On chains of orthosymplectic Lie superalgebras and the *n*-dimensional quantum harmonic oscillator

J. Beckers
Theoretical and Mathematical Physics, Institute of Physics (B.5), University of Liège, Sart Tilman, B-4000 Liège 1, Belgium
J. F. Cornwell
Department of Physics and Astronomy, University of St. Andrews, St. Andrews, Fife KY16 9SS, Scotland

(Received 27 October 1988; accepted for publication 22 February 1989)

Triplets of embedded orthosymplectic Lie superalgebras are singled out and analyzed in terms of their respective even and odd root systems. The corresponding chains of embeddings are considered for arbitrary integers m and n (for both $m \neq n$ and $m = n$). It is shown that the second member inside each triplet is always in a one-to-one correspondence with the semidirect sum of the third member *and* the associated Heisenberg superalgebra that is characterized by the same entries m and n. For arbitrary $m = n$, invariance superalgebras of the n-dimensional harmonic oscillator are recovered and their one-to-one correspondence associated with a recent "character reversal" phenomenon linking these invariance superstructures is rigorously explained. Specific and general cases as well as conclusions are presented.

I. INTRODUCTION

During the 1970's supersymmetry[1,2] in elementary particle physics motivated the development of a theory of graded Lie algebras whose properties and characteristics are now well known,[3,4] and which are now usually referred to as Lie superalgebras. Subsequently, supersymmetric quantum mechanics was initiated by Witten[5] in the 1980's and developed by many authors,[6–9] with the harmonic oscillator in n arbitrary spatial dimensions being treated as one of its fundamental applications. The study of this application has led to the investigation of the *invariance* Lie superalgebras of the harmonic oscillator, which belong mainly to the orthosymplectic series $\mathrm{osp}(p/q)$ [or to $B(r/s)$, $C(s)$, and $D(r/s)$ in Kac's notation[3]]. Such *simple* Lie superalgebras[6,7] have also been extended to certain *nonsimple* Lie superalgebras, which are called the "maximal" or "*largest*" invariance superalgebras[10–12] for the quantum harmonic oscillator. Indeed, for an n-dimensional harmonic oscillator, de Crombrugghe and Rittenberg[7] first showed that $\mathrm{osp}(2n/2n)$ was the invariance superalgebra, while Beckers and Hussin[10] demonstrated the role of the semidirect sum $\mathrm{osp}(2n/2n) \oplus \mathrm{sh}(2n/2n)$. The inclusion

$$[\mathrm{osp}(2n/2n) \oplus \mathrm{sh}(2n/2n)] \supset \mathrm{osp}(2n/2n) \tag{1.1}$$

enabled these latter authors[10] to point out the major role played by the generators of the nonsimple Heisenberg superalgebra $\mathrm{sh}(2n/2n)$, which is generated by $2n + 2n(+1)$ operators [where the $(+1)$ refers to the identity generator whose role is rather different from the others]. Indeed, Beckers and Hussin have shown that all the orthosymplectic generators are defined in terms of bilinear products of the Heisenberg ones. This largest superalgebra $\mathrm{osp}(2n/2n) \oplus \mathrm{sh}(2n/2n)$ has the advantage that it contains[11,12] kinematical[13] as well as dynamical[14,15] (super) symmetries, so that it was recognized as the largest spectrum generating superalgebra of the n-dimensional harmonic oscillator taking into account its "conformal" [13,16] as well as its "superconformal" [9] properties.

Nearly simultaneously with the above contribution,[12] Englefield[17] proposed a new invariance superalgebra for the n-dimensional harmonic oscillator, namely the (simple) orthosymplectic superalgebra $\mathrm{osp}(3/2n)$, which he demonstrated appears as a generalization of the developments of Van der Jeugt[18] concerning the superalgebra $\mathrm{osp}(3/2)$. The astonishing feature that becomes apparent when one compares $\mathrm{osp}(3/2n)$ with $\mathrm{osp}(2n/2n) \oplus \mathrm{sh}(2n/2n)$, as well as with $\mathrm{osp}(2n/2n)$, is that $\mathrm{osp}(3/2n)$ has a specific and natural $\mathrm{so}(3)$ content that seems difficult to explain from the $\mathrm{so}(2n)$ content of the other superstructures *for arbitrary n*. (For example this is impossible for $n = 1$!). Nevertheless, Beckers, Dehin, and Hussin[19] have observed that the interchange of the *even* and *odd* characters of the nontrivial generators belonging to the Heisenberg superalgebra links in a "unique" way some of the above superalgebras and subalgebras.

