhere (it is a global solution) and its signature is negative; in other words, it is a solution of the model on the Riemannian symmetric space $G_{20,11} = \mathrm{U}(2,1)/\mathrm{U}(2)\times\mathrm{U}(1)$.

(iii) Finally, with $G_1 = \{f, \partial_+ f\}$, $E_1 = \{\partial_+^2 f\}$, we get the *anti-self-dual* solution P_3, also singular on $|x_+| = 1$, and with positive signature; of course, $P_1 + P_2 + P_3 = 1$.

The (anti-)self-dual solutions P_1, P_3 are the noncompact equivalent of the two-(anti-)instantons[12]: up to normalization, the corresponding vectors have polynomial components of degree at most two, in x_+ and x_-, respectively, but their action is infinite.

Example 2: Consider now the model on $G_{21,21} = \mathrm{U}(2,2)/\mathrm{U}(1,1)\times\mathrm{U}(1,1)$, with $t = I_{22}$, and start with the two neutral vectors,

$$f^T = (r, r+1, r, r+1), \quad g^T = (s, s+1, s+1, s),$$

where $r = r(x_+)$, $s = s(x_+)$ are two arbitrary polynomials. One has

$$\langle f|f\rangle = \langle g|g\rangle = 0, \quad \langle f|g\rangle = 1.$$

Gram–Schmidt is inapplicable, but the algorithm of Proposition A3 may be used. We define indeed

$$h = f + g,$$
$$\check{h} = f - \langle g|f\rangle\langle f|g\rangle^{-1}g = f - g,$$

and check

$$\langle h | h \rangle = 2, \quad \langle \check{h} | \check{h} \rangle = -2, \quad \langle h | \check{h} \rangle = 0 .$$

The corresponding self-dual solution of rank 2 has indeed signature (1,1) and it is globally defined.

[1]A. Neveu and P. Papanicolaou, Commun. Math. Phys. **53**, 31 (1978).
[2]V. V. Nesterenko, Dubna preprints JINR-E2-82-745 and JINR-E2-82-761; Lett. Math. Phys. **7**, 287 (1983); M. Arik and F. Neyzi, J. Math. Phys. **29**, 594 (1988).
[3]P. O. Mazur, J. Phys. A: Math. Gen. **15**, 3173 (1982); Acta Phys. Pol. B **14**, 219 (1983); Phys. Lett. A **100**, 341 (1984); P. Breitenlohner, D. Maison, and G. Gibbons, "Four-dimensional black holes from Kaluza–Klein theories," preprint MPI-PAE/PTh27/87–LPTENS87/9.
[4]E. Cremmer and B. Julia, Nucl. Phys. B **159**, 141 (1979); J. Ellis, C. Kounas, and D. V. Nanopoulos, *ibid.* B **247**, 373 (1984).
[5]A. Kundu, J. Phys. A: Math. Gen. **19**, 1303 (1986).
[6]See the recent review of A. M. Perelomov, Phys. Rep. **146**, 135 (1987).
[7]L. Schäfer and F. Wegner, Z. Phys. B **38**, 113 (1980); S. Hikami, Phys. Rev. B **24**, 2175 (1981); K. B. Efetov, Adv. Phys. **32**, 53 (1983).
[8]R. Oppermann, Nucl. Phys. B **280** [FS **18**], 753 (1987).
[9]F. E. Burstall and J. C. Wood, J. Diff. Geom. **23**, 255 (1986); J. C. Wood, "The explicit construction and parametrization of all harmonic maps from the two-sphere to a complex Grassmannian," University of Leeds preprint, 1986.
[10]J.-P. Antoine and B. Piette, J. Math. Phys. **28**, 2753 (1987).
[11]J. Harnad, Y. Saint-Aubin, and S. Shnider, Commun. Math. Phys. **92**, 329 (1984); **93**, 33 (1984).
[12]H. J. Borchers and W. D. Garber, Commun. Math. Phys. **72**, 77 (1980).
[13]For a review and references to earlier papers, see W. J. Zakrzewski, in "Integrable quantum field theories (Proc. Tvärminne 1981)," edited by J. Hietarinta and C. Montonen, *Lecture Notes in Physics*, Vol. 151 (Springer, Berlin, 1982), pp. 160–188; W. J. Zakrzewski, J. Geom. Phys. **1**, 39 (1984).
[14]R. Sasaki, Z. Phys. C **24**, 163 (1984).
[15]A. M. Din, J. Lukierski, and W. J. Zakrzewski, Nucl. Phys. B **194**, 157 (1982).
[16]K. Fujii, T. Koikawa, and R. Sasaki, Prog. Theor. Phys. **71**, 388 (1984).
[17]K. Fujii and R. Sasaki, Prog. Theor. Phys. **71**, 3831 (1984).
[18]J. Bognar, *Indefinite Inner Product Spaces* (Springer, Berlin, 1974).
[19]R. Deheuvels, *Formes quadratiques et groupes classiques* (PUF, Paris, 1981).
[20]A. J. Macfarlane, Phys. Lett. B **82**, 239 (1979).
[21]A. M. Din and W. J. Zakrzewski, Lett. Math. Phys. **5**, 505 (1983).
[22]B. O'Neill, *Semi-Riemannian Geometry with Applications to Relativity* (Academic, New York, 1983), especially Chap. 11.
[23]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, NJ, 1951).
[24]R. A. Shapiro, Comment. Math, Helv. **46**, 529 (1971).
[25]See, for instance, the review of M. Bordemann, M. Forger, and H. Römer, Commun. Math. Phys. **102**, 605 (1986).

Quantization of conformally invariant Bargmann–Wigner equations with gauge freedom

W. F. Heidenreich
Institut für Theoretische Physik A, TU Clausthal, West Germany

M. Lorente
Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

(Received 19 January 1988; accepted for publication 16 March 1988)

Massless, conformally invariant Bargmann–Wigner equations with gauge freedom are found. An invariant indefinite metric quantization is described.

I. INTRODUCTION

The opinion gains ground that if local fields are fundamental, they should be massless. Having no scale, their space-time invariance group should be the conformal group SU(2,2). Another feature peculiar to massless particles is their gauge freedom, as most prominently documented in the case of electrodynamics. This in turn is considered as a major tool to formulate interactions between the fields.

Massless integer spin fields are customarily described by employing fully symmetric tensors.[1,2] For mass 0 they have a gradient type gauge freedom, like $A_\nu \to A_\nu + \partial_\nu \Lambda$. These equations are in general not conformally invariant, even if the corresponding equations for field strengths are. The irreducible massless representations of the Poincaré group can be extended to the conformal group, but not the indecomposable ones that appear in the gauge theories.

Manifestly conformally invariant gauge theories using tensors for spin 1 and 2 have been formulated[3,4] but to do the general case requires new techniques for dealing with indecomposable representations.

Half-integer spin fields can be described very similarly using Rarita–Schwinger fields.[5]

A different approach uses fully symmetric multispinors. The massless Bargmann–Wigner equations[6] describe particles in an irreducible representation, without gauge freedom; they are also conformally invariant.[7] For spin $\frac{1}{2}$ another formulation is possible, which does have gauge freedom of the current type.[3] It was originally found and discussed in conformal space.[8–11] The object of this paper is to extend this approach to arbitrary spins. For Poincaré-invariant Bargmann–Wigner types of equations with gauge freedom (see Ref. 12).

Our notation will be the same as in Ref. 10. The quantum numbers to label states are given by the maximal compact subgroup $U(1)\times SU(2)\times SU(2)$ of a covering of the conformal group. The U(1) eigenvalues are called (conformal) energy; the SU(2) eigenvalues (j_i,m_i), $i=1,2$, give the spin properties of a state.

Irreducible lowest weight representations can be uniquely labeled as $D(E_0,j_1,j_2)$ by the quantum numbers of the lowest weight. The representations with $E_0-(j_1+j_2)=1$, $j_1\cdot j_2=0$ give, when restricted to the Poincaré group, the massless representations with helicity $\lambda=j_1-j_2$.

In Sec. II we first review conformal space and some properties of the representations used here. Then we find manifestly conformally invariant field equations for all spins $s\geqslant\frac{1}{2}$. They have a current type gauge freedom. In Sec. III we find invariant two-point functions of these fields, which give an invariant indefinite metric quantization. Section IV transforms the field equations in flat Minkowski space so that comparison with earlier work is simplified. The principal tools for Secs. II and III are the tensor products between certain finite and infinite representations of the conformal group, which are reduced in Appendix A. Appendix B gives some formulas used in calculations.

II. WAVE EQUATIONS FOR MASSLESS CONFORMAL MULTISPINORS

A. Transformation of fields in conformal space

The nonlinear action of the conformal group on flat space causes unnecessarily complicated transformation properties of fields over Minkowski space. Therefore we first prefer to discuss manifestly conformally invariant field equations in conformal space, which will be transformed later (see Sec. IV) to flat space.

On four-dimensional conformal space

$$x^2=\eta^{ab}x_a x_b=x_1^2+x_2^2+x_3^2-x_4^2+x_5^2-x_6^2=0,$$
$$x_a\equiv\lambda x_a,\quad \lambda>0, \tag{1}$$

the conformal group acts linearly.[8,13] To avoid any dynamics along the rays, we fix the degree of homogeneity of all fields by imposing

$$x\cdot\partial\Psi\equiv x^a\partial_a\Psi=n\Psi. \tag{2}$$

The scalar fields with positive energy and degree $n<-1$ carry a representation $D(-n,0,0)$.

As there are two inequivalent fundamental (four-dimensional) representations $D(-\frac{1}{2},\frac{1}{2},0)$ and $D(-\frac{1}{2},0,\frac{1}{2})$ of SU(2,2), we have two inequivalent spinor fields which we denote by $\Psi_a(x)$ and $\Psi_{\dot a}(x)$. The generators of the conformal group are

$$\begin{aligned} J_{ab}&=M_{ab}+S_{ab},\\ M_{ab}&=-i(x_a\partial_b-x_b\partial_a),\\ S_{ab}&=-(i/4)(\beta_a\gamma_b-\beta_b\gamma_a)\end{aligned} \tag{3}$$

on the undotted spinors. For dotted spinors, the spin part of the generators is

 0022-2488/88/071698-07$02.50

$$\dot{S}_{ab} = -(i/4)(\gamma_a \beta_b - \gamma_b \beta_a). \tag{4}$$

The 4×4 matrices β,γ have to satisfy

$$\beta_a \gamma_b + \beta_b \gamma_a = 2\eta_{ab}. \tag{5}$$

A possible realization is

$$\beta_i = \gamma_i = \begin{pmatrix} 0 & i\sigma_i \\ -i\sigma_i & 0 \end{pmatrix}, \quad i = 1,2,3,$$
$$\beta_4 = \gamma_4 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \beta_5 = \gamma_5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{6}$$
$$\beta_6 = -\gamma_6 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

On multispinors $\Psi_{\alpha_1\cdots\alpha,\dot{\alpha}_1,\cdots\dot{\alpha}_s}$, the spin part is

$$S_{ab} = \sum_{i=1}^{r} \underset{(i)}{S_{ab}} + \sum_{i=1}^{s} \underset{(i)}{\dot{S}_{ab}},$$

where $\underset{(i)}{S_{ab}}$ acts on the ith undotted and $\underset{(i)}{\dot{S}_{ab}}$ on the ith dotted index.

Massless particles with helicity $\lambda = \frac{1}{2}$ (resp. $\lambda = -\frac{1}{2}$) can be described with undotted (resp. dotted) spinor fields. Therefore we expect to describe massless particles with helicity $\lambda > 0$ by fully symmetric multispinor fields $\Psi_\lambda(x) \equiv \Psi_{(\alpha_1\cdots\alpha_{2\lambda})}(x)$ and those with helicity $\lambda < 0$ by $\Psi_\lambda(x) \equiv \Psi_{(\dot{\alpha}_1\cdots\dot{\alpha}_{|2\lambda|})}(x)$.

In the sequel we will give details only for the $\lambda > 0$ cases. The $\lambda < 0$ cases can be obtained by exchanging dotted and undotted spinors, β and γ, j_1 and j_2, and by replacing λ by $-\lambda$.

The constant multispinor $\Psi_{(\alpha_1\cdots\alpha_{2|\lambda|})}$ carries a finite representation $D(-\lambda,\lambda,0)$ of the conformal group. So the field $\Psi_\lambda(x)$ with degree of homogeneity $n < -1$ carries the tensor product

$$D(-\lambda,\lambda,0) \otimes D(-n,0,0). \tag{7}$$

In Appendix A it is shown that this tensor product contains for $n = -(2\lambda + 1)$ the Gupta–Bleuler triplet

$$D(2+\lambda,\lambda-\tfrac{1}{2},\tfrac{1}{2}) \to D(1+\lambda,\lambda,0) \to D(2+\lambda,\lambda-\tfrac{1}{2},\tfrac{1}{2}). \tag{8}$$

We will refer to the representations (and their states) as

scalar → physical → gauge.

The physical massless helicity λ representation $D(1+\lambda,\lambda,0)$ has gauge freedom $D(2+\lambda,\lambda-\frac{1}{2},\frac{1}{2})$. The conventionally termed "scalar" representation is necessary for an indefinite metric quantization, as we will describe in Sec. III.

B. Field equations

In order to construct wave equations for the fields we will use the second-order Casimir operator

$$\tfrac{1}{2}J^2 \equiv \tfrac{1}{2}J^{ab}J_{ab} = \tfrac{1}{2}M^2 + MS + \tfrac{1}{2}S^2. \tag{9}$$

Each term commutes with the others and with the operator $x\cdot\partial$. Therefore they must have a common set of eigenfunctions. For the orbital part we get

$$\tfrac{1}{2}M^2 = x^2\partial^2 + (x\cdot\partial)^2 + 4(x\cdot\partial). \tag{10}$$

On conformal space it only depends on the degree n, for which we have

$$(x\cdot\partial)\Psi_\lambda = -(2|\lambda| + 1)\Psi_\lambda. \tag{11}$$

The eigenvalue of the second-order Casimir operator is, for the lowest weight representation $D(E_0,j_1,j_2)$,

$$E_0(E_0-4) + 2j_1(j_1+1) + 2j_2(j_2+1). \tag{12}$$

For acting on $\Psi_\lambda(x)$ we get the eigenvalue of $D(-\lambda,\lambda,0)$, i.e.,

$$\tfrac{1}{2}S^2 = 3\lambda(\lambda+2). \tag{13}$$

We further use

$$MS = 2\lambda(x\cdot\partial) - \sum_{i=1}^{2\lambda} \underset{(i)}{(\beta\cdot x)}(\gamma\cdot\partial) \equiv 2\lambda(x\cdot\partial) - Q. \tag{14}$$

The physical and gauge modes of the Gupta–Bleuler triplet (8) are eigenfunctions of $\frac{1}{2}J^2$ with eigenvalues $3(\lambda^2-1)$. So they satisfy the equation

$$Q\Psi_\lambda = -[\tfrac{1}{2}J^2 - 3(\lambda^2-1)]\Psi_\lambda = 0, \tag{15}$$

where Q is defined in Eq. (14). For the scalar modes we have to expect

$$Q\Psi_\lambda^{\text{scalar}} = \Psi_\lambda^{\text{gauge}},$$

since the irreducible gauge and scalar representations (on a quotient space) are equivalent. That this actually happens can be checked by applying Q to the ground state (A17). So the full Gupta–Bleuler triplet satisfies the field equation

$$Q^2\Psi_\lambda(x) = 0, \quad \text{for } \lambda \geqslant 1. \tag{16}$$

For $\lambda = \frac{1}{2}$, we have $Q^2 = 0$ automatically. In this case the tensor product (7) contains the Gupta–Bleuler triplet only; it is not necessary to project on a subspace.

For negative helicities the same equations [(15) and (16)] hold, with

$$Q = \sum_{i=1}^{-2\lambda} \underset{(i)}{(\gamma\cdot x)}(\beta\cdot\partial). \tag{17}$$

The field Ψ_λ describes massless particles with gauge freedom. In the usual electrodynamics it corresponds to the field potential A_μ. Equation (15) corresponds to the Lorentz condition. Next we want to find a field χ that describes massless particles irreducibly as the field strength $F_{\mu\nu}$ does. It must be zero for the pure gauge states. Inspecting the state (A14) suggests

$$\chi_\lambda = \prod_{i=1}^{2\lambda} \underset{(i)}{(\gamma\cdot x)}\,\Psi_\lambda, \quad \text{for } \lambda > 0, \tag{18}$$

$$\chi_\lambda = \prod_{i=1}^{-2\lambda} \underset{(i)}{(\beta\cdot x)}\,\Psi_\lambda, \quad \text{for } \lambda < 0. \tag{19}$$

Note that χ_λ has dotted indices for $\lambda > 0$ and undotted ones for $\lambda < 0$.

Clearly χ_λ has a degree of homogeneity

$$x\cdot\partial\chi_\lambda = -\chi_\lambda, \tag{20}$$

and satisfies

$$\underset{(i)}{(\beta\cdot x)}\,\chi_\lambda = 0, \quad \text{for } \lambda > 0,$$
$$\underset{(i)}{(\gamma\cdot x)}\,\chi_\lambda = 0, \quad \text{for } \lambda < 0, \quad i = 1,\ldots,2|\lambda|. \tag{21}$$

For fields with degree (-1) the wave operator ∂^2 is intrinsically defined on conformal space. Using the ground state (A13) of the physical modes it is straightforward to show that

$$\partial^2 \chi = 0 \tag{22}$$

holds. This can also be derived from the "Lorentz condition" (15).

In usual electrodynamics, the pure gauge modes have the form $A_\mu = \partial_\mu \Lambda$. Inspecting the ground state (A14) we see that in our cases the corresponding relation is

$$\Psi_{(\alpha_1 \cdots \alpha_{2\lambda})} = \sum_{i=1}^{2\lambda} (\beta \cdot x)_{\alpha\dot{\beta}} \Phi_{\dot{\beta}(\alpha_1 \cdots \overline{\alpha}_i \cdots \alpha_{2\lambda})}, \quad \text{for } \lambda > 0, \tag{23}$$

$$\Psi_{(\dot{\alpha}_1 \cdots \dot{\alpha}_{-2\lambda})} = \sum_{i=1}^{-2\lambda} (\gamma \cdot x)_{\dot{\alpha}\beta} \Phi_{\beta(\dot{\alpha}_1 \cdots \overline{\dot{\alpha}}_i \cdots \dot{\alpha}_{2\lambda})}, \quad \text{for } \lambda < 0, \tag{24}$$

where $\overline{\alpha}$ means that α is missing.

The field $\Phi(x)$ has degree

$$(x \cdot \partial)\Phi_\lambda = -(2|\lambda| + 2)\Phi_\lambda. \tag{25}$$

From the Casimir operator we get the field equation

$$\left[\underset{(1)}{\gamma \cdot x}\, \beta \cdot \partial + \sum_{i=2}^{2\lambda} \beta \cdot x\, \underset{(i)}{\gamma \cdot \partial} \right] \Phi_\lambda = 0, \quad \text{for } \lambda > 0, \tag{26}$$

$$\left[\underset{(1)}{\beta \cdot x}\, \gamma \cdot \partial + \sum_{i=2}^{-2\lambda} \gamma \cdot x\, \underset{(i)}{\beta \cdot \partial} \right] \Phi_\lambda = 0, \quad \text{for } \lambda < 0. \tag{27}$$

Thus for all helicities $\lambda \neq 0$ we have a field potential Ψ_λ with gauge freedom, a field strength χ_λ, which carries the irreducible representation, and a field Φ_λ, which describes pure gauge only. For the field potentials we have field equations (11) and (16) whose solutions carry the Gupta–Bleuler triplets (8). The first-order equations (15) project on physical and gauge modes only; they are the generalized "Lorentz conditions." This structure gives an indefinite metric quantization, which we discuss next.