Recalling that the dimension of the orthosymplectic superalgebra $\mathrm{osp}(m/n)$ is given by

$$d = \tfrac{1}{2}[(m+n)^2 + n - m], \tag{1.2}$$

we notice that, from a dimensional point of view, we have the *strict* inclusion (for $n \neq 1$)

$$[\mathrm{osp}(2n/2n) \oplus \mathrm{sh}(2n/2n)] \supset \mathrm{osp}(3/2n), \tag{1.3}$$

as well as the *possible* identification (for $n = 1$)

$$\mathrm{osp}(2/2) \oplus \mathrm{sh}(2/2) \leftrightarrow \mathrm{osp}(3/2). \tag{1.4}$$

With such starting elements, Beckers *et al.*[19] then studied the generalization of (1.4) to n arbitrary dimensions, i.e., to the correspondences

$$\mathrm{osp}(2/2n) \oplus \mathrm{sh}(2/2n) \leftrightarrow \mathrm{osp}(3/2n). \tag{1.5}$$

In order to be complete, it should be mentioned that a Heisenberg superalgebra $\mathrm{sh}(2m/2n)$ has the dimension

$$d = 2m + 2n(+1), \tag{1.6}$$

where again the unity $(+1)$ refers to the identity operator of the central extension included in all the Heisenberg alge-

 0022-2488/89/081655-07$02.50

bras, this extra $(+1)$ term being not included in the dimensions for the correspondences (1.4) and (1.5).

The "character reversal" phenomenon observed by Beckers *et al.*[19] has now been rigorously explained[20] in the simplest case, i.e., for the correspondence (1.4) associated with the invariance superalgebras of the one-dimensional quantum harmonic oscillator. This analysis can be extended to the case of arbitrary n, for which the associated correspondences are those of (1.5), so that certain remarkable new properties of superalgebras become apparent.

The purpose of the present paper is not merely to present these arguments and results, but to go appreciably further. We shall show (entirely within the framework of Lie superalgebra theory) that the character reversal phenomenon also works for the set of correspondences

$$[\mathrm{osp}(2n/2n) \ni\!\!\!\!+ \mathrm{sh}(2n,2n)] \leftrightarrow \mathrm{osp}(2n+1/2n) \qquad (1.7)$$

[which generalize (1.5)], *and also* for *all* the sets of correspondences

$$\mathrm{osp}(2m/2n) \ni\!\!\!\!+ \mathrm{sh}(2m,2n) \leftrightarrow \mathrm{osp}(2m+1/2n) \qquad (1.8)$$

where m and n are *arbitrary* positive integers. Clearly, (1.8) reduces to (1.5) in the case $m=1$, (1.8) reduces to (1.7) for $m=n$, and (1.7) reduces to (1.4) for the special case $m=n=1$. Consequently our intention here is to concentrate on the last set of correspondences (1.8) for *arbitrary* positive integers m and n. In investigating these we are led to remarkable chains of orthosymplectic Lie superalgebras that are not only of physical interest, but which also possess some distinctive and characteristic properties of their subalgebra contents that we believe to be new. In our study[20] of the correspondence (1.4) we have already pointed out the interesting chain:

$$\mathrm{osp}(3/4) \supset \mathrm{osp}(3/2) \supset \mathrm{osp}(2/2). \qquad (1.9)$$

In the general situation corresponding to (1.8) we will be interested in the following chains:

$$\mathrm{osp}(2m+1/2n+2) \supset \mathrm{osp}(2m+1/2n) \supset \mathrm{osp}(2m/2n). \qquad (1.10)$$

The contents of this paper are accordingly arranged as follows. In Sec. II we shall establish the notations and conventions we are using together with some of the basic properties of the orthosymplectic Lie superalgebras. The emphasis is put particularly on the systems of even and odd roots in terms of the simple ones. In Sec. III the chains (1.10) are characterized and two propositions are established by pointing out the particularly important role of the complement of $\mathrm{osp}(2m/2n)$ in $\mathrm{osp}(2m+1/2n)$. Section IV deals with the embedding of the Heisenberg superalgebra $\mathrm{sh}(2m/2n)$ in $\mathrm{osp}(2m+1/2n+2)$ as well as with the properties of such an embedding. Two more propositions are stated and are shown to be sufficient in order to give a complete proof of the character reversal phenomenon already mentioned for the invariance superalgebras of the harmonic oscillator. Finally Sec. V is devoted to some comments and conclusions.