III. QUANTIZATION OF THE GAUGE FIELDS

In the last section tensor products between finite- and infinite-dimensional representations of the conformal group were used to discuss the field equations. Such a structure is also very helpful to find the invariant two-point functions, of which the Pauli–Jordan commutation functions are built. It can provide the commutators of the free quantum fields, even in the case of a Hilbert space with indefinite metric, as it is encountered in gauge theories.

The invariant two-point functions of the scalar, unitary, positive-energy representations $D(n,0,0)$ are

$$D_n^+ = (x \cdot x')_+^{-n} = [x \cdot x' + i\epsilon(x_4 x_6' - x_6 x_4')]^{-2}, \tag{28}$$

and the invariant two-point functions of the tensor products (7) are

$$F_{2\lambda,0} = \frac{\delta^{(\beta_1} \cdots \delta^{\beta_{2\lambda})}_{\alpha_{2s})}}{(x \cdot x')_+^{1+2\lambda}}. \tag{29}$$

We have raised the second index of each Kronecker delta, so we can denote the independent symmetrization of the α's and β's most easily. The symmetrizations stem from the use of fully symmetric spinors. The notation $F_{2\lambda,0}$ is motivated by using more general invariant functions later; see Eq. (39) below. For the tensor product (A2) we have the same expression with dotted indices.

In the case $s = \frac{1}{2}$ these are precisely the invariant two-point functions of the Gupta–Bleuler triplets (see Ref. 10).

For $s = 1$ we have additional terms $D(4,0,1)$ [resp. $D(4,1,0)$] in the tensor products (A1) and (A2). Using Eq. (40) below we can show that

$$F_{2,2} \equiv \frac{(\beta \cdot x\ \gamma \cdot x')^{(\beta_1}_{(\alpha_1}\, (\beta \cdot x\ \gamma \cdot x')^{\beta_2)}_{\alpha_2)}}{(x \cdot x')^5} \tag{30}$$

satisfies

$$(Q-4)F_{2,2} = 0, \tag{31}$$

and therefore it is the invariant function of $D(4,0,1)$. The irreducible pure gauge representation $D(3,\frac{1}{2},\frac{1}{2})$ has the invariant function

$$G = F_{2,2} - F_{2,1}, \quad \text{with } F_{2,1} = \frac{(\beta \cdot x\ \gamma \cdot x')^{(\beta_1}_{(\alpha_1}\, \delta^{\beta_2)}_{\alpha_2)}}{(x \cdot x')^4}. \tag{32}$$

It satisfies [see Eq. (40)]

$$QG = 0. \tag{33}$$

The invariant function D_1^+ of the helicity-one Gupta–Bleuler triplet

$$D(3,\tfrac{1}{2},\tfrac{1}{2}) \to D(2,1,0) \to D(3,\tfrac{1}{2},\tfrac{1}{2}) \tag{34}$$

must be a linear combination of $F_{2,0}$ and $F_{2,2}$, which fulfills the field equation $Q^2 D_1 = 0$. It is only defined up to a factor and the addition of the invariant function G of pure gauge. We obtain

$$D_1^+ = F_{2,0} - \tfrac{3}{2} F_{2,1}. \tag{35}$$

For the helicity-(-1) Gupta–Bleuler triplet

$$D(3,\tfrac{1}{2},\tfrac{1}{2}) \to D(2,0,1) \to D(3,\tfrac{1}{2},\tfrac{1}{2}), \tag{36}$$

the invariant function is

$$D_{-1}^+ = \dot{F}_{2,0} - \tfrac{3}{2} \dot{F}_{2,1}; \tag{37}$$

$\dot{F}$ is obtained from F by replacing $(\beta \cdot x\ \gamma \cdot x')_{\alpha\beta}$ by $(\gamma \cdot x\ \beta \cdot x')_{\dot{\alpha}\dot{\beta}}$ and $\delta_{\alpha\beta}$ by $\delta_{\dot{\alpha}\dot{\beta}}$.

In the cases with $|\lambda| > 1$, we have $2|\lambda| - 1$ terms in the reduction, in addition to the Gupta–Bleuler triplets. To get the invariant function of the helicity-λ Gupta–Bleuler triplet D_λ, we have to subtract from $F_{|\lambda|,0}$ invariant functions in such a way that

$$Q^2 D_\lambda = 0 \tag{38}$$

is satisfied. It is not necessary that we give explictly all invariant functions of the irreducible representation in the reductions (A1) and (A2), but we must be sure that the subtracted terms do not contain the physical massless modes.

This latter condition is satisfied for the invariant functions

$$F_{n,m} = \frac{(\ )^{(1} \cdots (\ )^m_{m} \delta^{m+1} \cdots \delta^{n)}_{m+1}}{(x \cdot x')_+^{n+m+1}},$$

where

$$(\ ) \equiv (\beta \cdot x\ \gamma \cdot x'), \quad n = 2\lambda > 0, \quad m \geqslant 1. \tag{39}$$

To calculate $QF_{n,m}$, we first apply Q to one of the terms in

$F_{n,m}$ (without symmetrization) and then symmetrize, with the result

$$QF_{n,m} = -(n-m)(n+m+1)F_{n,m+1} - 2m(m-1)F_{n,m}. \tag{40}$$

To solve Eq. (16), for $\lambda > 0$, we make the ansatz

$$D_\lambda^+ = \sum_{m=0}^{2\lambda-1} a_m F_{2\lambda,m}, \tag{41}$$

and obtain the recursion relation

$$a_{m-1} = [m^2 d_m/\{(2\lambda+m)(2\lambda-m+1)\}]a_m, \tag{42}$$

with

$$d_{2\lambda-1} = -4, \quad d_{m-1} = -4(1+1/d_m). \tag{43}$$

For helicity $\frac{3}{2}$ we get explicitly

$$D_{3/2}^+ = F_{3,0} - 4F_{3,1} + \tfrac{2}{5}F_{3,2}, \tag{44}$$

and for helicity 2 we get

$$D_2^+ = F_{4,0} - \tfrac{15}{2}F_{4,1} + \tfrac{45}{4}F_{4,2} - \tfrac{35}{8}F_{4,3}. \tag{45}$$

The commutation relations between the quantum field operator

$$\Psi_\lambda(x) = \Psi_{(\alpha_1\cdots\alpha_{2\lambda})}(x), \quad \lambda > 0, \tag{46}$$

and its conjugate

$$\bar{\Psi}_\lambda \equiv \Psi^+_{(\dot\alpha_1\cdots\dot\alpha_{2\lambda})}(\gamma_4)_{\dot\alpha_1\beta_1}\cdots(\gamma_4)_{\dot\alpha_{2\lambda}\beta_{2\lambda}} \tag{47}$$

are

$$[\Psi_\lambda(x), \bar{\Psi}_\lambda(x')]_\pm = D_\lambda^+(x,x') + D_\lambda^-(x,x'), \quad D^-(x,x') = (D^+(x,x'))^*. \tag{48}$$

The sign in the commutator can be chosen as $-$ for integer λ and $+$ for half-integer λ, to conform with the spin statistics theorem. For $\lambda < 0$ replace D by $\dot{D}$, and γ_4 in Eq. (47) by β_4.

IV. THE FIELD EQUATIONS IN MINKOWSKI SPACE

Conformal space (1) is a compactification of flat Minkowski space. The usual flat space coordinates can be written as

$$y_\nu = x_\nu/(x_5+x_6), \quad \nu = 1,2,3,4. \tag{49}$$

Calculations are simplified if we also introduce

$$B = x^2/(x_5+x_6)^2, \quad y_+ = x_5 + x_6, \tag{50}$$

which satisfy $B = 0$, $y_+ \equiv \lambda y_+$ in conformal space.

Spinor fields in flat space have translational generators $P_\nu = -i\partial/\partial y^\nu$, and they do not depend on y_+. These requirements are met by

$$\chi_{M,\lambda} = y_+ U^{2\lambda}\chi_\lambda, \quad \Psi_{M,\lambda} = y_+^{2\lambda+1}U^{2\lambda}\chi_\lambda, \quad \lambda > 0, \tag{51}$$

where

$$U^{2\lambda} = \prod_{i=1}^{2\lambda} U_{\alpha_i\beta_i}, \quad U_{\alpha_i\beta_i} = (1 + y^\nu\gamma_\nu\Gamma^-)_{\alpha_i\beta_i}, \quad \Gamma^\pm = (\beta_6 \pm \beta_5)/2. \tag{52}$$

We also will need the transformation matrix of the dotted spinors, which has U replaced by

$$\dot{U}_{\dot\alpha\dot\beta_i} = (1 - y^\nu\gamma_\nu\Gamma^+)_{\dot\alpha\dot\beta_i}. \tag{53}$$

In the sequel we again give details only for $\lambda > 0$ (the equations for $\lambda < 0$ are obtained by $\gamma \leftrightarrow \beta$, $\Gamma^\pm \to \Gamma^\mp$, $\lambda \to -\lambda$).

First we want to write the field equations of the χ field in flat space; they describe massless particles in irreducible representations of the conformal group.

The subsidiary conditions $(\underset{(i)}{\gamma\cdot x})\chi_\lambda = 0$ [Eq. (21)] give in flat space [see Eq. (B7)]

$$\underset{(i)}{\Gamma}\chi_{M,\lambda} = 0. \tag{54}$$

So the only nonvanishing components are

$$\prod_{i=1}^{2\lambda}\Gamma^-_{(i)}\chi_{M,\lambda} \equiv :\chi^{-\cdots-}_{M,\lambda}. \tag{55}$$

The field equation (22) becomes

$$0 = U^{2\lambda}\partial^2(U^{2\lambda})^{-1}\chi_{M,\lambda} \tag{56}$$

$$= y_+^{-1}\Big\{\Box + 2\sum_{k=1}^{2\lambda}\partial_{(k)}\Gamma^+_{(k)} + \sum_{j\neq k}\gamma^\nu_{(j)}\Gamma^-_{(j)}\gamma_{\nu(k)}\Gamma^-_{(k)}\Big\}\chi_{M,\lambda}. \tag{57}$$

Using Eq. (54) above we get, by applying appropriate projection operators,

$$\partial_{(k)}\chi^{-\cdots-}_{M,\lambda} = 0, \quad \text{where } \partial_{(k)} \equiv \gamma^\nu_{(k)}\frac{\partial}{\partial y^\nu}, \quad \lambda > 0, \tag{58}$$

$$\underset{(k)}{\gamma^\nu}\underset{(j)}{\gamma_\nu}\chi^{-\cdots-}_{M,\lambda} = 0. \tag{59}$$

The latter equations are automatically satisfied for symmetric spinors [see (B9)]. If we use a basis for the γ matrices (6) in which γ_6 and γ_5 are diagonal, e.g.,

$$\gamma_4^M = i\gamma_5, \quad \gamma_5^M = i\gamma_4, \tag{60}$$

then the projection operators become

$$\Gamma^+ = \begin{pmatrix}1&0\\0&0\end{pmatrix}, \quad \Gamma^- = \begin{pmatrix}0&0\\0&1\end{pmatrix}. \tag{61}$$

So χ^- becomes a two-spinor and Eq. (58) is the Weyl equation acting on each two-spinor index. We have the usual massless Bargmann–Wigner equations for symmetric spinors.[6] The corresponding field equations for negative helicities are

$$\partial_{(j)}\chi^{+\cdots+}_{M,\lambda} = 0, \quad \lambda < 0, \quad j = 1,\ldots,2|\lambda|.$$

The two multispinors for $\lambda = s$, $\lambda = -s$ can be taken together to a Dirac multispinor, which satisfies the massless Dirac equation and describes massless helicity $\lambda = \pm s$ particles. Next we rewrite the field equations of the Ψ field, which has gauge freedom, in flat space.

The pure gauge states satisfy $(\gamma\cdot x)\chi_\lambda = 0$, which becomes, by using (B7),

$$\Psi^{+\cdots+}_{M,\lambda} = 0, \quad \lambda > 0. \tag{62}$$

The "Lorentz condition" (15) becomes [see (B8)]

$$\sum_{j=1}^{2s}\Gamma^-_{(j)}\Big\{\partial_{(j)} + 2\sum_{k\neq j}\Gamma^-_{(k)}\Big\}\Psi_{M,\lambda} = 0. \tag{63}$$

From this we obtain, by acting with appropriate projection operators,

$$\partial_{(j)}\Psi^{+\cdots+}_{M,\lambda} = 0, \quad \lambda > 0, \tag{64}$$

i.e., the Weyl equation acting on each two-spinor index of $\Psi^{+\cdots+}_{M,\lambda}$.

In the case of helicity $\frac{1}{2}$ there is no further equation, and any function Ψ^{-} describes gauge. For helicity 1 there is, in addition to Eqs. (64),

$$\partial_{(1)}\Psi^{+-}_{M,1} + \partial_{(2)}\Psi^{-+}_{M,1} + 4\Psi^{--}_{M,1} = 0, \tag{65}$$

and similar equations for higher helicities. They project on those functions $\Psi^{+-\cdots}$, which carry pure gauge, but not the additional representations in the tensor products (A1). The full Gupta–Bleuler triplet has for spin $\frac{1}{2}$ no field equation, and for higher spins we have $Q^2\Psi = 0$; see Eq. (16). This can be transformed in flat space as before. For helicity 1 we obtain, e.g.,

$$\partial_{(1)}\partial_{(2)}\Psi^{++}_{M,1} + 2\partial_{(1)}\Psi^{+-}_{M,1} + 2\partial_{(2)}\Psi^{-+}_{M,1} + 8\Psi^{--}_{M,1} = 0. \tag{66}$$

This equation has the "scalar" modes as a solution, in addition to the physical and gauge modes that already solve Eqs. (64) and (65).

Again the equations for negative helicity are obtained by exchanging $\Gamma^{+} \leftrightarrow -\Gamma^{-}$. So, e.g., the equivalent of Eq. (64) is

$$\partial_{(j)}\Psi^{-\cdots-}_{M,\lambda} = 0, \quad \lambda > 0. \tag{67}$$

It is also possible to rewrite the invariant two-point functions of Sec. III in flat space, as was done in Ref. 10 for spin $\frac{1}{2}$.

ACKNOWLEDGMENTS

We have the pleasure of thanking Professor H. D. Doebner for encouragement and advice.

We further thank the DAAD and Spanish Ministerio de Educacion Ciencia for financial support that made the present collaboration possible.

APPENDIX A: REDUCTION OF THE TENSOR PRODUCTS

The aim of this appendix is to give the reduction of the tensor product

$$\begin{aligned} &D(-\lambda,\lambda,0)\otimes D(1+2\lambda,0,0) \\ &\quad = \{D(2+\lambda,\lambda-\tfrac{1}{2},\tfrac{1}{2}) \to D(1+\lambda,\lambda,0) \\ &\qquad \to D(2+\lambda,\lambda-\tfrac{1}{2},\tfrac{1}{2})\} \\ &\qquad \oplus \sum_{i=2}^{2\lambda} D(1+\lambda+i,\lambda-i/2,i/2), \quad \lambda>0. \end{aligned} \tag{A1}$$

Then also a similar formula holds, with the two angular momenta j_1, j_2 exchanged:

$$\begin{aligned} &D(-s,0,s)\otimes D(1+2s,0,0) \\ &\quad = \{D(2+s,\tfrac{1}{2},s-\tfrac{1}{2}) \to D(1+s,0,s) \\ &\qquad \to D(2+s,\tfrac{1}{2},s-\tfrac{1}{2})\} \\ &\qquad \oplus \sum_{i=2}^{2s} D(1+s+i,i/2,s-i/2), \quad s>0. \end{aligned} \tag{A2}$$

Comparing weight diagrams as described in Ref. 10, Appendix B, gives the reduction except for the leaks between Weyl-equivalent representations, which we denote by an arrow ($\to$). Note that the massless representation $D(1+\lambda,\lambda,0)$ has a reduced weight diagram, which "lost" a $D(2+\lambda,\lambda-\frac{1}{2},\frac{1}{2})$ as compared to the general case.

To prove the leaks, we act with step operators on lowest weights of the representations. We call

$$J_{46} = -i(x_+\partial_+ - x_-\partial_-) + \frac{1}{2}\sum_{j=1}^{2\lambda}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}_{(j)},$$

$$\partial_\pm = \frac{\partial}{\partial x_\pm}, \quad x_\pm = x_4 \pm i x_6, \tag{A3}$$

the conformal energy. The generators of the two SU(2) subgroups are

$$J_i^{(1,2)} = \tfrac{1}{2}(\epsilon_{ijk}J_{jk} \pm J_{i5}). \tag{A4}$$

Their components are explicitly

$$\begin{aligned} J_i^{(1,2)} = &-\frac{i}{2}(x_j\partial_k - x_k\partial_j \pm x_i\partial_5 \mp x_5\partial_i) \\ &+\frac{1}{2}\sum_{m=1}^{2\lambda}\begin{cases} \begin{pmatrix} \sigma_i & 0 \\ 0 & 0 \end{pmatrix}_{(m)}, \\ \begin{pmatrix} 0 & 0 \\ 0 & \sigma_i \end{pmatrix}_{(m)}, \end{cases} \quad i,j,k \text{ cyclic.} \end{aligned} \tag{A5}$$

The other noncompact generators of SU(2,2) we write as energy raising and lowering operators

$$J_r^{\pm} = J_{r5} \pm iJ_{r4}, \quad r = 1,2,3,5, \tag{A6}$$

or explicitly

$$J_i^{+} = -2x_i\partial_+ - x_-\partial_i + \sum_{k=1}^{2\lambda}\begin{pmatrix} 0 & 0 \\ \sigma_i & 0 \end{pmatrix}_{(k)}, \quad i = 1,2,3, \tag{A7}$$

$$J_i^{-} = 2x_i\partial_- + x_+\partial_i - \sum_{k=1}^{2\lambda}\begin{pmatrix} 0 & \sigma_i \\ 0 & 0 \end{pmatrix}_{(k)}, \quad i = 1,2,3, \tag{A8}$$

$$J_5^{+} = -2x_5\partial_+ - x_-\partial_5 + \sum_{k=1}^{2\lambda}\begin{pmatrix} 0 & 0 \\ i & 0 \end{pmatrix}_{(k)}, \tag{A9}$$

$$J_5^{-} = 2x_5\partial_- + x_+\partial_5 + \sum_{k=1}^{2\lambda}\begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix}_{(k)}. \tag{A10}$$

The lowest weight of the infinite representation $D(E_0,0,0)$ is $x_+^{-E_0}$ and the states at energy $E = E_0 + 1$ are $x_r x_+^{-(E_0+1)}$, $r = 1,2,3,5$. For finite representation $D(-\lambda,\lambda,0)$, the lowest weight with eigenvalues $(E,(j_1,j_1^3),(j_2,j_2^3)) = (-\lambda, (\lambda,-\lambda),(0,0))$ is

$$\underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \times \cdots \times \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{2\lambda \text{ terms}}, \tag{A11}$$

and the state with eigenvalues $(1-\lambda,(\lambda-\frac{1}{2},-\lambda-\frac{1}{2}),(\frac{1}{2},-\frac{1}{2}))$ is

$$(s)\equiv\begin{pmatrix}0\\0\\1\end{pmatrix}\times\overbrace{\begin{pmatrix}0\\1\\0\end{pmatrix}\times\cdots\times\begin{pmatrix}0\\1\\0\end{pmatrix}}^{2\lambda-1\text{ terms}}+\cdots+\begin{pmatrix}0\\1\\0\end{pmatrix}\times\cdots\times\begin{pmatrix}0\\1\\0\end{pmatrix}\times\begin{pmatrix}0\\0\\1\end{pmatrix}\quad\text{(}2\lambda\text{ terms)}. \tag{A12}$$

So the lowest state in the tensor product (A1) is

$$|p_0>_\lambda = x_+^{-2\lambda-1}\underbrace{\begin{pmatrix}0\\1\\0\end{pmatrix}\times\cdots\times\begin{pmatrix}0\\1\\0\end{pmatrix}}_{2\lambda-1\text{ terms}}. \tag{A13}$$

If this lowest weight belongs to an irreducible representation, acting with the raising operators J_r^+ on it would only produce states at $(E,j_1,j_2)=(2+\lambda,\lambda+\frac{1}{2},\frac{1}{2})$. But in our case here it also gives states at $(2+\lambda,\lambda-\frac{1}{2},\frac{1}{2})$, e.g., the state with $j_1^3=-\lambda+\frac{1}{2}$, $j_2^3=-\frac{1}{2}$:

$$|g_0>_\lambda = x_+^{-2\lambda-2}\times\left[(\beta\cdot x)\begin{pmatrix}0\\0\\1\end{pmatrix}\times\overbrace{\begin{pmatrix}0\\1\\0\end{pmatrix}\times\cdots\times\begin{pmatrix}0\\1\\0\end{pmatrix}}^{2\lambda-1\text{ terms}}+\cdots+\begin{pmatrix}0\\1\\0\end{pmatrix}\times\cdots\times\begin{pmatrix}0\\1\\0\end{pmatrix}\times(\beta\cdot x)\begin{pmatrix}0\\0\\1\end{pmatrix}\right]\quad\text{(}2\lambda\text{ terms)}. \tag{A14}$$

It can, e.g., be obtained by

$$|g_0>_\lambda = [1/(2\lambda+1)]\{\tfrac{1}{2}(iJ_1^+ + J_2^+)J_+^{(1)} - \lambda(iJ_3^+ - J_5^+)\}|p_0\rangle_\lambda, \tag{A15}$$

where $J_+^{(1)}=J_1^{(1)}+iJ_2^{(1)}$ is the raising operator in the first SU(2).[14] The state $|g_0\rangle_\lambda$ itself is a lowest weight

$$J_r^-|g_0\rangle_\lambda = 0 \tag{A16}$$

of the irreducible representation $D(2+\lambda,\lambda-\frac{1}{2},\frac{1}{2})$.

There is another independent state in the tensor product (A1) with the same quantum number as $|g_0\rangle_\lambda$, namely

$$|s_0\rangle_\lambda = x_+^{-2\lambda-1}(s), \tag{A17}$$

where (s) is the state from Eq. (A12). Acting with the energy lowering operators J_r^- on it gives the state $|p_0>_\lambda$. It is only a relative lowest weight.

So we could prove the reduction (A1). The explicit form of the lowest weights given above is also used in Sec. II.

APPENDIX B: SOME USEFUL RELATIONS

Using the property (5) of the β, γ matrices we get

$$(\beta\cdot x)(\gamma\cdot x) = (\gamma\cdot x)(\beta\cdot x) = x^2, \tag{B1}$$

$$(\beta\cdot x\ \gamma\cdot\partial)^2 = -x^2\partial^2 + (4+2x\cdot\partial)\beta\cdot x\ \gamma\cdot\partial. \tag{B2}$$

The trace of matrices acting on different spinor indices is

$$\beta^a_{\beta_1\dot\alpha_1}(\gamma_a)_{\dot\beta_2\alpha_2}\psi_{\dot\alpha_1\alpha_2} = 2\psi_{\dot\beta_2\beta_1}, \tag{B3}$$

$$\gamma^A_{\beta_1\alpha_1}(\gamma_a)_{\dot\beta_2\alpha_2}\psi_{\alpha_1\alpha_2} = 2(\psi_{\dot\beta_1\dot\beta_2} - \psi_{\dot\beta_2\beta_1}). \tag{B4}$$

To show this, transform into a matrix product $\beta^a\psi\gamma_a^T$ and use the explicit basis of Eq. (6). With (B4) we get

$$(\beta\cdot x\ \gamma^A)^{\beta_1}_{(\alpha_1}(\beta\cdot x'\ \gamma_a)^{\beta_2}_{\alpha_2)} = 0. \tag{B5}$$

The transformations U in Sec. IV satisfy

$$U^{-1} = 1 - \not\beta\Gamma^-, \tag{B6}$$

$$(1-\not\beta\Gamma^+)(\gamma\cdot x)(1-\not\beta\Gamma^-) = y_+\Gamma^+, \tag{B7}$$

$$U\beta\cdot x\ \gamma\cdot\partial U^{-1} = -\Gamma^-\left[\gamma^\nu\frac{\partial}{\partial y_\nu} - 2(2+y_+\partial_+)\right]. \tag{B8}$$

The trace of γ matrices acting on different spinor indices of χ (or ψ) is, for the various components $\underset{(1)}{\Gamma^\pm}\underset{(2)}{\Gamma^\pm}\psi = \psi^{\pm\pm}$,

$$\underset{(1)(2)}{\gamma^\nu\gamma_\nu}(\chi^{++},\chi^{--},\chi^{+-},\chi^{-+}) = (0,0,2\chi^{+-},2\chi^{-+}). \tag{B9}$$

Again we transform in a matrix product, e.g., $\gamma^\nu\Gamma^+\chi(\Gamma^+)^T\chi_\nu^T$ and use the explicit matrices (6) and γ_4^M,γ_5^M from Eq. (60).

[1] M. Fierz and W. Pauli, Proc. R. Soc. London Ser. A **173**, 211 (1939).
[2] C. R. Hagen and L. P. S. Singh, Phys. Rev. D **9**, 898 (1974); C. Fronsdal, *ibid.* **18**, 3624 (1978).
[3] B. Binegar, C. Fronsdal, and W. Heidenreich, J. Math. Phys. **24**, 2828 (1983).
[4] B. Binegar, C. Fronsdal, and W. Heidenreich, Phys. Rev. D **27**, 2249 (1983).
[5] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941); L. P. S. Singh and C. R. Hagen, Phys. Rev. D **9**, 910 (1974); F. Fang and C. Fronsdal, *ibid.* **18**, 3630 (1978).
[6] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. (USA) **34**, 211 (1948).
[7] M. A. Rodriguez and M. Lorente, Nuovo Cimento A **83**, 249 (1984); W.

Rindler and R. Penrose, *Spinors and Spacetime* (Cambridge U.P., Cambridge, 1984).
[8]P. A. M. Dirac, Ann. Math. **37**, 429 (1936).
[9]L. Castell, Nuovo Cimento A **42**, 160 (1977).
[10]W. Heidenreich, Nuovo Cimento A **80**, 220 (1984).
[11]M. Lorente, in *Conformal Groups and Related Symmetries*, Clausthal, 1985, *Springer Lecture Notes in Physics*, Vol. 261, edited by A. O. Barut and H. D. Doebner (Springer, Berlin, 1986), p. 101.
[12]W. F. Heidenreich, Math. Phys. **27**, 2154 (1986); G. P. Collins, N. A. Doughty, J. Math. Phys. **28**, 448 (1987).
[13]L. Castell, Nucl. Phys. B **13**, 231 (1969).
[14]M. Lorente and B. Gruber, J. Math. Phys. **25**, 1674 (1984); B. Gruber, R. Lenczewski, and M. Lorente, "Induced scalar products and unitarization of representations," University of Carbondale preprint. In the formulation of this paper the representation with highest weight vector $|p_0\rangle_\lambda$ is indecomposable, because it contains the external vector $|g_0\rangle_\lambda$.

On the Heisenberg and orthosymplectic superalgebras of the harmonic oscillator

J. Beckers and D. Dehin[a)]
Physique Théorique et Mathématique, Institut de Physique au Sart Tilman, B.5, Université de Liège, B-4000 Liège 1, Belgium

V. Hussin[b)]
Centre de Recherches Mathématiques, Université de Montréal, CP 6128, Succ. A, Montréal, Québec H3C 3J7, Canada

(Received 29 December 1987; accepted for publication 16 March 1988)

The new invariance superalgebra $\mathrm{osp}(2,2n)\oplus\mathrm{sh}(2,2n)$ for an n-dimensional quantum harmonic oscillator is proposed, and the corresponding supersymmetric Hamiltonian is constructed. This superalgebra is related to the structure $\mathrm{osp}(3,2n)$ recently studied by Englefield [J. Phys. A **21**, 1309 (1988)]. By interchanging the even and odd characters of the fundamental generators belonging to the Heisenberg superalgebra $\mathrm{sh}(2,2n)$, the connection between both structures is established and its consequences exploited. The specific cases $n = 1,2,3$ are discussed and compared with recent contributions. Some general conclusions are also summarized.

I. INTRODUCTION

The invariance superalgebra of the n-dimensional harmonic oscillator was first recognized by de Crombrugghe and Rittenberg[1] as the orthosymplectic structure $\mathrm{osp}(2n,2n)$ after the advent of supersymmetric quantum mechanics.[2] Since this proposal two other superstructures have been pointed out. *On the one hand*, Beckers, Dehin, and Hussin[3] have incorporated the harmonic oscillator in the superconformal context and have shown that the *largest* superalgebra of the harmonic oscillator is the semidirect sum of the de Crombrugghe–Rittenberg structure and the Heisenberg superalgebra $\mathrm{sh}(2n,2n)$, i.e., the superalgebra $\mathrm{osp}(2n,2n)\oplus\mathrm{sh}(2n,2n)$. They have extended the conformal symmetries displayed by Niederer[4] to the supercontext by combining Witten's[2] and Fubini–Rabinovici's[5] studies leading to type-Q and type-S supercharges. This largest superalgebra is the maximal dynamical invariance algebra of the n-dimensional harmonic oscillator containing dynamical[6] as well as kinematical[7] symmetries and supersymmetries.[3] *On the other hand*, Englefield[8] has just proposed the superalgebra $\mathrm{osp}(3,2n)$ by starting from the metaplectic representation of $\mathrm{osp}(3,2)$ with a one-dimensional harmonic oscillator with spin,[9] and by extending it to the supersymmetric context in n dimensions.

Here our main purpose is to construct a new superalgebra $\mathrm{osp}(2,2n)\oplus\mathrm{sh}(2,2n)$ for an n-dimensional harmonic oscillator, appearing as a substructure of the superalgebra $\mathrm{osp}(2n,2n)\oplus\mathrm{sh}(2n,2n)$, which will be clearly related to the Englefield superstructure $\mathrm{osp}(3,2n)$. Then we want to take advantage of this relation between $\mathrm{osp}(2,2n)\oplus\mathrm{sh}(2,2n)$ and $\mathrm{osp}(3,2n)$ in order to understand both structures in connection with the physics of the harmonic oscillator and with different procedures of supersymmetrization already used in the literature. In particular, we think of the standard procedure[2] and the spin-orbit coupling procedure[10,11] first discussed by Balantekin in the three-dimensional context and generalized by Kostelecky, Nieto, and Truax[12] to the arbitrary n case. We will see the fundamental role played by the Heisenberg superalgebras exhibited in the above structures.

In Sec. II we construct the superalgebra $\mathrm{osp}(2,2n)\oplus\mathrm{sh}(2,2n)$ by looking for two degrees of freedom in the fermionic variables and $2n$ in the bosonic variables describing an n-dimensional supersymmetric harmonic oscillator. From previous results[3] we easily get all the generators in terms of the operators of the Heisenberg superalgebra, and we explicitly construct the corresponding supersymmetric Hamiltonian. Section III is devoted to the connection between our superalgebra and $\mathrm{osp}(3,2n)$, the one proposed and discussed by Englefield.[8] We also give the explicit form of the corresponding Hamiltonian in this $\mathrm{osp}(3,2n)$ context, as well as a realization of all its generators. In Sec. IV we then come back to the specific physical cases $n = 1,2,3$ and make some comments in connection with different procedures of supersymmetrization and with other recent contributions.[13–17] Finally, Sec. V contains some general conclusions, putting the accent on the Heisenberg generators and their fundamental role in the understanding of these superalgebras.

In order to fix our conventions and units let us mention that the mass m is taken as unity, although we maintain ω, the angular frequency of the harmonic oscillator, in our formulas. In general, the summations on Latin indices go from 1 to n except for specific mentions; they are always understood on repeated indices. In connection with the Heisenberg superalgebras and their dimensions, we frequently omit the $(+1)$ dimension associated with the central extension they always exhibit due to the presence of the identity operator in the corresponding structures.

[a)] Chercheur Institut Interuniversitaire des Sciences Nucléaires.
[b)] On leave from Liège University, Liège, Belgium.

 0022-2488/88/071705-07$02.50

II. THE SUPERALGEBRA osp(2,2n) ⊕ sh(2,2n) AND THE HARMONIC OSCILLATOR

After the first $N=2$ supersymmetric quantum-mechanical system was given and recognized by Witten[2] (with a suitable superpotential) as a supersymmetric version of the 1-dimensional harmonic oscillator, it was a trivial task to extend these considerations to arbitrary (n) spatial dimensions, maintaining the famous cancellation between the Bose and Fermi contributions to the ground-state energy in such a supersymmetric theory. Then the extension to osp($2n$,$2n$) with respect to *dynamical*[6] symmetries and supersymmetries has also been proposed.[1] Moreover, the corresponding discussion within the so-called nonrelativistic *conformal* context[7] has been pointed out by the present authors,[3,11] leading to the largest dynamical invariance superalgebra osp($2n$,$2n$) ⊕ sh($2n$,$2n$) when the standard procedure of supersymmetrization[2] is enhanced. Such considerations include dynamical as well as *kinematical*[4] (super)symmetries and refer to the *same* number ($2n$) of bosonic *and* fermionic degrees of freedom.

Let us now proceed to construct osp(2,2n) ⊕ sh(2,2n) as the invariance superalgebra of a harmonic oscillator by starting with osp($2n$,$2n$) ⊕ sh($2n$,$2n$), the superalgebra[3] of the harmonic oscillator in the well-known standard procedure. Let us recall that the above $(8n^2+4n(+1))$-dimensional superalgebra is formed by the $8n^2$ generators T_{kl}, $C_{\pm,kl}$, Y_{kl}, $Z_{\pm,kl}$, $Q_{\pm,kl}$, and $S_{\pm,kl}$ of osp($2n$,$2n$), and the $(4n(+1))$ generators $P_{\pm,k}$ and $T_{\pm,k}$ (and I) of sh($2n$,$2n$). We have noticed[3] that the first $8n^2$ generators can be expressed only in terms of those generating the Heisenberg superalgebra sh($2n$,$2n$). Indeed, we have

$$T_{kl}=\tfrac{1}{4}\{P_{-,k},P_{+,l}\}, \tag{2.1a}$$

$$C_{\pm,kl}=\mp(i/4)\{P_{\pm,k},P_{\pm,l}\}, \tag{2.1b}$$

$$Y_{kl}=(\omega/2)[T_{+,k},T_{-,l}], \tag{2.1c}$$

$$Z_{\pm,kl}=\pm(i\omega/2)[T_{\pm,k},T_{\pm,l}], \tag{2.1d}$$

$$Q_{\pm,kl}=(1/\sqrt{2})P_{\mp,k}T_{\pm,l}=(1/2\sqrt{2})\{P_{\mp,k},T_{\pm,l}\}, \tag{2.1e}$$

$$S_{\pm,kl}=(1/\sqrt{2})P_{\pm,k}T_{\pm,l}=(1/2\sqrt{2})\{P_{\pm,k},T_{\pm,l}\}, \tag{2.1f}$$

where the Heisenberg generators are given by

$$P_{\pm,k}=\pm i\sqrt{2\omega}e^{\mp i\omega t}a_{\pm,k} \tag{2.2a}$$

and

$$T_{\pm,k}=e^{\mp i\omega t}\xi_{\pm,k}. \tag{2.2b}$$

Here the $a_{\pm,k}$'s are the components of the creation ($+$) and annihilation ($-$) operators in the n-dimensional space ($k=1,\ldots,n$) satisfying

$$[a_{-,k},a_{+,l}]=\delta_{kl}, \tag{2.3}$$

while the $\xi_{\pm,k}$'s are Grassmannian variables such that

$$\{\xi_{+,k},\xi_{-,l}\}=\delta_{kl},\quad \{\xi_{\pm,k},\xi_{\pm,l}\}=0. \tag{2.4}$$

These operators clearly refer to $2n$ bosonic and $2n$ fermionic degrees of freedom, which enter, respectively, into the bosonic and fermionic parts of the supersymmetric Hamiltonian for the harmonic oscillator given by

$$H^{\rm SS}_{\rm H.O.}=(\omega/2)\{\mathbf{a}_+,\mathbf{a}_-\}+(\omega/2)[\boldsymbol{\xi}_+,\boldsymbol{\xi}_-]=\{Q_+,Q_-\}, \tag{2.5}$$

where the supercharges

$$Q_\pm=Q_{\pm,kk}=\mp i\sqrt{\omega}\mathbf{a}_\mp\cdot\boldsymbol{\xi}_\pm=(1/\sqrt{2})\mathbf{P}_\mp\cdot\mathbf{T}_\pm \tag{2.6}$$

are conserved (as expected):

$$[Q_\pm,H^{\rm SS}_{\rm H.O.}]=0. \tag{2.7}$$

Let us now restrict ourselves to two fermionic degrees of freedom and seek to obtain the superalgebra osp(2,2n) ⊕sh(2,2n) in a similar manner as before, but with only two Grassmannian variables, ξ_+ and ξ_-. So we will be working with the simplest two-dimensional Clifford algebra,[18]

$$\{\varphi^a,\varphi^b\}=\delta^{ab},\quad a,b=1,2, \tag{2.8a}$$

with

$$\varphi^1=\xi_++\xi_-,\quad \varphi^2=i(\xi_--\xi_+). \tag{2.8b}$$

The generators (2.2b) then become the only two generators T_+ and T_-, so that the Heisenberg superalgebra becomes a $[2+2n(+1)]$-dimensional structure that we call sh(2,2n). Correspondingly we immediately see through Eqs. (2.1) that the algebra sp($2n$) generated by the $(2n^2+n)$ operators (2.1a) and (2.1b) is unchanged, while the $(2n^2-n)$ generators (2.1c) and (2.1d) of so($2n$) reduce to the single generator

$$Y=(\omega/2)[T_+,T_-], \tag{2.9}$$

associated with a so(2) algebra. Moreover, the $4n^2$ charges (2.1e) and (2.1f) reduce to the $4n$ operators

$$Q_{\pm,k}=(1/\sqrt{2})P_{\mp,k}T_\pm,\quad S_{\pm,k}=(1/\sqrt{2})P_{\pm,k}T_\pm. \tag{2.10}$$

In this way we get $(2n^2+7n+3(+1))$ generators, which form the superalgebra osp(2,2n)⊕sh(2,2n). Let us just take note here of the structure relations of the Heisenberg superalgebra sh(2,2n):

$$[P_{-,k},P_{+,l}]=2\omega\delta_{kl}I,\quad [P_{\pm,k},P_{\pm,l}]=0, \tag{2.11}$$

$$[P_{\pm,k},T_+]=0,\quad [P_{\pm,k},T_-]=0, \tag{2.12}$$

$$\{T_+,T_-\}=I,\quad \{T_\pm,T_\pm\}=0. \tag{2.13}$$

The other structure relations for osp(2,2n)⊕sh(2,2n) can evidently be obtained from the preceding ones due to the definitions of the osp(2,2n) generators in terms of those of the Heisenberg superalgebra.