As quantum physics is involved in all these applications of these superstructures, we shall choose our units in this present paper to be such that $\hbar=1$, and consider harmonic oscillators with mass $m=1$, but with angular frequency ω. These conventions will evidently have consequences in the structure relations displayed in the following sections.

II. BASIC PROPERTIES OF THE ORTHOSYMPLECTIC LIE SUPERALGEBRAS

Before demonstrating the various subalgebra embeddings mentioned in Sec. I, it is necessary to establish the notations and conventions that will be used.[21] Suppose that p and q are any two positive integers, and consider a $(p+q)\times(p+q)$ matrix $\mathbf{M}$ with complex entries and with the partitioning

$$\mathbf{M}=\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix},$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are submatrices with dimensions $p\times p$, $p\times q$, $q\times p$, and $q\times q$, respectively. As usual, if $\mathbf{B}=\mathbf{0}$ and $\mathbf{C}=\mathbf{0}$, then $\mathbf{M}$ is said to be even and to have degree 0, whereas if $\mathbf{A}=\mathbf{0}$ and $\mathbf{D}=\mathbf{0}$, then $\mathbf{M}$ is said to be odd and to have degree 1. The set of all complex linear combinations of these matrices form the complex associative superalgebra $M(p/q;\mathbb{C})$. Frequent use will be made of the $(p+q)\times(p+q)$ matrices $\mathbf{e}_{kl}$ that are defined by

$$(\mathbf{e}_{kl})_{ij}=\delta_{ki}\delta_{lj} \text{ (for } i,j,k,l=1,2,\ldots,p+q). \qquad (2.1)$$

The complex orthosymplectic Lie superalgebra $\mathrm{osp}(p/q)$ is obtained by considering matrices of the set $M(p/q;\mathbb{C})$ with $p\geqslant 1$ and with q positive and *even*. Let $\mathbf{K}$ be the member of $M(p/q;\mathbb{C})$ which is such that

$$\mathbf{K}=\begin{bmatrix}\mathbf{G}_p & \mathbf{0}\\ \mathbf{0} & \mathbf{J}_q\end{bmatrix}, \qquad (2.2)$$

where $\mathbf{G}_p=\mathbf{1}_p$ (the $p\times p$ unit matrix) and $\mathbf{J}_q$ is the $q\times q$ matrix

$$\mathbf{J}_q=\begin{bmatrix}\mathbf{0} & \mathbf{1}_{(1/2)q}\\ -\mathbf{1}_{(1/2)q} & \mathbf{0}\end{bmatrix}. \qquad (2.3)$$

Such a matrix $\mathbf{K}$ will play the role of a metric so that the subset of matrices $\mathbf{M}$ of $M(p/q;\mathbb{C})$ that satisfy the condition

$$\mathbf{M}^{\mathrm{st}}\mathbf{K}+(-1)^{\deg \mathbf{M}}\mathbf{K}\mathbf{M}=\mathbf{0} \qquad (2.4)$$

form the complex orthosymplectic Lie superalgebra $\mathrm{osp}(p/q)$. Here $\mathbf{M}^{\mathrm{st}}$ indicates the supertranspose of $\mathbf{M}$

$$\mathbf{M}=\begin{bmatrix}\mathbf{A}^t & -\mathbf{C}^t\\ \mathbf{B}^t & \mathbf{D}^t\end{bmatrix},$$

and $\deg \mathbf{M}$ denotes the degree of $\mathbf{M}$. The even elements of $\mathrm{osp}(p/q)$ are therefore of the form

$$\mathbf{M}=\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{bmatrix},$$

and so satisfy the constraint $\mathbf{A}^t\mathbf{G}+\mathbf{G}\mathbf{A}=\mathbf{0}$ (which, with $\mathbf{G}=\mathbf{1}_p$, reduces to $\mathbf{A}^t+\mathbf{A}=\mathbf{0}$), together with the constraint $\mathbf{D}^t\mathbf{J}+\mathbf{J}\mathbf{D}=\mathbf{0}$. Similarly the odd elements of $\mathrm{osp}(p/q)$ are of the form

$$\mathbf{M}=\begin{bmatrix}\mathbf{0} & \mathbf{B}\\ \mathbf{C} & \mathbf{0}\end{bmatrix},$$

and so satisfy the constraint $\mathbf{B}^t\mathbf{G}=\mathbf{J}\mathbf{C}$. The dimensions of the even and odd parts of $\mathrm{osp}(p/q)$ are $\frac{1}{2}p(p-1)+\frac{1}{2}q(q+1)$ and pq, respectively.