Let us summarize by saying that we are dealing with $(2n^2+3n+1)$ *even* generators T_{kl}, $C_{\pm,kl}$, Y, and $P_{\pm,k}$, as well as with $(2+4n)$ *odd* generators $T_\pm$, $Q_{\pm,k}$, and $S_{\pm,k}$. These even ($\mathscr{E}$) or odd ($\mathscr{O}$) properties of the operators are closely related to the fact that they effectively satisfy the structure relations

$$[\mathscr{E},\mathscr{E}]=\mathscr{E},\quad [\mathscr{E},\mathscr{O}]=\mathscr{O},\quad \{\mathscr{O},\mathscr{O}\}=\mathscr{E}. \tag{2.14}$$

So we can look for the explicit form of the supersymmetric Hamiltonian for a harmonic oscillator with the invariance superalgebra osp (2,2n)⊕sh(2,2n). From the type-Q supercharges (2.10), we define in this $N=2$ supersymmetric theory[2] the two other supercharges,

$$Q_\pm=\alpha_{\pm,k}Q_{\pm,k}=(1/\sqrt{2})\alpha_{\pm,k}P_{\mp,k}T_\pm, \tag{2.15}$$

where the $\alpha_{\pm,k}$'s are actually arbitrary coefficients. Then by using the relations [see Eqs. (2.1a) and (2.11)]

$$P_{-,k}P_{+,l} = 2T_{kl} + \omega\delta_{kl}I, \quad P_{+,l}P_{-,k} = 2T_{kl} - \omega\delta_{kl}I,$$

and [see Eqs. (2.9) and (2.13)]

$$T_{\pm}\,T_{\mp} = \tfrac{1}{2}(I \pm (2/\omega)Y),$$

we easily get

$$\begin{aligned}\{Q_+,Q_-\} &= \tfrac{1}{2}\alpha_{+,k}\alpha_{-,l}\{P_{-,k}T_+,P_{+,l}T_-\} \\ &= \alpha_{+,k}\alpha_{-,l}T_{kl} + \alpha_{+,k}\alpha_{-,k}Y = H^{\mathrm{SS}}.\end{aligned} \tag{2.16}$$

Due to the constraints[2] on the supercharges (2.15) in an $N=2$ supersymmetric theory, we then require that the coefficients λ satisfy

$$\alpha_{-,k} = \alpha^*_{+,k}, \quad \forall k = 1,\ldots,n, \tag{2.17a}$$

so that the purely fermionic part of Eq. (2.16) becomes (as expected)

$$H_F = nY, \tag{2.18a}$$

if we require, moreover,

$$|\alpha_{+,k}|^2 = 1, \quad \forall k = 1,\ldots,n. \tag{2.17b}$$

The last condition immediately results if the purely bosonic Hamiltonian[6] has to be included (as expected) in the expression of the supersymmetric Hamiltonian:

$$H_B = T_{kk}. \tag{2.18b}$$

In fact, by introducing for $k \neq l$ the skew-symmetric L_{kl} and symmetric M_{kl} operators defined by

$$L_{kl} = (i/\omega)(T_{kl} - T_{lk}), \quad M_{kl} = (1/\omega)(T_{kl} + T_{lk}), \tag{2.19}$$

we obtain

$$\alpha_{+,k}\alpha_{-,l}T_{kl} = T_{kk} + \frac{\omega}{2}\sum_{k\neq l}\beta_{kl}L_{kl} + \frac{\omega}{2}\sum_{k\neq l}\gamma_{kl}M_{kl}, \tag{2.20}$$

where the skew-symmetric β_{kl} and symmetric γ_{kl} coefficients are defined, respectively, by

$$\begin{aligned}\beta_{kl} &= -(i/2)(\alpha_{+,k}\alpha_{-,l} - \alpha_{+,l}\alpha_{-,k}), \\ \gamma_{kl} &= \tfrac{1}{2}(\alpha_{+,k}\alpha_{-,l} + \alpha_{+,l}\alpha_{-,k}).\end{aligned} \tag{2.21}$$

Finally, the supersymmetric Hamiltonian (2.16) takes the form

$$H^{\mathrm{SS}} = H_B + nY + \frac{\omega}{2}\sum_{k\neq l}\beta_{kl}L_{kl} + \frac{\omega}{2}\sum_{k\neq l}\gamma_{kl}M_{kl}. \tag{2.22}$$

Let us note that, with the properties (2.17), we can write the coefficients (2.21) in the form

$$\beta_{kl} = \sin(a_k - a_l), \quad \gamma_{kl} = \cos(a_k - a_l), \tag{2.23}$$

where the numbers a_k are defined by

$$\alpha_{+,k} = \exp(ia_k). \tag{2.24}$$

For $n > 1$ we immediately conclude that all the β_{kl} and γ_{kl} cannot be taken equal to zero simultaneously. Finally, realizing the Grassmannian variables in terms of the usual Pauli matrices of the form

$$\xi_+ \equiv \sigma_+ = \tfrac{1}{2}(\sigma_1 + i\sigma_2), \quad \xi_- \equiv \sigma_- = \tfrac{1}{2}(\sigma_1 - i\sigma_2), \tag{2.25a}$$

or, conversely,

$$\xi_+ \equiv \sigma_+, \quad \xi_- \equiv \sigma_+, \tag{2.25b}$$

we get the two specific Hamiltonians

$$H^{\mathrm{SS}}_{(\varepsilon)} = H_B + \frac{n\varepsilon}{2}\omega\sigma_3 + \frac{\omega}{2}\sum_{k\neq l}\beta_{kl}L_{kl} + \frac{\omega}{2}\sum_{k\neq l}\gamma_{kl}M_{kl}. \tag{2.26}$$

The cases $\varepsilon = +1$ and $\varepsilon = -1$ correspond to the choices (2.25a) and (2.25b), respectively.

III. A CONNECTION WITH THE SUPERALGEBRA osp(3,2n)

Starting from the metaplectic representation of the superalgebra osp(3,2) discussed by Van der Jeugt[9] for a one-dimensional harmonic oscillator with spin, Englefield[8] has very recently proposed its generalization to the orthosymplectic superalgebra osp(3,2n) for an n-dimensional supersymmetric harmonic oscillator. In fact, Englefield has shown that the bound states of such an n-dimensional supersymmetric oscillator span an irreducible representation of osp(3,2n) containing the supersymmetric Hamiltonian.

In connection with this result we want to point out that the dimension of osp(3,2n) is

$$d = \tfrac{1}{2}[(2n+3)^2 + 2n - 3] = 2n^2 + 7n + 3, \tag{3.1}$$

so that it coincides with the dimension of the superalgebra osp(2,2n)$\oplus$sh(2,2n) discussed in Sec. II (if we omit the generator I of the central extension).

In order to go deeply into such a remark let us first come back to the one-dimensional context and compare the superalgebras osp(3,2) *and* osp(2,2)$\oplus$sh(2,2), characterized by the same dimension, 12. The last structure corresponds to the *twelve* generators $H_B, C_{\pm}$ [associated with so(2,1)$\sim$sp(2)], Y [associated with so(2)], $Q_{\pm}, S_{\pm}$—these eight operators being associated with osp(2,2)—and $P_{\pm}, T_{\pm}$ [associated with sh(2,2) up to the identity operator]. The main point is that the interchange between the even and odd characters of the nontrivial generators belonging only to the Heisenberg superalgebra sh(2,2) immediately transforms osp(2,2)$\oplus$sh(2,2) into osp(3,2). In fact, the even part of osp(3,2) will contain the algebra sp(2) $\oplus$ so(3), here generated by $\{H_B, C_{\pm}\}$ for sp(2), and by $\{Y, T_{\pm}\}$ for so(3) when we consider the generators $T_{\pm}$ as even [see Eq. (2.9)]. Then the odd part of osp(3,2) is generated by the supercharges $Q_{\pm}$ and $S_{\pm}$, as well as by the odd generators $P_{\pm}\sigma_3$ corresponding to the previous even $P_{\pm}$'s. In particular, these generators $P_{\pm}\sigma_3$ must now satisfy anticommutation relations according to Eqs. (2.14) so that ($\sigma_3^2 = I$)

$$\{P_+\sigma_3, P_-\sigma_3\} = \{P_+, P_-\} = 4H_B. \tag{3.2}$$

Let us here point out the fundamental role played by the new operators $P_{\pm}\sigma_3$ and $T_{\pm}$, since all the others can be obtained from their products as particular cases of Eqs. (2.1). For example, the supercharges

$$Q_{\pm} = (1/\sqrt{2})P_{\mp}T_{\pm}, \quad S_{\pm} = (1/\sqrt{2})P_{\pm}T_{\pm} \tag{3.3}$$

can be written as

$$Q_{+}^{(1)} = (1/\sqrt{2})(P_{-}\sigma_3)T_{+}^{(1)}, \quad Q_{-}^{(1)} = (1/\sqrt{2})T_{-}^{(1)}(P_{+}\sigma_3), \tag{3.4a}$$

$$S_{+}^{(1)} = (1/\sqrt{2})(P_{+}\sigma_3)T_{+}^{(1)}, \quad S_{-}^{(1)} = (1/\sqrt{2})T_{-}^{(1)}(P_{-}\sigma_3), \tag{3.4b}$$

according to the choice (2.25a) with $\varepsilon = +1$, or as

$$\begin{aligned} Q_{+}^{(-1)} &= (1/\sqrt{2})T_{+}^{(-1)}(P_{-}\sigma_3), \\ Q_{-}^{(-1)} &= (1/\sqrt{2})(P_{+}\sigma_3)T_{-}^{(-1)}, \end{aligned} \tag{3.5a}$$

$$\begin{aligned} S_{+}^{(-1)} &= (1/\sqrt{2})T_{+}^{(-1)}(P_{+}\sigma_3), \\ S_{-}^{(-1)} &= (1/\sqrt{2})(P_{-}\sigma_3)T_{-}^{(-1)}, \end{aligned} \tag{3.5b}$$

according to the choice (2.25b) with $\varepsilon = -1$. Indeed, the expressions (3.3) immediately result from (3.4) or (3.5) by the fact that we have

$$\sigma_3\sigma_{\pm} = \pm\sigma_{\pm}, \quad \sigma_{\pm}\sigma_3 = \mp\sigma_{\pm}. \tag{3.6}$$

Let us also note that the relations (3.4) and (3.5) immediately show the correspondences $Q_{+}^{(1)} \leftrightarrow S_{-}^{(-1)}$, $Q_{-}^{(1)} \leftrightarrow S_{+}^{(-1)}$ in connection with the choices $\varepsilon = +1$ or $\varepsilon = -1$, interesting information for the evaluation of the supersymmetric Hamiltonian proposed by different authors.[3,8,11]

By combining these results with those of Sec. II, it is now easy to extend the present considerations to the n-dimensional case and to the corresponding superalgebra osp(3,2n) containing the even algebra sp($2n$) $\oplus$ so(3). By maintaining the preceding so(3) algebra and extending the two odd generators ($P_{\pm}\sigma_3$) to $2n$ generators ($P_{\pm,k}\sigma_3$), we get the whole set of $(2n^2+7n+3)$ generators of osp(3,2n). Once again we point out the main role played in such a construction by the (odd) $P_{\pm,k}\sigma_3$ and (even) $T_{\pm}$ generators belonging *originally* to the fundamental Heisenberg superalgebra sh(2,2n).

For arbitrary n we effectively have $(2n^2+n+3)$ even operators [explicitly, the $n(2n+1)$ T_{kl} and $C_{\pm,kl}$ of sp($2n$) and the three generators of so(3)] *and* ($6n$) odd operators (explicitly, the $2n$ $P_{\pm,k}\sigma_3$ and the $4n$ "supercharges" $Q_{\pm,k}$ and $S_{\pm,k}$).

In correspondence with Eqs. (2.2), we clearly define

$$P_{\pm,k}\sigma_3 = \pm i\sqrt{2\omega}e^{\mp i\omega t}a_{\pm,k}\sigma_3 \tag{3.7a}$$

and

$$T_{\pm} = e^{\mp i\omega t}\xi_{\pm}, \tag{3.7b}$$

so that the definitions (2.1a) and (2.1b) of the T_{kl}'s and $C_{\pm,kl}$'s are unchanged (due to the fact that $\sigma_3^2 = I$), while the generator Y is still given by Eq. (2.9). Now if for brevity we take (2.25a) corresponding to $\varepsilon = +1$ for the variables $\xi_{\pm}$, we have (without superscript) the $4n$ charges extended from Eqs. (3.4):

$$\begin{aligned} Q_{+,k} &= (1/\sqrt{2})(P_{-,k}\sigma_3)T_{+}, \\ Q_{-,k} &= (1/\sqrt{2})T_{-}(P_{+,k}\sigma_3) \end{aligned} \tag{3.8a}$$

and

$$\begin{aligned} S_{+,k} &= (1/\sqrt{2})(P_{+,k}\sigma_3)T_{+}, \\ S_{-,k} &= (1/\sqrt{2})T_{-}(P_{-,k}\sigma_3). \end{aligned} \tag{3.8b}$$

The structure relations of osp(3,2n) can then be summarized by taking into account the new even or odd character of the above generators. Let us only insist here on the relations between the generators $T_{\pm}$ and $P_{\pm,k}\sigma_3$:

$$\begin{aligned} \{P_{-,k}\sigma_3, P_{+,l}\sigma_3\} &= 4T_{kl}, \\ \{P_{\pm,k}\sigma_3, P_{\pm,l}\sigma_3\} &= \pm 4iC_{\pm,kl}, \end{aligned} \tag{3.9}$$

$$\begin{aligned} [P_{\pm,k}\sigma_3, T_{\pm}] &= \pm 2\sqrt{2}S_{\pm,k}, \\ [P_{\mp,k}\sigma_3, T_{\pm}] &= \pm 2\sqrt{2}Q_{\pm,k}, \end{aligned} \tag{3.10}$$

and

$$[T_{+}, T_{-}] = (2/\omega)Y \tag{3.11}$$

[where (2.9) $\equiv$ (3.11)]. In fact, they can be used to define the generators $\{T_{kl}, C_{\pm,kl}, Y, Q_{\pm,k}, S_{\pm,k}\}$ of osp(2,2n) *inside* osp(3,2n). They clearly show that the operator I of a central extension does not appear here.

From these considerations and those developed in Sec. II, we clearly get the explicit form of the supersymmetric Hamiltonian for osp(3,2n), new information with respect to the Englefield contribution.[8]

IV. SOME SPECIFIC PHYSICAL CASES AND COMMENTS

Having at our disposal the arbitrary n-dimensional context, let us now come back to the specific cases $n = 1$ (Sec. IV A), $n = 2$ (Sec. IV B), and $n = 3$ Sec. IV C), which cover an important number of contributions dealing with physical applications in supersymmetric quantum mechanics.

A. The one-dimensional context

If $n = 1$, the numbers of bosonic and fermionic variables are the same, and we are working within the standard procedure of supersymmetrization "*á la* Witten."[2] Our superalgebra is the osp(2,2)$\oplus$sh(2,2) structure already discussed,[11,17] which corresponds to the largest or maximal dynamical invariance superalgebra of the one-dimensional harmonic oscillator. Through the considerations of Sec. III, it is related to the superalgebra osp(3,2) studied by Van der Jeugt[9] and Englefield.[8] Let us point out once again the inclusion

$$\text{osp}(3,2) \supset \text{so}(2,1) \oplus \text{so}(3). \tag{4.1}$$

In order to make clear the connections between our generators and those defined in Englefield,[8] let us mention *once* the twelve correspondences:

$$\begin{aligned} &(H_B, C_{+}, C_{-}) \leftrightarrow (2t_0, 2it_{+}, 2it_{-}), \\ &(T_{+}, T_{-}, Y) \leftrightarrow (s_{-}, s_{+}, -s_0), \\ &(Q_{+}, Q_{-}, S_{+}, S_{-}) \leftrightarrow (-iR_{-1-}, iR_{1+}, -iR_{-1+}, iR_{1-}), \\ &(P_{\pm}\sigma_3) \leftrightarrow (2iR_{0,\pm}). \end{aligned}$$

We also notice that the Englefield choice of the Grassmannian variables corresponds to $\varepsilon = -1$ [cf. Eq. (2.25b)], so that the supercharges leading to the supersymmetric Hamiltonian are of type S. The Englefield results[8] clearly apply here, and it is easy to recognize the osp(1,2) and osp(2,2) substructures, etc.

From Secs. II and III we get the well-known and expected Hamiltonian

$$H^{SS}_{(n=1)} = H_B + Y, \quad H_B = T_{11}, \tag{4.2}$$

whose specific form (with $\varepsilon = +1$) is, according to Ravndal,[13]

$$H^{SS}_{(n=1)} = \tfrac{1}{2}(p^2 + \omega^2 x^2) + (\omega/2)\sigma_3. \tag{4.3}$$

This appears to be a particular case of Eq. (2.22) or (2.26) where there are neither L_{kl} nor M_{kl} terms ($k = l = 1$).

B. The two-dimensional context

If the dynamical *and* kinematical (super)symmetries of the two-dimensional harmonic oscillator are considered, we know[3] that the maximal dynamical invariance algebra is osp(4,4)⊕sh(4,4) within the standard procedure, while it is osp(2,4)⊕sh(2,4) within the spin-orbit procedure. The latter superstructure may clearly be put in correspondence with the Englefield superalgebra osp(3,4) in the way developed in Sec. III. The 25 generators of both contexts can be associated with each other, and all the above conclusions apply here (dimension, even and odd generators, fundamental role of sh(2,2) generators, etc.). Let us note that we clearly obtain now eight "supercharges" of type Q and type S, and there are four Heisenberg operators that have to become odd ones through multiplication by σ_3, besides the two even operators $T_\pm$.

This $n = 2$ context of the harmonic oscillator is particularly interesting from a physical point of view; it is directly connected[3,14,19,20] with the problem of an electron interacting with an external *constant* magnetic field. Such an application enhances the spin-orbit coupling supersymmetrization procedure in a very natural way.[3] Indeed, in this two-dimensional case the supersymmetric Hamiltonian (2.22) contains at most one β and one γ coefficient, defined by Eqs. (2.21), such that Eqs. (2.23) are satisfied. It is clearly impossible to suppress both coefficients simultaneously: either the term L_{12} or the term M_{12} will survive.

Let us consider the first case, the most interesting one, since $L_{12} \equiv L$ is clearly the third component of the orbital angular momentum. Then we define the following supercharges according to Eqs. (2.15) with (2.17). Constrained by the embedding of the $n = 1$ context, they become

$$Q_\pm = Q_{\pm,1} \pm iQ_{\pm,2}. \tag{4.4}$$

This corresponds to the choice of coefficients

$$\alpha_{+,1} = \alpha_{-,1} = 1, \quad \alpha_{+,2} = -\alpha_{-,2} = i, \tag{4.5a}$$

ensuring the cancellation of the symmetric terms ($\gamma_{12} = \gamma_{21} = 0$) and the reduction of the skew-symmetric ones to a single term ($\beta_{12} = -\beta_{21} = -1$) in the Hamiltonian. We then get

$$H^{SS(a)}_{(n=2)} = H_B + 2Y - \omega L. \tag{4.6a}$$

Let us also note that the choice (4.5a) is not unique. If we take, for example,

$$\alpha_{+,1} = \alpha_{-,1} = 1, \quad \alpha_{+,2} = -\alpha_{-,2} = -i, \tag{4.5b}$$

we correspondingly get

$$H^{SS(b)}_{(n=2)} = H_B + 2Y + \omega L. \tag{4.6b}$$

Moreover, if we combine the above contexts with the possible realizations (2.25) of the variables $\xi_\pm$, we can construct four different types of Hamiltonians (4.6). In particular, with $\varepsilon = -1$, and the context (a), we get the already studied Hamiltonian[3] that coincides with the Englefield one[8] in this $n = 2$ case. It is written

$$H^{SS(a)}_{(n=2)(\varepsilon=-1)} = H_B - \omega\sigma_3 - \omega L. \tag{4.7}$$

All the results (4.6) and (4.7) are typical of the spin-orbit coupling procedure of supersymmetrization for which we have associated an algebra[3] modified with respect to the usual Clifford algebra (2.8). Indeed, we have shown that in this $N = 2$ supersymmetric quantum theory we have

$$\{\xi_{+,k}, \xi_{-,l}\} = \delta_{kl} - i\Xi_{kl}, \tag{4.8}$$

where $\Xi_{kl} = -\Xi_{lk}$, $(\Xi)^\dagger = \Xi$. It is easy to identify these quantities in the present context. For example, in connection with Eq. (4.7) and $\varepsilon = -1$ we get

$$\xi_{+,1} = -i\xi_{+,2} = \sigma_-, \quad \xi_{-,1} = i\xi_{-,2} = \sigma_+,$$
$$\Xi_{12} = -\Xi_{21} = I.$$

Further discussions on corresponding Pauli Hamiltonians can be found elsewhere.[3,14,21]

C. The three-dimensional context

This case is physically the most interesting one and has evidently been considered in different contributions from the supersymmetric point of view (see the references previously mentioned and compared).[3,10,11,16]

Within the standard procedure,[2] we have extended the de Crombrugghe–Rittenberg result[1] [osp(6,6)] to the superalgebra osp(6,6)⊕sh(6,6) as the maximal dynamical invariance algebra of the three-dimensional harmonic oscillator.

Besides these considerations we also know, since the Balantekin contribution,[10] that the spin-orbit coupling procedure leads to the maximal superstructure osp(2,2) ⊕ so(3) [which has been generalized to osp(2,2) ⊕ so(n) by Kostelecky *et al.*[12] in the n-dimensional case]. In particular, we know[11] that the symmetries of $\mathbf{P}_\pm$ and $\mathbf{T}_\pm$ types belonging to sh(6,6) are broken due to the presence of the spin-orbit coupling term in the Hamiltonian.

Following the arguments of Sec. II, we propose here the new superalgebra osp(2,6)⊕sh(2,6) and determine its supersymmetric Hamiltonian. In addition, we relate this structure to the superalgebra osp(3,6) proposed by Englefield.[8] So we immediately get the 42 generators and their structure relations. We notice that the six even generators $\mathbf{P}_\pm$ belonging to sh(2,6) are replaced by the odd generators $\mathbf{P}_\pm\sigma_3$ in order to give osp(3,6).

From the above arguments it is evident that neither the osp(2,6)⊕sh(2,6) *nor* the osp(3,6) superalgebras can be associated with the spin-orbit coupling procedure. Another interesting way to show this is to study once again the expressions (2.23) of the coefficients β_{kl} and γ_{kl} in this $n = 3$ context in a way parallel to the one discussed for $n = 2$. It is easy to show that the maximal simplification always maintains three terms in L_{kl} and M_{kl} in the Hamiltonian, but that at most two of them are angular momentum operators.

In particular, by embedding the $n=1,2$ cases in the three-dimensional context, let us construct here the two supercharges according to the choice (4.5a) as follows:

$$Q_{\pm} = Q_{\pm,1} \pm iQ_{\pm,2} \pm iQ_{\pm,3}. \tag{4.9}$$

The corresponding coefficients β_{kl} and γ_{kl} ($k \neq l$, $k,l = 1,2,3$) are easily determined, and we get Eq. (2.22) in the form

$$H^{SS(a)}_{(n=3)} = H_B + 3Y - \omega(L_{12} + L_{13}) + \omega M_{23}, \tag{4.10}$$

and, more precisely, with the choice $\varepsilon = +1$ [cf. Eq. (2.25a)],

$$H^{SS(a)}_{(n=3)(\varepsilon=1)} = H_B + \tfrac{1}{2}\omega\sigma_3 - \omega(L_{12} + L_{13}) + \omega M_{23}. \tag{4.11}$$

Such supersymmetric Hamiltonians admit the superalgebra $\mathrm{osp}(2,6) \oplus \mathrm{sh}(2,6)$ *or* $\mathrm{osp}(3,6)$ as an invariance superalgebra.

V. GENERAL CONCLUSIONS

Since our first contribution[17] to the symmetry superalgebras of the harmonic oscillator we have insisted on the prominent part played by the Heisenberg operators. Such a role is once again remarkable here not only in the structures $\mathrm{osp}(2n,2n) \oplus \mathrm{sh}(2n,2n)$ and, consequently, in $\mathrm{osp}(2,2n) \oplus \mathrm{sh}(2,2n)$, but also through the relation with $\mathrm{osp}(3,2n)$. Remember (as detailed in Sec. III) that by considering only the operators belonging originally to the fundamental Heisenberg superalgebra, we have explained the connection between $\mathrm{osp}(2,2n) \oplus \mathrm{sh}(2,2n)$ and $\mathrm{osp}(3,2n)$. We have also shown that *all* the other generators belonging to these structures can be expressed in terms of the Heisenberg generators. Here, moreover, we want to point out that these expressions in terms of the Heisenberg operators can be considered as *definitions* in the context of $\mathrm{osp}(2,2n) \oplus \mathrm{sh}(2,2n)$, but that they become *structure relations* of $\mathrm{osp}(3,2n)$. As an example, compare the *definitions* given by Eqs. (2.1a) and (2.1b) in $\mathrm{osp}(2,2n) \oplus \mathrm{sh}(2,2n)$ and the *structure relations* summarized in Eqs. (3.9) for $\mathrm{osp}(3,2n)$.

Another conclusion here lies in summarizing much information concerning the invariance superalgebras for the n-dimensional harmonic oscillator. The superalgebra $\mathrm{osp}(2n,2n) \oplus \mathrm{sh}(2n,2n)$ is *the* largest or maximal *dynamical* invariance superstructure containing dynamical as well as kinematical (super)symmetries when the standard procedure of supersymmetrization is under study.[3] We also have the following "dynamical $\supset$ kinematical" inclusion[3]:

$$[\mathrm{osp}(2n,2n) \oplus \mathrm{sh}(2n,2n)] \supseteq [[\mathrm{osp}(2,2) \oplus \mathrm{so}(n)] \oplus \mathrm{sh}(2n,2n)], \tag{5.1}$$

where the last superalgebra is the maximal *kinematical* one. Let us also insist on the (trivial) inclusion

$$[[\mathrm{osp}(2,2) \oplus \mathrm{so}(n)] \oplus \mathrm{sh}(2n,2n)] \supset \mathrm{osp}(2,2) \oplus \mathrm{so}(n), \tag{5.2}$$

where we recognize the superalgebra found by Kostelecky *et al.*[12] as the (kinematical) invariance structure obtained within the spin-orbit coupling procedure, as a result generalizing the Balantekin study.[10] Moreover, let us mention the other (trivial) inclusion

$$\mathrm{osp}(2n,2n) \oplus \mathrm{sh}(2n,2n) \supseteq \mathrm{osp}(2,2n) \oplus \mathrm{sh}(2,2n) \tag{5.3}$$

[the equalities in Eqs. (5.1) and (5.3) being ensured only for $n=1$], where the last superalgebra appears (see Sec. II) if we require a minimum number of fermionic degrees of freedom. Finally, from Sec. III we also have

$$\mathrm{osp}(2,2n) \oplus \mathrm{sh}(2,2n) \sim \mathrm{osp}(3,2n), \tag{5.4}$$

connecting the Englefield structure[8] with ours and showing that $\mathrm{osp}(3,2n)$ *does contain* dynamical as well as kinematical (super)symmetries.

If $n=1$, then only the standard procedure is meaningful, and in this context the dynamical or kinematical superalgebra is[8,17]

$$\mathrm{osp}(2,2) \oplus \mathrm{sh}(2,2) \sim \mathrm{osp}(3,2), \tag{5.5}$$

the Kostelecky *et al.* result[12] being excluded for $n=1$.

If $n=2$, then the standard and spin-orbit procedures differ and lead, respectively, to the $\mathrm{osp}(4,4) \oplus \mathrm{sh}(4,4)$ and $\mathrm{osp}(2,4) \oplus \mathrm{sh}(2,4)$ superalgebras,[3] which evidently are such that

$$\mathrm{osp}(4,4) \oplus \mathrm{sh}(4,4) \supset \mathrm{osp}(2,4) \oplus \mathrm{sh}(2,4) \sim \mathrm{osp}(3,4). \tag{5.6}$$

In addition, we also have

$$\mathrm{osp}(3,4) \supset \mathrm{osp}(2,2) \oplus \mathrm{so}(2), \tag{5.7}$$

an inclusion that allocates the place of the Kostelecky *et al.* contribution and ignores the Heisenberg (super)symmetries.

If $n=3$, then both procedures once again differ, and from the largest dynamical superalgebra[3] we have

$$\mathrm{osp}(6,6) \oplus \mathrm{sh}(6,6) \supset \mathrm{osp}(2,6) \oplus \mathrm{sh}(2,6) \tag{5.8}$$

and

$$\mathrm{osp}(2,6) \oplus \mathrm{sh}(2,6) \sim \mathrm{osp}(3,6). \tag{5.9}$$

From the results of Sec. IV C, we immediately see that the Balantekin structure $[\mathrm{osp}(2,2) \oplus \mathrm{so}(3)]$ has nothing to do with the structure (5.9), but only, through (5.2), with

$$[\mathrm{osp}(2,2) \oplus \mathrm{so}(3)] \oplus \mathrm{sh}(6,6) \supset \mathrm{osp}(2,2) \oplus \mathrm{so}(3). \tag{5.10}$$

This says that for the spin-orbit coupling procedure all the Heisenberg (super)symmetries are broken[11] in this three-dimensional context. Let us recall[11] that an easy way to recover these spin-orbit coupling results[10] is the construction of 4×4 supercharges Q_+ and Q_- defined by

$$Q_{\pm} = \sigma_k \otimes Q_{\pm,k}, \tag{5.11}$$

which can be compared with our present definitions Eqs. (2.15). We immediately see that, with $\varepsilon = +1$, we get here four Grassmannian variables

$$\xi_{\pm,k} = \sigma_k \otimes \sigma_{\pm}, \tag{5.12}$$

satisfying (4.8) instead of (2.4). They lead to the Balantekin Hamiltonian

$$H^{SS}_{(n=3)} = \begin{pmatrix} H_B + \frac{1}{2}\omega + \omega\boldsymbol{\sigma}\cdot\mathbf{L} & 0 \\ 0 & H_B - \frac{1}{2}\omega - \omega\boldsymbol{\sigma}\cdot\mathbf{L} \end{pmatrix}, \tag{5.13}$$

admitting only $\mathrm{osp}(2,2) \oplus \mathrm{so}(3)$ as the invariance superalgebra.

ACKNOWLEDGMENT

We want to thank Dr. M. J. Englefield for sending us his results prior to publication.

[1]M. de Crombrugghe and V. Rittenberg, Ann. Phys. (NY) **151**, 99 (1983).
[2]E. Witten, Nucl. Phys. B **188**, 513 (1981).
[3]J. Beckers, D. Dehin, and V. Hussin, J. Phys. A **21**, 651 (1988).
[4]U. Niederer, Helv. Phys. Acta **46**, 191 (1973).
[5]S. Fubini and E. Rabinovici, Nucl. Phys. B **245**, 17 (1984).
[6]B. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[7]U. Niederer, Helv. Phys. Acta **45**, 802 (1972); C. R. Hagen, Phys. Rev. D **5**, 377 (1972).
[8]M. J. Englefield, J. Phys. A **21**, 1309 (1988).
[9]J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).
[10]A. B. Balantekin, Ann. Phys. (NY) **164**, 277 (1985).
[11]J. Beckers, D. Dehin, and V. Hussin, J. Phys. A **20**, 1137 (1987).
[12]V. A. Kostelecky, M. M. Nieto, and D. R. Truax, Phys. Rev. D **32**, 2627 (1985).
[13]F. Ravndal, *Proceedings of the 1984 Cern School of Physics* (CERN, Geneva, 1985), 85-11, p. 300.
[14]S. Durand, M.S. thesis, Université de Montréal, 1985.
[15]E. Gozzi, Phys. Lett. B **129**, 432 (1983).
[16]J. Gamboa and J. Zanelli, Phys. Lett. B **165**, 91 (1985).
[17]J. Beckers and V. Hussin, Phys. Lett. A **118**, 319 (1986).
[18]D. H. Sattinger and D. L. Weaver, *Lie Groups and Algebras with Applications to Physics, Geometry and Mechanics* (Springer, New York, 1986).
[19]D. Dehin and V. Hussin, Helv. Phys. Acta **60**, 552 (1987).
[20]D. Dehin and V. Hussin, Int. J. Mod. Phys. A **3**, 487 (1988).
[21]E. D'Hoker, L. Vinet, and V. A. Kostelecky, "Spectrum Generating Algebras," in *Dynamical Groups and Spectrum Generating Algebras*, edited by A. Barut, A. Bohm, and Y. Ne'eman (World Scientific, Singapore, 1987).

Three-dimensional scattering of impulsive acoustic waves by a semi-infinite crack in the plane interface of a half-space and a layer

Bert Jan Kooij and Dirk Quak
Delft University of Technology, Faculty of Electrical Engineering, Laboratory of Electromagnetic Research, P.O. Box 5031, 2600 GA Delft, The Netherlands

(Received 15 July 1987; accepted for publication 9 March 1988)

The three-dimensional diffraction of a pulsed acoustic wave by a semi-infinite crack located in the interface of a uniform layer and a half-space is investigated theoretically. In the analysis the influence of shear stresses in the material structure is neglected. The incident acoustic wave is taken to be generated by an impulsive compressional point source located at the top boundary of the layer. With the aid of the Wiener–Hopf technique and a modified version of the Cagniard–de Hoop technique implemented in an iterative scheme, closed-form expressions for the particle velocity within any finite time window anywhere at the top boundary of the layer are obtained.

I. INTRODUCTION

The problem of diffraction of acoustic waves by a crack in the interface of two different solids is of interest in the field of nondestructive evaluation of material structures. With this application in mind, we investigate the diffraction of acoustic waves by a semi-infinite crack in the interface of a uniform layer and a half-space. The configuration is excited by an impulsive compressional point source located at the top boundary of the layer, where the receiver is also located. The resulting acoustic wave motion is analyzed as far as its compressional components are concerned, i.e., the influence of shear stresses is neglected (the equivalent fluid model).

Considerable work has already been done in the field of diffraction of transient waves by a semi-infinite crack in a homogeneous medium of infinite extent. In this respect we mention the paper by Sommerfeld[1] on diffraction of transient electromagnetic waves by a semi-infinite screen and the papers by Cagniard,[2] De Hoop,[3] and Du Cloux[4] on diffraction of transient acoustic waves by a semi-infinite crack in a medium of infinite extent. The time-domain diffraction of electromagnetic waves by a semi-infinite screen in the interface of two different half-spaces has been studied by Du Cloux.[5] However, unlike in Du Cloux's[5] problem, in our case the presence of the layer next to the diffracting edge gives rise to multiple reflections and repeated diffraction, with considerable consequences for the analysis. Multiple edge diffraction has been studied by Shirai and Felsen,[6] who investigated the diffraction of pulsed electromagnetic waves by a strip. We solve the problem by a proper combination of a modified version of the Cagniard–de Hoop technique[2–4,7–9] and the Wiener–Hopf technique, or factorization method,[10,11] for three-dimensional configurations. These techniques are implemented in an iterative scheme that has the property that an exact analytic expression for the time-domain solution is obtained within any finite time window. In this paper the emphasis lies on the methodological aspects in the analysis.

II. DESCRIPTION OF THE CONFIGURATION

The configuration consists of a uniform layer D_1 and a semi-infinite medium occupying the half-space D_2: Their acoustic properties are characterized by a positive constant volume density of mass ρ and a positive constant compressibility κ. The corresponding compressional wave speed is $c = (\rho\kappa)^{-1/2}$. We let

$$\rho = \rho_1, \quad \kappa = \kappa_1, \quad c = c_1 \quad \text{in } D_1, \tag{1}$$

$$\rho = \rho_2, \quad \kappa = \kappa_2, \quad c = c_2 \quad \text{in } D_2. \tag{2}$$

To locate a point in the configuration, we employ the coordinates $\{x, y, z\}$ with respect to a given orthogonal Cartesian reference frame that is specified by an origin 0 and three mutually perpendicular base vectors of unit length $\{\mathbf{i}_x, \mathbf{i}_y, \mathbf{i}_z\}$ that in the given order form a right-hand system. The position vector is denoted by $\mathbf{r} = x\mathbf{i}_x + y\mathbf{i}_y + z\mathbf{i}_z$. The reference frame is chosen such that (see Fig. 1)

$$D_1 = \{\mathbf{r}\in\mathbb{R}^3 | (x,y)\in\mathbb{R}^2, \ z\in(0,d)\}, \tag{3}$$

$$D_2 = \{\mathbf{r}\in\mathbb{R}^3 | (x,y)\in\mathbb{R}^2, \ z\in(d,\infty)\}. \tag{4}$$

The semi-infinite gap S coincides with the half-plane

$$S = \{\mathbf{r}\in\mathbb{R}^3 | x\in(0,\infty), \ y\in\mathbb{R}, \ z = d\}. \tag{5}$$

The time coordinate is denoted by t. Partial differentiation is denoted by ∂ and $\nabla = \mathbf{i}_x \partial_x + \mathbf{i}_y \partial_y + \mathbf{i}_z \partial_z$.

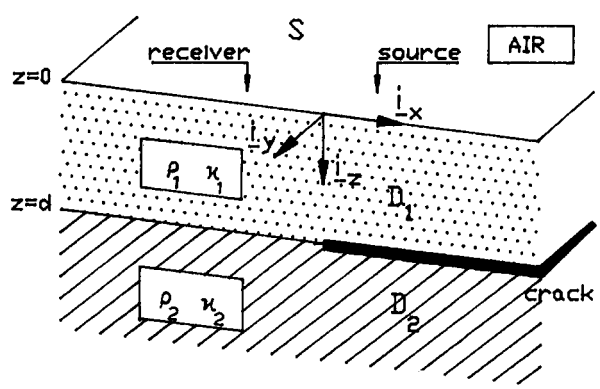


FIG. 1. Geometrical configuration.