In the following analysis instead of taking $\mathbf{G}_p = \mathbf{1}_p$ it will be convenient to take for $\mathbf{G}_p$ either

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{0} & \mathbf{1}_r \\ \mathbf{1}_r & \mathbf{0} \end{bmatrix} \tag{2.5}$$

if $p(=2r)$ is even, or

$$\mathbf{G}_p = \begin{bmatrix} \mathbf{0} & \mathbf{1}_r & \mathbf{0} \\ \mathbf{1}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_1 \end{bmatrix} \tag{2.6}$$

if $p(=2r+1)$ is odd, which merely corresponds to applying to the matrices $\mathbf{M}$ of $M(p/q;\mathbb{C})$ a transformation of the type $\mathbf{M} \to \mathbf{SMS}'$ with an appropriate $(p+q)\times(p+q)$ matrix $\mathbf{S}$. The resulting matrices form a complex Lie superalgebra that is isomorphic to $\mathrm{osp}(p/q)$, and which will be denoted by the same set of symbols.

The Lie superalgebras $\mathrm{osp}(p/q)$ are all simple. In the notation of Kac[3] $\mathrm{osp}(2r+1/2s)$ is denoted by $B(r/s)$ (for $r\geqslant 0$ and $s\geqslant 1$), $\mathrm{osp}(2/2s-2)$ is denoted by $C(s)$ (for $s\geqslant 2$), and $\mathrm{osp}(2r/2s)$ is denoted by $D(r/s)$ (for $r\geqslant 2$ and $s\geqslant 1$).

The structure of $B(r/s)$ $[=\mathrm{osp}(2r+1/2s)]$ will now be described in some detail for the case in which $r\geqslant 1$ and $s\geqslant 1$, which is the situation of main interest in the applications of this paper. (Of course most of the properties that will be mentioned here, including the explicit expressions for the roots, have appeared previously in the papers listed in Ref. 3.) The dimensions of the even and odd parts of $\mathrm{osp}(2r+1/2s)$ are $r(2r+1)+s(2s+1)$ and $2s(2r+1)$, respectively. The even subalgebra of $\mathrm{osp}(2r+1/2s)$ is $C_s \oplus B_r$, and the rank of l of $\mathrm{osp}(2r+1/2s)$ is given by $l=r+s$. A convenient basis of the Cartan subalgebra $\mathscr{H}_s$ of $B(r/s)$ is provided by

$$\mathbf{h}_j^{\,1} = \mathbf{e}_{j+2r+1,j+2r+1} - \mathbf{e}_{j+2r+s+1,j+2r+s+1}, \quad j=1,2,\ldots,s \tag{2.7}$$

(which provides a basis for the Cartan subalgebra $\mathscr{H}_0^{\,1}$ of C_s), and

$$\mathbf{h}_j^{\,2} = \mathbf{e}_{j,j} - \mathbf{e}_{j+r,j+r}, \quad j=1,2,\ldots,r \tag{2.8}$$

(which provides a basis for the Cartan subalgebra $\mathscr{H}_0^{\,2}$ of B_r). Two useful sets of linear functionals $\epsilon_1^{\,1}, \epsilon_2^{\,1},\ldots,\epsilon_s^{\,1}$ and $\epsilon_1^{\,2}, \epsilon_2^{\,2},\ldots,\epsilon_r^{\,2}$ on $\mathscr{H}_s$ are defined by

$$\epsilon_p^{\,1}(\mathbf{h}) = \begin{cases} \delta_{jp}, & \text{if } \mathbf{h}=\mathbf{h}_j^{\,1}, \quad \text{for } j=1,2,\ldots,s, \\ 0, & \text{if } \mathbf{h}\in\mathscr{H}_0^{\,2}, \end{cases} \tag{2.9}$$