 0022-2488/88/071712-10$02.50

III. THE ACOUSTIC WAVE FIELD IN THE CONFIGURATION

The acoustic wave field in the configuration is described in terms of the acoustic pressure p and the particle velocity $\mathbf{v}$. In the interior of the domains D_1 and D_2, these quantities satisfy the source-free acoustic equations

$$\nabla p + \rho\, \partial_t \mathbf{v} = 0\,, \tag{6}$$

$$\nabla\cdot\mathbf{v} + \kappa\, \partial_t p = 0\,. \tag{7}$$

In accordance with the notation adopted in Sec. II, we write

$$p = p_1 \quad \text{and} \quad \mathbf{v} = \mathbf{v}_1 \quad \text{in } D_1\,, \tag{8}$$

$$p = p_2 \quad \text{and} \quad \mathbf{v} = \mathbf{v}_2 \quad \text{in } D_2\,. \tag{9}$$

At the upper boundary of the layer a pressure-type source is located at $x = x_s, y = y_s, z = 0$. Then, we have the boundary condition

$$\lim_{z\downarrow 0} p_1(x,y,z,t) = P(t)\delta(x - x_s, y - y_s)\,. \tag{10}$$

At the interface of the two media the following boundary conditions hold:

$$\lim_{z\uparrow d}\{p_1, v_{1,z}\} = \lim_{z\downarrow d}\{p_2, v_{2,z}\} \text{ when } -\infty < x < 0,\ y\in\mathbb{R}\,, \tag{11}$$

and

$$\lim_{z\uparrow d} p_1 = \lim_{z\downarrow d} p_2 = 0 \quad \text{when } 0 < x < \infty,\ y\in\mathbb{R}\,. \tag{12}$$

We assume that the source starts to act at the instant $t = 0$. Prior to this instant, no acoustic wave field is present in the configuration (the initial condition).

IV. FORMULATION OF THE SPECTRAL-DOMAIN ACOUSTIC WAVE FIELD PROBLEM

To carry out our analysis, we cast the field representations in a particular form characteristic of the modified Cagniard technique. First we subject the field quantities to a one-sided Laplace transformation with respect to time. One of the characteristic features of the Cagniard technique is that the relevant transform variable s is taken to be real and positive. To show the notation we write the transformation of the acoustic pressure as

$$\hat{p}(x,y,z,s) = \int_0^\infty \exp(-st)p(x,y,z,t)dt\,. \tag{13}$$

Next, we subject the quantities to a spatial Fourier transformation with respect to the coordinates x and y. In view of the future application of the modified Cagniard–de Hoop technique we write this Fourier transformation as a two-sided Laplace transformation (Van der Pol and Bremmer[12]) for which the transform variables α and β are purely imaginary. For the acoustic pressure we have

$$\tilde{p}(\alpha,\beta,z,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{p}(x,y,z,s)\exp[s(\alpha x + \beta y)]dx\,dy,$$

$$\text{with } \alpha\in I,\quad \beta\in I\,, \tag{14}$$

with the inverse transformation

$$\hat{p}(x,y,z,s) = \left(\frac{s}{2\pi i}\right)^2 \int_{-i\infty}^{i\infty} d\beta \int_{-i\infty}^{i\infty} \tilde{p}(\alpha,\beta,z,s) \times \exp[-s(\alpha x + \beta y)]d\alpha\,. \tag{15}$$

In (14) and (15), s is real and positive.

After applying the transformations (13) and (14) to Eqs. (6) and (7) and eliminating $\tilde{v}_x$ and $\tilde{v}_y$ we are left with a system of two first-order differential equations in z in which $\tilde{p}$ and $\tilde{v}_z$ occur as unknowns. These equations have constant coefficients and hence have exponential functions as their solutions. We write

$$\tilde{p}_1 = A_1^+ \exp(-s\gamma_1 z) + A_1^- \exp[-s\gamma_1(d-z)]\,, \quad \text{with } 0<z<d\,, \tag{16}$$

$$\tilde{p}_2 = A_2^+ \exp[-s\gamma_2(z-d)], \text{ with } d<z<\infty\,, \tag{17}$$

$$\tilde{v}_{1,z} = Y_1\{A_1^+ \exp(-s\gamma_1 z) - A_1^- \exp[-s\gamma_1(d-z)]\}\,, \quad \text{with } 0<z<d\,, \tag{18}$$

$$\tilde{v}_{2,z} = Y_2 A_2^+ \exp[-s\gamma_2(z-d)], \quad \text{with } d<z<\infty\,, \tag{19}$$

in which $A_{1,2}^{\pm} = A_{1,2}^{\pm}(\alpha,\beta,s)$ do not depend on z and where

$$Y_{1,2} = \gamma_{1,2}/\rho_{1,2} \tag{20}$$

is the vertical acoustic wave admittance and

$$\gamma_{1,2} = (1/c_{1,2}^2 - \alpha^2 - \beta^2)^{1/2}, \quad \text{with } \mathrm{Re}(\gamma_{1,2}) > 0 \tag{21}$$

is the vertical slowness. Using the boundary condition (10) and the continuity of the pressure at $z = d$ [cf. (11) and (12)] it follows from (16) and (17) that

$$A_1^+ = [\tilde{P}^i - A_2^+ \exp(-s\gamma_1 d)]/[1 - \exp(-2s\gamma_1 d)] \tag{22}$$

and

$$A_1^- = [A_2^+ - \tilde{P}^i \exp(-s\gamma_1 d)]/[1 - \exp(-2s\gamma_1 d)]\,, \tag{23}$$

with

$$\tilde{P}^i = \hat{P}(s)\exp[s(\alpha x_s + \beta y_s)]\,. \tag{24}$$

Next we introduce the two-media plane-wave reflection coefficient R as

$$R = (Y_1 - Y_2)/(Y_1 + Y_2) \tag{25}$$

and the two-media plane-wave transmission coefficient T as

$$T = 2Y_1/(Y_1 + Y_2)\,. \tag{26}$$

In order for the boundary conditions that are x dependent to be satisfied sufficient conditions are further provided by [cf. (11) and (12)]

$$A_2^+ = G^R \tag{27}$$

and

$$T\tilde{P}^i \exp(-s\gamma_1 d) - A_2^+ [1 + R\exp(-2s\gamma_1 d)] = H^L[1 - \exp(-2s\gamma_1 d)]/[\rho_1(Y_1 + Y_2)]\,, \tag{28}$$

in which $G^R = G^R(\alpha,\beta,s)$ is an analytic function of α in $D^R\subset\mathbb{C}$ with $G^R = o(1)$ as $|\alpha|\to\infty$ in D^R and $H^L = H^L(\alpha,\beta,s)$ is an analytic function of α in $D^L\subset\mathbb{C}$ with $H^L = o(1)$ as $|\alpha|\to\infty$ in D^L while $\beta\in I$. The regions D^R and D^L are shown in Fig. 2. After substituting A_2^+ from (27) into (28), the resulting equation constitutes the Wiener–Hopf equation

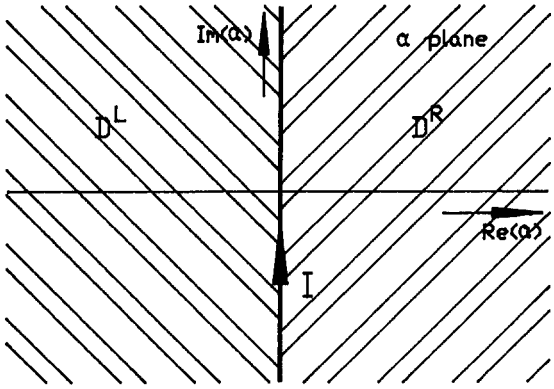


FIG. 2. The contour I that divides the complex α plane into the regions D^R and D^L.

$$T\tilde{P}^i \exp(-s\gamma_1 d) - G^R[1 + R\exp(-2s\gamma_1 d)] = H^L[1 - \exp(-2s\gamma_1 d)]/[\rho_1(Y_1 + Y_2)], \quad \text{with } \alpha \in I. \tag{29}$$

V. SOLUTION OF THE WIENER–HOPF PROBLEM

In this section we solve (29) in an iterative way. To this end we write

$$G^R = \sum_{n=0}^{\infty} G_n^R, \tag{30}$$

$$H^L = \sum_{n=0}^{\infty} H_n^L, \tag{31}$$

in which the terms $G_n^R = G_n^R(\alpha, \beta, s)$ and $H_n^L = H_n^L(\alpha, \beta, s)$ will be shown to be of the exponential order $\exp[-(2n+1)s\gamma_1 d]$. We start by solving

$$G_0^R + H_0^L/[\rho_1(Y_1 + Y_2)] = T\tilde{P}^i \exp(-s\gamma_1 d), \quad \text{with } \alpha \in I. \tag{32}$$

Next, we relate G_n^R and H_n^L to G_{n-1}^R and H_{n-1}^L via

$$G_n^R + H_n^L/[\rho_1(Y_1 + Y_2)] = \tilde{U}_n^i, \quad \text{with } \alpha \in I, \tag{33}$$

in which

$$\tilde{U}_n^i = \{H_{n-1}^L/[\rho_1(Y_1 + Y_2)] - RG_{n-1}^R\} \times \exp(-2s\gamma_1 d), \quad \text{with } n \in \{1,2,3,\ldots\}. \tag{34}$$

The function $\tilde{U}_n^i = \tilde{U}_n^i(\alpha, \beta, s)$ can be viewed as the *excitation* function in the nth Wiener–Hopf equation (32). By defining $\tilde{U}_0^i$ as

$$\tilde{U}_0^i = T\tilde{P}^i \exp(-s\gamma_1 d), \tag{35}$$

we can rewrite (33) as

$$G_n^R = \rho_2 H_n^L/[(\rho_1 + \rho_2)\gamma_1 K] + \tilde{U}_n^i, \quad \text{with } \alpha \in I \text{ and } n \in \{0,1,2,\ldots\}, \tag{36}$$

in which

$$K = (\gamma_2/\gamma_1 + \rho_2/\rho_1)/(1 + \rho_2/\rho_1). \tag{37}$$

The kernel function $K = K(\alpha, \beta)$ has been constructed such that $|K(\alpha, \beta)| \to 1$ as $|\alpha| \to \infty$ uniformly in the complex α plane. The formal solution of the Wiener–Hopf problem (29) follows from (36) provided that the series expansions in (30) and (31) converge. The proof of this convergence depends on the detailed structure of the solutions to the Wiener–Hopf equations of the type (36). The solutions of (36) follow by the application of the standard Wiener–Hopf method (see Appendix A). In this method we need the factorized parts of the kernel function K and the propagation coefficient γ_1. The factorized parts $K^{R,L}$ of the kernel K, i.e., the representations analytic in $D^{R,L}$, respectively, can be obtained in the manner indicated by Du Cloux,[5] who also derived useful approximate representations for them. The factorized parts $\gamma_1^{R,L}$ of γ_1, i.e., the representations analytic in $D^{R,L}$, respectively, can be found by inspection and are given by

$$\gamma_1^{R,L} = [\Omega_1(\beta) \pm \alpha]^{1/2}, \quad \text{when } \alpha \in I \cup D^{R,L}, \quad \beta \in I, \tag{38}$$

in which

$$\Omega_{1,2}(\beta) = [(1/c_{1,2}^2) - \beta^2]^{1/2}. \tag{39}$$

In (38) the upper sign corresponds to the quantities with the superscript R and the lower sign to the quantities with the superscript L. By introducing the function $\tilde{Q}_n = \tilde{Q}_n(\alpha, \beta, s)$ as

$$\tilde{Q}_n = \tilde{U}_n^i K^R \gamma_1^R, \quad \text{when } \alpha \in I, \text{ with } n \in \{0,1,2,\ldots\}, \tag{40}$$

the solutions of the Wiener–Hopf equations (36) become

$$G_n^R = \tilde{Q}_n^R/K^R\gamma_1^R, \quad \text{when } \alpha \in I \cup D^R, \text{ with } n \in \{0,1,2,\ldots\} \tag{41}$$

and

$$H_n^L = (1 + \rho_1/\rho_2)\gamma_1^L K^L \tilde{Q}_n^L, \quad \text{when } \alpha \in I \cup D^L, \text{ with } n \in \{0,1,2,\ldots\}, \tag{42}$$

where for any $\tilde{Q}_n$, $\tilde{Q}_n^{R,L} = \tilde{Q}_n^{R,L}(\alpha, \beta, s)$ are defined as

$$\tilde{Q}_n^{R,L} = \Phi_{(I)}^{R,L}(\tilde{Q}_n), \quad \text{with } n \in \{0,1,2,\ldots\}. \tag{43}$$

In (43) we have used the *integral operator* $\Phi_{(I)}^{R,L}$ that is defined as

$$\Phi_{(L)}^{R,L}(\tilde{Q}_n(\alpha,\beta,s)) = \mp\left(\frac{1}{2\pi i}\right)\int_L \frac{\tilde{Q}_n(q,\beta,s)}{q - \alpha}\, dq, \quad \text{when } \alpha \in D_L^{R,L} \tag{44}$$

and

$$\Phi_{(L)}^{R,L}(\tilde{Q}_n(\alpha,\beta,s)) = \frac{1}{2}\tilde{Q}_n(\alpha,\beta,s) \mp \frac{1}{2\pi i}\int_L \frac{\tilde{Q}_n(q,\beta,s)}{q - \alpha}\, dq, \quad \text{when } \alpha \in L \tag{45}$$

for any contour L that divides the complex plane into a part D_L^R to the right and a part D_L^L to the left of it and any function $\tilde{Q}_n$ that satisfies a Hölder condition on L. In our case the contour L thus far corresponds to the imaginary axis I; in this case $D_L^{R,L} = D^{R,L}$. In (44) and (45) the upper signs correspond to the quantities with the superscript R and the lower signs to the quantities with the superscript L and in (45) the integral is a Cauchy principal-value integral. The expression for $Q_n^{R,L}$ with $\alpha \in I$ [cf. (43) and (45)] can be derived from (44) by applying a limiting procedure when α approaches L via the regions D_L^R or D_L^L, respectively. In Figs. 3 and 4 the location of the integral contour on the

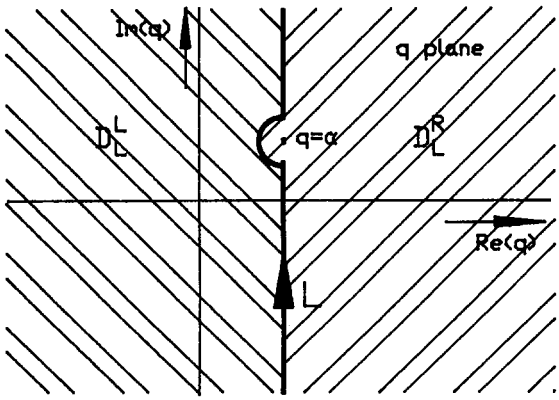


FIG. 3. The contour L pertaining to the operator $\Phi_{(L)}^R$ with $\alpha \in L$, which divides the complex q plane into the regions D_L^R and D_L^L.

imaginary axis and the semicircle around the pole on I are shown pertaining to the functions Q_n^R and Q_n^L, respectively.

VI. TRANSFORM-DOMAIN WAVE CONSTITUENTS

For the transformation of the particle velocity $\tilde{v}_{1,z}$ [cf. (18)] back to the space domain it is necessary to distinguish between the different *wave constituents*, i.e., the different terms in the representation of $\tilde{v}_{1,z}$ that have a particular exponential function of the type $\exp(-2ms\gamma_1 d)$, with $m \in N$. First we derive from (18), using (22)–(24) and (27), an expression for the z component of the particle velocity taken at the surface of the layer. When expressed in terms of G^R, we have

$$\tilde{v}_{1,z}(\alpha,\beta,0,s) = Y_1\{\tilde{P}^i[1+\exp(-2s\gamma_1 d)] - 2G^R \exp(-s\gamma_1 d)\} \times [1-\exp(-2s\gamma_1 d)]^{-1}, \quad \text{with } \alpha \in I \cup D^R \tag{46}$$

and in terms of H^L, using (29) in (46), we have

$$\tilde{v}_{1,z}(\alpha,\beta,0,s) = Y_1\{\tilde{P}^i[1-R\exp(-2s\gamma_1 d)] + 2H^L \exp(-s\gamma_1 d)/[\rho_1(Y_1+Y_2)]\} \times [1+R\exp(-2s\gamma_1 d)]^{-1}, \quad \text{with } \alpha \in I \cup D^L. \tag{47}$$

We expand $\tilde{v}_{1,z}$ in the series

$$\tilde{v}_{1,z}(\alpha,\beta,0,s) = \sum_{n=0}^{\infty} \tilde{v}_{1,z}^{(n)}(\alpha,\beta,0,s), \tag{48}$$

in which $\tilde{v}_{1,z}^{(n)}$ has a particular exponential function of the type $\exp(-2ns\gamma_1 d)$. Next we rewrite (46) and (47) by using the expansions (30), (31), and (48) in such a way that we can capture the different wave constituents. This yields

$$[1-\exp(-2s\gamma_1 d)]\sum_{n=0}^{\infty} \tilde{v}_{1,z}^{(n)} = Y_1\left\{\tilde{P}^i[1+\exp(-2s\gamma_1 d)] - 2\sum_{n=0}^{\infty} G_n^R \exp(-s\gamma_1 d)\right\}, \quad \text{with } \alpha \in I \cup D^R \tag{49}$$

and

$$[1+R\exp(-2s\gamma_1 d)]\sum_{n=0}^{\infty} \tilde{v}_{1,z}^{(n)} = Y_1\left\{\tilde{P}^i[1-R\exp(-2s\gamma_1 d)] + 2\sum_{n=0}^{\infty} \frac{H_n^L \exp(-s\gamma_1 d)}{\rho_1(Y_1+Y_2)}\right\}, \quad \text{with } \alpha \in I \cup D^L, \tag{50}$$

respectively. From Eqs. (49) and (50) we can derive the following recurrence relations by capturing the wave constituents. The recurrence relation pertaining to $\tilde{v}_{1,z}^{(n)}$ in terms of G_n^R is given by

$$\tilde{v}_{1,z}^{(n)} = \tilde{v}_{1,z}^{(n-1)} \exp(-2s\gamma_1 d) - 2Y_1 G_{n-1}^R \exp(-s\gamma_1 d), \quad \text{with } n \in \{2,3,\ldots\} \text{ and } \alpha \in I \cup D^R, \tag{51}$$

while

$$\tilde{v}_{1,z}^{(0)} = Y_1 \tilde{P}^i, \quad \text{with } \alpha \in \mathbb{C}, \tag{52}$$

and

$$\tilde{v}_{1,z}^{(1)} = 2Y_1 \tilde{P}^i \exp(-2s\gamma_1 d) - 2Y_1 G_0^R \exp(-s\gamma_1 d), \quad \text{with } \alpha \in I \cup D^R, \tag{53}$$

in which [cf. (41) and (43)]

$$G_n^R = \Phi_{(I)}^R(\tilde{Q}_n)/K^R \gamma_1^R, \quad \text{when } \alpha \in I \cup D^R. \tag{54}$$

Alternatively, the recurrence relation pertaining to $\tilde{v}_{1,z}^{(n)}$ in terms of H_n^L is given by

$$\tilde{v}_{1,z}^{(n)} = -R\,\tilde{v}_{1,z}^{(n-1)} \exp(-2s\gamma_1 d) + 2Y_1 H_n^L \exp(-s\gamma_1 d)/[\rho_1(Y_1+Y_2)], \quad \text{with } \alpha \in I \cup D^L \text{ and } n \in \{2,3,\ldots\}, \tag{55}$$

while

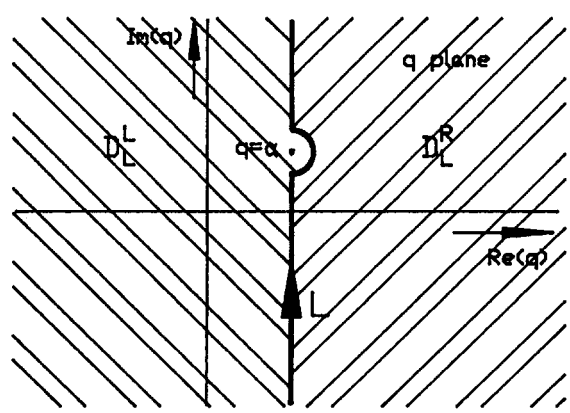


FIG. 4. The contour L pertaining to the operator $\Phi_{(L)}^L$ with $\alpha \in L$, which divides the complex q plane into the regions D_L^R and D_L^L.

$$\tilde{v}_{1,z}^{(0)} = Y_1 \tilde{P}^i, \quad \text{with } \alpha \in \mathbb{C} \tag{56}$$

and

$$\tilde{v}_{1,z}^{(1)} = -2Y_1 \tilde{P}^i R \exp(-2s\gamma_1 d) + 2Y_1 H_0^L \exp(-s\gamma_1 d)/[\rho_1(Y_1+Y_2)], \quad \text{with } \alpha \in I \cup D^L, \tag{57}$$

in which [cf. (42) and (43)]

$$H_n^L = (1+\rho_1/\rho_2)\gamma_1^L K^L \Phi_{(I)}^L(Q_n), \quad \text{when } \alpha \in I \cup D^L. \tag{58}$$

VII. THE SPACE-TIME DOMAIN ACOUSTIC WAVE FIELD

The particle velocity $\hat{v}_{1,z}(x,y,0,s)$ in the Laplace transform-domain can, with the aid of (48), be written as

$$\hat{v}_{1,z}(x,y,0,s) = \sum_{n=0}^{\infty} \hat{v}_{1,z}^{(n)}(x,y,0,s), \tag{59}$$

in which

$$\hat{v}_{1,z}^{(n)}(x,y,0,s) = \left(\frac{s}{2\pi i}\right)^2 \int_{-i\infty}^{i\infty} \exp(-s\beta y)d\beta \int_{-i\infty}^{i\infty} \exp(-s\alpha x) \times \tilde{v}_{1,z}^{(n)}(\alpha,\beta,0,s)d\alpha. \tag{60}$$

To transform the acoustic wave field back to the space-time domain we apply the modified Cagniard–de Hoop method: This method can only be applied successfully if we know the s dependence of the integrand explicitly. The representation of $\hat{v}_{1,z}(x,y,0,s)$ consists of a series, the terms of which are multidimensional integrals of the second or higher order. The integrand in each of the terms contains the s dependence only through the exponential functions $\exp(-sm\gamma_1 d)$, with $m \in N$. This invites the application of the modified Cagniard method to each integral in each term separately, each with its own "time variable."

A. Analytic continuation

It is anticipated that difficulties will arise during the deformations in the path of integration that occur in the modified Cagniard technique with respect to the wave constituents (51), (53) and (55), (57). This is due to the fact that the wave constituents contain the decomposition integral operator $\Phi_{(I)}^{R,L}$ [cf. (44) and (45)]. From Eqs. (51)–(54), (40), (24), and (B1) we obtain the following recurrence relations:

$$\tilde{v}_{1,z}^{(1)} = 2Y_1 \tilde{P}^i \exp(-2s\gamma_1 d) - 2Y_1 \exp(-s\gamma_1 d)\Phi_{(I)}^R(\tilde{Q}_0)/K^R\gamma_1^R,$$
$$\text{with } \tilde{Q}_0 = T\hat{P}(s)\exp[s(\alpha x_s + \beta y_s)]\exp(-s\gamma_1 d)K^R\gamma_1^R \text{ and } \alpha \in I \cup D^R, \tag{61}$$

$$\tilde{v}_{1,z}^{(n)} = \tilde{v}_{1,z}^{(n-1)} \exp(-2s\gamma_1 d) - 2Y_1 \exp(-s\gamma_1 d)\Phi_{(I)}^R(\tilde{Q}_{n-1})/K^R\gamma_1^R, \quad \text{with } \alpha \in I \cup D^R \text{ and } n \in \{2,3,\ldots\}, \tag{62}$$

and

$$\tilde{Q}_n = \tilde{Q}_{n-1} \exp(-2s\gamma_1 d) - T\exp(-2s\gamma_1 d)\Phi_{(I)}^R(\tilde{Q}_{n-1}), \quad \text{with } \alpha \in I \cup D^R \text{ and } n \in \{1,2,\ldots\}. \tag{63}$$

Alternatively, from (55)–(58), (40), (24), (B1), and (B3) we obtain the recurrence relations

$$\tilde{v}_{1,z}^{(1)} = -2Y_1 \tilde{P}^i R \exp(-2s\gamma_1 d) + 2Y_1 \exp(-s\gamma_1 d)\Phi_{(I)}^L(\tilde{Q}_0)K^L\gamma_1^L/K\gamma_1,$$
$$\text{with } \tilde{Q}_0 = T\exp(-s\gamma_1 d)\hat{P}(s)\exp[s(\alpha x_s + \beta y_s)]K\gamma_1/K^L\gamma_1^L \text{ and } \alpha \in I \cup D^L, \tag{64}$$

$$\tilde{v}_{1,z}^{(n)} = -R\tilde{v}_{1,z}^{(n-1)} \exp(-2s\gamma_1 d) + 2Y_1 \exp(-s\gamma_1 d)\Phi_{(I)}^L(Q_{n-1})K^L\gamma_1^L/K\gamma_1, \quad \text{with } \alpha \in I \cup D^L \text{ and } n \in \{2,3,\ldots\}, \tag{65}$$

and

$$\tilde{Q}_n = R\tilde{Q}_{n-1} \exp(-2s\gamma_1 d) + T\exp(-2s\gamma_1 d)\Phi_{(I)}^L(\tilde{Q}_{n-1}), \quad \text{with } \alpha \in I \cup D^L \text{ and } n \in \{1,2,\ldots\}. \tag{66}$$

Each time we apply the recurrence relation we have to deal with the integration contour pertaining to the decomposition operator. These contours will be denoted collectively by $\{C_i;\ i \in \mathbb{N}\}$. It is observed that in the expression for $\tilde{v}_{1,z}^{(n)}$ at most n contours occur. For the analytic continuation of $\tilde{v}_{1,z}^{(n)}$ and $\tilde{Q}_n$ we have to distinguish between the different locations that the source and receiver have with respect to the edge of the diffracting crack; these cases are (i) $x>0, x_s>0$; (ii) $x<0, x_s<0$; (iii) $x>0, x_s<0$; and (iv) $x<0, x_s>0$.

1. The case $x>0$, $x_s>0$

Since $x>0$, the Cagniard–de Hoop contours associated with the different terms in $\tilde{v}_{1,z}^{(n)}$ in the complex α plane are located in the part D^R to the right of the imaginary α axis; hence (61)–(63) are to be used. Further, since $x_s>0$, those Cagniard–de Hoop contours pertaining to the decomposition operators $\Phi_{(I)}^R$ that act on $\tilde{Q}_0$ are located in the part D^L to the left of the imaginary axis. The relevant decomposition operators transform into $\Phi_{(C_i)}^R$, in accordance with the definitions (44) and (45). The subsequent decomposition operators, those that do not act directly on $\tilde{Q}_0$, require no contour deformation in view of the fact that the corresponding Cagniard contour coincides with the imaginary axis. Carrying out these steps, the recurrence relations (61)–(63) result in an expression for $\tilde{v}_{1,z}^{(n)}$ with $n \in N$ in which no singularities are encountered during the deformation of the original contours into the Cagniard–de Hoop ones provided that the functions K^R and γ_1^R as they occur in $\tilde{Q}_0$ [cf. (61)] are continued analytically into D^L according to

$$\{K^R(q),\gamma_1^R(q)\} = \{K(q)/K^L(q),\gamma_1(q)/\gamma_1^L(q)\}, \quad \text{when } q \in I \cup D^L. \tag{67}$$

We also assume that K and γ_1 are analytic and free from zeros in the part of the complex q plane between the original and deformed contours.

2. The case $x<0$, $x_s<0$

Since $x<0$, the Cagniard–de Hoop contours associated with the different terms in $\tilde{v}_{1,z}^{(n)}$ in the complex α plane are located in the part D^L to the left of the imaginary α axis; hence (64)–(66) are to be used. Further, since $x_s<0$, those Cagniard–de Hoop contours pertaining to the decomposition operators $\Phi_{(I)}^L$ that act on $\tilde{Q}_0$ are located in the part D^R to the right of the imaginary axis. The relevant decomposition operators transform into $\Phi_{(C_i)}^L$, in accordance with the definitions (44) and (45). The subsequent decomposition operators, those that do not act directly on $\tilde{Q}_0$, require no contour deformation in view of the fact that the corresponding Cagniard contour coincides with the imaginary axis. Carrying out these steps, the recurrence relations (64)–(66) result in an expression for $\tilde{v}_{1,z}^{(n)}$ with $n\in N$ in which no singularities are encountered during the deformation of the original contours into the Cagniard–de Hoop ones provided that the functions K^L and γ_1^L as they occur in $\tilde{Q}_0$ [cf. (64)] are continued analytically into D^R according to

$$\{K^L(q),\gamma_1^L(q)\} = \{K(q)/K^R(q),\gamma_1(q)/\gamma_1^R(q)\}, \quad \text{when } q\in I\cup D^R. \tag{68}$$

We also assume that K and γ_1 are analytic and free from zeros in the part of the complex q plane between the original and deformed contours.

3. The case $x>0$, $x_s<0$

Since $x>0$, the Cagniard–de Hoop contours associated with the different terms in $\tilde{v}_{1,z}^{(n)}$ in the complex α plane are located in the part D^R to the right of the imaginary α axis; hence (61)–(63) are to be used. Further, since $x_s<0$, those Cagniard–de Hoop contours pertaining to the decomposition operators $\Phi_{(I)}^R$ that act on $\tilde{Q}_0$ are also located in the part D^R to the right of the imaginary axis. The relevant decomposition operators transform into $\Phi_{(C_i)}^R$, in accordance with the definitions (44) and (45). However, depending on the location of the different contours with respect to each other, we also need analytic continuation of $\Phi_{(C_i)}^R$ into the part $D_{C_i}^L$. The subsequent decomposition operators, those that do not act directly on $\tilde{Q}_0$, require no contour deformation in view of the fact that the corresponding Cagniard contour coincides with the imaginary axis.

For a systematic approach we start from the recurrence relations (61) and (63), from which we derive analytically continued expressions for $\tilde{v}_{1,z}^{(1)}$ and $\tilde{Q}_1$. The analytic continuations of $\tilde{v}_{1,z}^{(n)}$ and $\tilde{Q}_n$ with $n\in\{2,3,\ldots\}$ are derived recursively from the analytically continued expressions of $\tilde{v}_{1,z}^{(n-1)}$ and $\tilde{Q}_{n-1}$. In the procedure of recursive analytic continuation, only the decomposition operators $\Phi_{(C_i)}^{R,L}$ that act on $\tilde{Q}_0$ and map into the corresponding parts $D_{C_i}^{L,R}$ of the complex α plane require analytic continuation into those parts. The analytic continuations of those decomposition operators $\Phi_{(C_i)}^{R,L}$ that act on a product of the type $\tilde{Q}_0\exp(-is\gamma_1 d)\tilde{W}$, where $\tilde{W}=\tilde{W}(\alpha,\beta)$ is some analytic amplitude function, follow from Plemelj's theorem as

$$\Phi_{(C_i)}^{R,L}[\tilde{Q}_0\exp(-is\gamma_1 d)\tilde{W}] = \begin{cases} \Phi_{(C_i)}^{R,L}[\tilde{Q}_0\exp(-is\gamma_1 d)\tilde{W}], & \text{when } \alpha\in D_{C_i}^{R,L}\cup C_i, \\ \tilde{Q}_0\exp(-is\gamma_1 d)\tilde{W} - \Phi_{(C_i)}^{L,R}[\tilde{Q}_0\exp(-is\gamma_1 d)\tilde{W}], & \text{when } \alpha\in D_{C_i}^{L,R}\cup C_i. \end{cases} \tag{69}$$

The regions $D_{C_i}^{R,L}$ and one of the deformed contours C_i are shown in Fig. 5.

4. The case $x<0$, $x_s>0$

Since $x<0$, the Cagniard–de Hoop contours associated with the different terms in $\tilde{v}_{1,z}^{(n)}$ in the complex α plane are located in the part D^L to the left of the imaginary α axis; hence (64)–(66) are to be used. Further, since $x_s>0$, those Cagniard–de Hoop contours pertaining to the decomposition operators $\Phi_{(I)}^L$ that act on $\tilde{Q}_0$ are also located in the part D^L to the left of the imaginary axis. The relevant decomposition operators transform into $\Phi_{(C_i)}^L$, in accordance with the definitions (44) and (45). However, depending on the location of the different contours with respect to each other, we also need the analytic continuation of $\Phi_{(C_i)}^L$ into the part $D_{C_i}^R$. The subsequent decomposition operators, those that do not act directly on $\tilde{Q}_0$, require no contour deformation in view of the fact that the corresponding Cagniard contour coincides with the imaginary axis.

For a systematic approach we start from the recurrence relations (64) and (66), from which we derive analytically continued expressions for $\tilde{v}_{1,z}^{(1)}$ and $\tilde{Q}_1$. The analytic continuations of $\tilde{v}_{1,z}^{(n)}$ and $\tilde{Q}_n$ with $n\in\{2,3,\ldots\}$ are derived recursively from the analytically continued expressions of $\tilde{v}_{1,z}^{(n-1)}$ and $\tilde{Q}_{n-1}$. In the procedure of recursive analytic continuation, only the decomposition operators $\Phi_{(C_i)}^{R,L}$ that act on $\tilde{Q}_0$ and map into the corresponding parts $D_{C_i}^{L,R}$ of the complex α plane require analytic continuation into those parts. The analytic continuations of those decomposition operators $\Phi_{(C_i)}^{R,L}$ that act on a product of the type $\tilde{Q}_0\exp(-is\gamma_1 d)\tilde{W}$, where $\tilde{W}=\tilde{W}(\alpha,\beta)$ is some analytic amplitude function, follow from Plemelj's theorem [cf. (69)].

B. Time inversion by the modified Cagniard–de Hoop method

In the analytically continued integral expression of the particle velocity in the space–Laplace domain we can decompose the field into a *geometric* part that contains twofold integral terms and a *diffracted* part that contains higher-order multiple-integral terms. The expression of the wave constituent $\hat{v}_{1,z}^{(n)}(x,y,0,s)$ [cf. (60)] is obtained from the suitable recurrence relations (61)–(66) using the recursive procedure of analytic continuation outlined in Sec. VII A. A typical term $\hat{U}^g$ in the geometric part of $\tilde{v}_{1,z}^{(n)}$ is of the shape

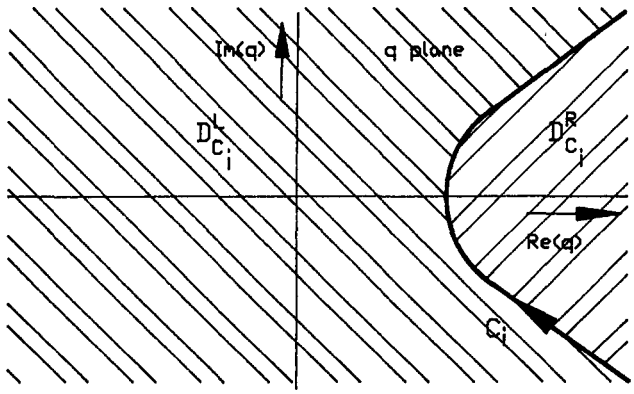


FIG. 5. A typical Cagniard–de Hoop contour pertaining to the decomposition operators $\Phi^{R,L}_{(C_i)}$ in the complex q plane.

$$\hat{U}^g(x,y,0,s) = s^2\hat{P}(s)\int_{-i\infty}^{i\infty} \exp[-s\beta(y-y_s)]\,d\beta \times \int_{-i\infty}^{i\infty} \tilde{A}^g(\alpha,\beta)\exp[-s\alpha(x-x_s)] \times \exp[-ns\gamma_1(\alpha)d]\,d\alpha\,, \tag{70}$$

in which $\tilde{A}^g(\alpha,\beta)$ represents a spectral amplitude that is independent of s. In view of the fact that the time inversion by the modified Cagniard–de Hoop method of the geometric part of $\hat{v}^{(n)}_{1,z}$ runs along the same lines as the time inversion of the diffracted part of $\tilde{v}^{(n)}_{1,z}$, we only consider the time inversion of the diffracted part of $\hat{v}^{(n)}_{1,z}$. A typical term $\hat{U}^d$ in the diffracted part of $\hat{v}^{(n)}_{1,z}$ in which the decomposition operators are written out in their integral form [cf. (44) and (45)] is given by

$$\hat{U}^d(x,y,0,s) = s^2\hat{P}(s)\int_{-i\infty}^{i\infty} \exp[-s\beta(y-y_s)]\,d\beta \int_{-i\infty}^{i\infty} \oint_{I_M}\oint_{I_{M-1}} \cdots \oint_{I_1}\oint_{I_0} \tilde{A}^d(\beta,q_0,q_1,\ldots,q_M,\alpha)\exp[-s(\alpha x - q_0 x_s)] \times \exp\left\{-s\left[\gamma_1(\alpha)k_\alpha d + \sum_{p=0}^{M}\gamma_1(q_p)k_{q_p}d\right]\right\} dq_0\,dq_1\cdots dq_M\,d\alpha, \quad \text{with } M\in\{0,1,2,\ldots,n-1\}\,, \tag{71}$$

in which

$$k_{q_1} = k_{q_2} = \cdots = k_{q_M} = 2 \text{ and } k_\alpha + k_{q_0} = 2n - 2M, \quad \text{with } k_\alpha, k_{q_0}\in N \text{ and } k_\alpha, k_{q_0} \text{ odd}\,. \tag{72}$$

In (71), $\tilde{A}^d(\beta, q_0, q_1,\ldots,q_M, \alpha)$ represents a spectral amplitude that is independent of s. Application of the modified Cagniard–de Hoop method (Du Cloux[4]) to each integral in (71) separately results in the following paths of integration:

$$C_\theta:\ \begin{cases}\mathrm{Re}[X_\theta\theta + \gamma_1(\theta)k_\theta d] = \tau_\theta\,,\\ \mathrm{Im}[X_\theta\theta + \gamma_1(\theta)k_\theta d] = 0\,,\end{cases} \tag{73}$$

where θ represents one of the variables of integration $\{\alpha,q_0,q_1,\ldots,q_M\}$ and the parameter X_θ is defined in Table I. The parametric expression of $\theta = \theta^d(\tau_\theta)$ is then given by

$$\theta^d(\tau_\theta) = \{X_\theta\tau_\theta + ik_\theta d\,[\tau_\theta^2 - (t^b_{d,\theta})^2]^{1/2}\} \times (X_\theta^2 + k_\theta^2 d^2)^{-1}, \quad \text{with } t^b_{d,\theta} \leqslant \tau_\theta < \infty\,, \tag{74}$$

where

$$t^b_{d,\theta} = \Omega_1(\beta)(X_\theta^2 + k_\theta^2 d^2)^{1/2} \tag{75}$$

is the value of τ_θ, where C_θ intersects the real θ axis. The contour pertaining to the parametric expression $\theta^d(\tau_\theta)$ [cf. (74)] is a branch of a hyperbola. Introduction of the normalized time variable ω_θ, defined by

TABLE I. Representation of X_θ.

θ	X_θ
α	x
q_0	$-x_s$
$q_1,q_2,\ldots,q_n$	0

$$\omega_\theta = \tau_\theta / t^b_{d,\theta} \tag{76}$$

as the variable of integration, changes expression (71) into

$$\hat{U}^d(x,y,0,s) = s^2\hat{P}(s)\int_{-i\infty}^{i\infty}\int_1^\infty\int_1^\infty\oint_1^\infty\cdots\oint_1^\infty\int_1^\infty \times H^d(\beta,\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha) \times \exp[-s\beta(y-y_s) - s\Omega_1(\beta)] \times Q(\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha)]\,d\omega_{q_0}\,d\omega_{q_1} \times \cdots d\omega_{q_M}\,d\omega_\alpha\,d\beta\,, \tag{77}$$

in which, using the identities (75) and (76),

$$Q(\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha) = \frac{1}{\Omega_1}\left[\sum_{p=0}^{M}\tau_{q_p} + \tau_\alpha\right] = \sum_{p=0}^{M}(X_{q_p}^2 + k_{q_p}^2 d^2)^{1/2}\omega_{q_p} + (X_\alpha^2 + k_\alpha^2 d^2)^{1/2}\omega_\alpha\,. \tag{78}$$

The amplitude function $H^d(\beta,\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha)$ is the result of the application of Schwarz's reflection principle to the product $J\tilde{A}^d$, where J denotes the Jacobian that occurs in the change of the integration variables. Its expression is determined recursively. By defining $\tilde{H}^d_M = \tilde{H}^d$ the simplest scheme starts in the α plane with

$$\tilde{H}^d_0(\beta,q_0^d,q_1^d,\ldots,q_M^d,\omega_\alpha) = \tilde{A}^d(\beta,q_0^d,q_1^d,\ldots,q_M^d,\alpha^d)\partial_{\omega_\alpha}\alpha^d - \tilde{A}^d(\beta,q_0^d,q_1^d,\ldots,q_M^d,\alpha^{d^*})\partial_{\omega_\alpha}\alpha^{d^*}\,. \tag{79}$$

The successive substitutions lead to the recursive relations

$$\tilde{H}^d_k(\beta,\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_k},q^d_{k+1},\ldots,q^d_M,\omega_\alpha)$$
$$= \tilde{H}^d_{k-1}(\beta,\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_{k-1}},q^d_k,q^d_{k+1},\ldots,q^d_M,\omega_\alpha)\partial_{\omega_{q_k}}q^d_k$$
$$-\tilde{H}^d_{k-1}(\beta,\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_{k-1}},q^{d*}_k,q^d_{k+1},\ldots,q^d_M,\omega_\alpha)$$
$$\times\partial_{\omega_{q_k}}q^{d*}_k, \quad \text{with } k\in\{1,2,3,\ldots,M\}, \tag{80}$$

in which the asterisk denotes the complex conjugate. Finally, the path of integration in the β plane is deformed into a Cagniard–de Hoop contour defined through

$$C_\beta: \begin{cases}\mathrm{Re}[(y-y_s)\beta+\Omega_1(\beta)Q]=\tau_\beta,\\ \mathrm{Im}[(y-y_s)\beta+\Omega_1(\beta)Q]=0,\end{cases} \tag{81}$$

where Q is given by (78). The parametric expression of $\beta=\beta^d(\tau_\beta)$ is

$$\beta^d(\tau_\beta)=\{(y-y_s)\tau_\beta+iQ[\tau_\beta^2-(t^b_{d,\beta})^2]^{1/2}\}$$
$$\times[(y-y_s)^2+Q^2]^{-1}, \quad \text{with } t^b_{d,\beta}\leqslant\tau_\beta<\infty, \tag{82}$$

where

$$t^b_{d,\beta}=t^b_{d,\beta}(\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha)$$
$$=(1/c_1)[(y-y_s)^2+Q^2(\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha)]^{1/2}. \tag{83}$$

The contour pertaining to the parametric expression of $\beta^d(\tau_\beta)$ [cf. (82)] is a branch of a hyperbola. Performing the β-contour deformation and rearranging the orders of integration in (77) yields, on account of Lerch's theorem, the space-time domain expression

$$U^d(x,y,0,t)=\partial_t^2\int_{t^b_{d,\beta}(1,1,\ldots,1,1)}^{t}p(t-\tau_\beta)g^d(x,y,0,\tau_\beta)d\tau_\beta, \tag{84}$$

in which

$$g^d(x,y,0,t)=2\int_1^{\omega^e_\alpha}\int_1^{\omega^e_M}\not\int_1^{\omega^e_{M-1}}\cdots\not\int_1^{\omega^e_1}\int_1^{\omega^e_0}\frac{\mathrm{Re}[\tilde{H}^d(\beta^d,\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha)\Omega_1(\beta^d)]}{[t^2-t^2_{d,\beta}(\omega_{q_0},\omega_{q_1},\ldots,\omega_{q_M},\omega_\alpha)]^{1/2}}$$
$$\times d\omega_{q_0}\,d\omega_{q_1}\cdots d\omega_{q_M}\,d\omega_\alpha\epsilon[t-t^b_{d,\beta}(1,1,\ldots,1,1)], \tag{85}$$

where $\epsilon[t]$ denotes the unit step function. In (85) the upper limits ω^e_i with $i\in\{0,1,\ldots,M\}$ are found by solving

$$t=t^b_{d,\beta}(1,\ldots,1,\omega^e_i,\omega_{q_{i+1}},\ldots,\omega_{q_M},\omega_\alpha), \quad \text{with } i\in\{0,1,\ldots,M\}, \tag{86}$$

and ω^e_α is found as the solution of

$$t=t^b_{d,\beta}(1,\ldots,1,\omega^e_\alpha). \tag{87}$$

If $c_1>c_2$ all contour deformations are admissible and joining arcs at infinity do not contribute.

C. Head waves

If $c_1<c_2$ the modified contours are tempted to intersect the real (α,q_0) axes at the branch cut associated with the vertical slowness γ_2 that is present in the spectral amplitude $\tilde{H}^d$. Whether or not this happens depends on the location of the points of observation and excitation; it only occurs in a finite region of β values. For points of observation in the region

$$1<x/(x^2+k^2_\alpha d^2)^{1/2}<c_1/c_2, \tag{88}$$

the integral contours with respect to α in the expressions for the geometric field and the diffracted field are tempted to intersect the real α axis in the intervals

$$\Omega_2(\beta)<\mathrm{Re}\{\alpha\}<\infty, \quad \text{when } x>0 \tag{89}$$

and

$$-\infty<\mathrm{Re}\{\alpha\}<-\Omega_2(\beta), \quad \text{when } x<0. \tag{90}$$

For points of excitation in the region

$$1<x_s/(x_s^2+k^2_{q_0}d^2)^{1/2}<c_1/c_2, \tag{91}$$

the integral contours with respect to q_0 in the expression for the diffracted field are tempted to intersect the real q_0 axis in the intervals

$$\Omega_2(\beta)<\mathrm{Re}\{q_0\}<\infty, \quad \text{when } x_s<0 \tag{92}$$

and

$$-\infty<\mathrm{Re}\{q_0\}<-\Omega_2(\beta), \quad \text{when } x_s>0. \tag{93}$$

The inequalities (89), (90) and (92), (93) are only satisfied if β satisfies the inequality

$$|\beta|<(\cos^2\phi/c_1^2-1/c_2^2)^{1/2}/\sin\phi, \tag{94}$$

where

$$\phi=\arccos[X_\theta/(X^2_\theta+k^2_\theta d^2)^{1/2}], \tag{95}$$

in which θ represents one of the variables $\{\alpha, q_0\}$. When the Cagniard–de Hoop contours are tempted to intersect the real axis in the regions indicated in (89), (90) and (92), (93), the path of integration has to be supplemented by a loop around the branch cut in the relevant complex plane (see Fig. 6). The parametric expression for the loop in the complex (α, q_0) plane is found from (73) and is

$$\theta^d_H(\tau_\theta)=\{X_\theta\tau_\theta-k_\theta d[(t^b_{d,\theta})^2-\tau_\theta^2]^{1/2}\}$$
$$\times(X^2_\theta+k^2_\theta d^2)^{-1}+i0,$$
$$\text{when } t^H_{d,\theta}\leqslant\tau_\theta<t^b_{d,\theta}, \tag{96}$$

where $t^H_{d,\theta}$ follows by substituting $\theta=0$ in (73):

$$t^H_{d,\theta}=k_\theta\,d\Omega_1(\beta). \tag{97}$$

In (96) and (97), θ represents one of the variables $\{\alpha, q_0\}$ and further $X_\alpha=x$, $X_{q_0}=-x_s$. The supplementary loops in the complex α and q_0 planes result in additional contributions to the wave functions; they represent the head wave

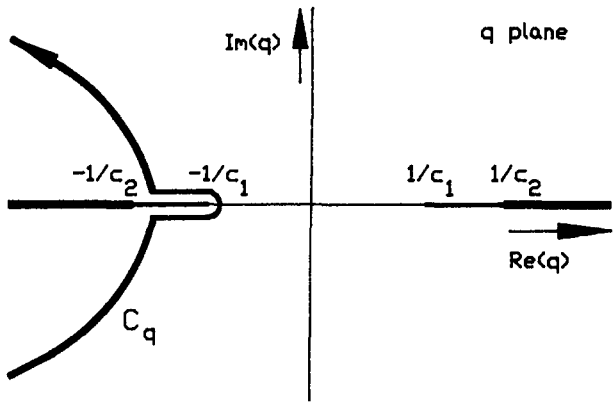


FIG. 6. A typical Cagniard–de Hoop contour with a supplementary loop integral around the branch cut of the branch point $-1/c_2$.

contributions. The head wave contributions can be separated into the terms

$$\hat{U}^{d,H}(x,y,0,s) = \hat{U}^{d,HH}(x,y,0,s) + \hat{U}^{d,BH}(x,y,0,s) + \hat{U}^{d,HB}(x,y,0,s) . \quad (98)$$

Here, the head wave contribution $\hat{U}^{d,HH}(x,y,0,s)$ arises from the coinciding paths of integration in the supplementary loop integrals that are present simultaneously in the complex α and q_0 planes. The head wave contribution $\hat{U}^{d,BH}(x,y,0,s)$ arises from the path of integration in the supplementary loop integral with respect to q_0 that does not coincide with the loop in the α plane. The head wave contribution $\hat{U}^{d,HB}(x,y,0,s)$ arises from the path of integration in the supplementary loop integral with respect to α that does not coincide with the loop in the q_0 plane. In the same way as we arrived at time-domain expressions for the body waves, we can find time-domain expressions for the head wave contributions. In the latter, the following normalizing substitutions for the supplementary loop integrals are used:

$$\omega_\theta = (\tau_\theta - t^H_{d,\theta})/(t_{d,\theta} - t^H_{d,\theta}), \quad \text{with } 0 \leqslant \omega_\theta \leqslant 1 , \quad (99)$$

in which θ represents one of the variables $\{\alpha, q_0\}$.

D. Time windowing

It is verified easily with the aid of the Cagniard–de Hoop method that each higher-order wave constituent arrives later than its previous one. Hence the analytically continued field expression obtained after n iteration steps in the recurrence relations (61)–(63) or (64)–(66) yields an expression in the time-domain that is exact within some finite time window. The upper limit of this window is determined by the minimal arrival time of the $(n+1)$st wave constituent.

VIII. CONCLUSION

In this paper, an analytic scheme is developed for the three-dimensional diffraction of acoustic waves by a semi-infinite crack in the interface of a uniform layer and a half-space. Both media consist of a material in which the shear stresses are neglected. The iterative scheme presented in this paper yields an exact solution for the time-domain particle velocity anywhere at the surface of the layer within a certain time window. The solution obtained from the scheme consists of a summation of multiple integrals which can be computed numerically within any degree of accuracy.

ACKNOWLEDGMENT

The authors wish to thank Professor A. T. de Hoop, Laboratory of Electromagnetic Research, Delft University of Technology, The Netherlands, for his helpful and stimulating discussions concerning the research presented in this paper.

APPENDIX A: STANDARD WIENER–HOPF TECHNIQUE

The solution of the Wiener–Hopf equation (36) is found by rewriting (36) in the form

$$[\tilde{U}^i_n - G^R_n]K = \rho_2 H^L_n/[(\rho_1+\rho_2)\gamma_1] , \quad (A1)$$

in which K is given by (37). Factorization of K and γ_1 leads to

$$\tilde{U}^i_n K^R \gamma^R_1 - G^R_n K^R \gamma^R_1 = \rho_2 H^L_n/[(\rho_1+\rho_2)\gamma^L_1 K^L] , \quad \text{when } \alpha \in I . \quad (A2)$$

We denote the first term on the lhs of (A2) by $\tilde{Q}_n$. Application of the additive decomposition to the function $\tilde{Q}_n$, i.e., $\tilde{Q}_n = \tilde{Q}^R_n + \tilde{Q}^L_n$ on I [cf. (43)–(45)] yields

$$\tilde{Q}^R_n - G^R_n K^R \gamma^R_1 = \rho_2 H^L_n/[(\rho_1+\rho_2)\gamma^L_1 K^L] - \tilde{Q}^L_n , \quad \text{when } \alpha \in I . \quad (A3)$$

The lhs of (A3) is regular in D^R and is at most $o(\alpha^{1/2})$ as $|\alpha| \to \infty$ in D^R and the rhs is regular in D^L and is at most $o(\alpha^{-1/2})$ as $|\alpha| \to \infty$ in D^L. According to Liouville's lemma, the lhs and the rhs are then identically zero in their domains of regularity. Hence

$$G^R_n = \tilde{Q}^R_n/(K^R \gamma^R_1), \quad \text{with } \alpha \in I \cup D^R \quad (A4)$$

and

$$H^L_n = (1+\rho_1/\rho_2)\gamma^L_1 K^L \tilde{Q}^L_n, \quad \text{with } \alpha \in I \cup D^L . \quad (A5)$$

APPENDIX B: PROOF OF THE CONVERGENCE OF THE RECURRENCE SCHEME

In order to prove the convergence of the recurrence scheme presented in Sec. V, we have to prove the convergence of the series expansions (30) and (31). In this Appendix we shall prove this convergence. First, we derive a recurrence relation for $\tilde{Q}_n$ using (36), with n replaced by $n-1$, and (34), (40), and (54):

$$\tilde{Q}_n = \tilde{Q}_{n-1} \exp(-2s\gamma_1 d) - T \exp(-2s\gamma_1 d)\Phi^R_{(I)}(\tilde{Q}_{n-1}) . \quad (B1)$$

The operator $\Phi^{R,L}_{(I)}$ is defined by (44) and (45). For $\alpha \in I$ we can write $\Phi^{R,L}_{(I)}$ [cf. (45)] as

$$\Phi^R_{(I)}(f(\alpha)) = \tfrac{1}{2}(I - iH)(f(\alpha)), \quad \text{when } \alpha \in I , \quad (B2)$$

and

$$\Phi^L_{(I)}(f(\alpha)) = \tfrac{1}{2}(I + iH)(f(\alpha)) = (I - \Phi^R_{(I)})(f(\alpha)), \quad \text{when } \alpha \in I , \quad (B3)$$

in which I denotes the *unit* operator and H the *Hilbert* opera-

tor on I. Since for the norm of these operators we have $\|I\| = 1$ and $\|H\| = 1$ (Hochstadt,[13] pp. 21 and 183), the norm of $\Phi_{(I)}^{R,L}(f(\alpha))$ with $\alpha \in I$ satisfies the following inequality:

$$\|\Phi_{(I)}^{R,L}(f(\alpha))\| \leqslant \|f(\alpha)\|, \quad \text{when } \alpha \in I. \tag{B4}$$

The function $\tilde{Q}_n$ in (B1) can be rewritten in terms of $\tilde{Q}_0$, using (B2), as

$$\tilde{Q}_n = \{\exp(-2s\gamma_1 d)[(1 - T/2)I + i(T/2)H]\}^{n-1}(\tilde{Q}_0), \quad \text{with } \alpha \in I. \tag{B5}$$

It is important to notice that the norm of the operator acting on $\tilde{Q}_0$ satisfies the inequality

$$\|\exp(-2s\gamma_1 d)[(1 - T/2)I + i(T/2)H]\| < 1 \tag{B6}$$

and hence this operator is a *contraction* operator (Hochstadt,[13] p. 26). Substitution of (B5) in (A4), using (B2), yields for G_n^R the following representation in terms of $\tilde{Q}_0$:

$$G_n^R = (2K^R\gamma_1^R)^{-1}(I - iH)\{\exp(-2s\gamma_1 d)[(1 - T/2)I + i(T/2)H]\}^{n-1}(\tilde{Q}_0), \quad \text{when } \alpha \in I. \tag{B7}$$

Similarly, substitution of (B5) in (A5), using (B3), yields for H_n^L the following representation in terms of $\tilde{Q}_0$:

$$H_n^L = [(1 + \rho_1/\rho_2)K^L\gamma_1^L/2](I + iH)\{\exp(-2s\gamma_1 d) \times [(1 - T/2)I + i(T/2)H]\}^{n-1}(\tilde{Q}_0), \quad \text{when } \alpha \in I. \tag{B8}$$

With the aid of (B7) and (B8) the series expansions (30) and (31), respectively, can be rewritten in terms of $\tilde{Q}_0$. Since the operators acting on $\tilde{Q}_0$ are contraction operators, the series expansions of (30) and (31) converge (Hochstadt[13]).

[1] A. Sommerfeld, "Theoretisches über die Beugung der Röntgenstrahlen," Z. Math. Phys. **46**, 11 (1901).
[2] L. Cagniard, *Reflexion et refraction des ondes seismiques progressives* (Gauthier-Villars, Paris, 1939), p. 55.
[3] A. T. de Hoop, Ph.D. thesis, Delft University of Technology, 1958.
[4] R. Du Cloux, Ph.D. thesis, Delft University of Technology, 1986.
[5] R. Du Cloux, "Pulsed electromagnetic radiation from a line source in the presence of a semi-infinite screen in the plane interface of two different media," Wave Motion **6**, 456 (1984).
[6] H. Shirai and L. B. Felsen, "Spectral method for multiple edge diffraction by a flat strip," Wave Motion **8**, 499 (1986).
[7] A. T. de Hoop, "A modification of Cagniard's method for solving seismic pulse problems," Appl. Sci. Res. B **8**, 349 (1960).
[8] A. T. de Hoop, "Pulsed electromagnetic radiation from a line source in a two-media configuration," Radio Sci. **14**, 253 (1979).
[9] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973), pp. 298–321 and 365–388.
[10] J. D. Achenbach, A. K. Gautesen, and H. McMaken, *Ray Methods for Waves in Elastic Solids*, Pitman Advanced Publishing Program No. 14 (Pitman, New York, 1982), pp. 109–148.
[11] L. A. Weinstein, *The Theory of Diffraction and the Factorization Method* (Golden, Boulder, CO, 1969), pp. 17–22.
[12] B. Van der Pol and H. Bremmer, *Operational Calculus Based on the Two-sided Laplace Integral* (Cambridge U.P., Princeton, NJ, 1950), pp. 7–21.
[13] H. Hochstadt, *Integral Equations, Pure & Applied Mathematics* (Wiley–Interscience, New York, 1973).