for $p=1,2,\ldots,s$, and

$$\epsilon_p^{\,2}(\mathbf{h}) = \begin{cases} \delta_{jp}, & \text{if } \mathbf{h}=\mathbf{h}_j^{\,2}, \quad \text{for } j=1,2,\ldots,r, \\ 0, & \text{if } \mathbf{h}\in\mathscr{H}_0^{\,1}, \end{cases} \tag{2.10}$$

for $p=1,2,\ldots,r$. As the Killing form $B(\ ,\ )$ of $B(r/s)$ is nondegenerate, for each linear functional α on $\mathscr{H}_s$ there exists a unique element h_α of $\mathscr{H}_s$ that is defined by $B(h_\alpha,h)=\alpha(h)$ for all h of $\mathscr{H}_s$. For the above linear functionals this implies that

$$h_\alpha = -[1/(2r-2s-1)]h_p^{\,1}, \text{ for } \alpha=\epsilon_p^{\,1} \quad (p=1,2,\ldots,s), \tag{2.11a}$$

and

$$h_\alpha = [1/(2r-2s-1)]h_p^{\,2} \text{ for } \alpha=\epsilon_p^{\,2} \ (p=1,2,\ldots,r). \tag{2.11b}$$

The superalgebra $B(r/s)$ has $2(r^2+s^2)$ *even* roots, which fall into eight classes. In the list of those that follows a realization of a basis element e_α of the corresponding one-dimensional root subspace $\mathscr{L}_{s\alpha}$ is provided for each root α:

$$\begin{aligned}
&(1)\ \alpha=\epsilon_p^{\,1}-\epsilon_q^{\,1} \text{ for } p,q=1,2,\ldots,s\ (p\neq q),\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{2r+1+p,2r+1+q} - \mathbf{e}_{2r+1+s+q,2r+1+s+p};\\
&(2)\ \alpha=\epsilon_p^{\,1}+\epsilon_q^{\,1} \text{ for } p,q=1,2,\ldots,s\ (p\leqslant q),\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{2r+1+p,2r+1+s+q} + \mathbf{e}_{2r+1+q,2r+1+s+p};\\
&(3)\ \alpha=-(\epsilon_p^{\,1}+\epsilon_q^{\,1}) \text{ for } p,q=1,2,\ldots,s\ (p\leqslant q),\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{2r+1+s+p,2r+1+q} + \mathbf{e}_{2r+1+s+q,2r+1+p};\\
&(4)\ \alpha=\epsilon_p^{\,2}-\epsilon_q^{\,2} \text{ for } p,q=1,2,\ldots,r\ (p\neq q)\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p,q} - \mathbf{e}_{q+r,p+r};\\
&(5)\ \alpha=\epsilon_p^{\,2}+\epsilon_q^{\,2} \text{ for } p,q=1,2,\ldots,r\ (p<q),\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p,q+r} - \mathbf{e}_{q,p+r};\\
&(6)\ \alpha=-(\epsilon_p^{\,2}+\epsilon_q^{\,2}) \text{ for } p,q=1,2,\ldots,r\ (p<q),\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p+r,q} - \mathbf{e}_{q+r,p};\\
&(7)\ \alpha=\epsilon_p^{\,2} \text{ for } p=1,2,\ldots,r,\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p,2r+1} - \mathbf{e}_{2r+1,p+r};\\
&(8)\ \alpha=-\epsilon_p^{\,2} \text{ for } p=1,2,\ldots,r,\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p+r,2r+1} - \mathbf{e}_{2r+1,p}.
\end{aligned} \tag{2.12}$$

The first three classes are extensions of roots of C_s, and the last five are extensions of roots of B_r. If $s=1$ the set (1) is empty, as are the sets (4), (5), and (6) if $r=1$.

$B(r/s)$ also has $2s(2r+1)$ *odd* roots, which fall into six classes. In the list of those that follows a realization of a basis element e_α of the corresponding 1-dimensional root subspace $\mathscr{L}_{s\alpha}$ is again provided for each root α:

$$\begin{aligned}
&(1)\ \alpha=\epsilon_q^{\,1} \text{ for } q=1,2,\ldots,s,\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{2r+1,2r+1+s+q} - \mathbf{e}_{2r+1+q,2r+1};\\
&(2)\ \alpha=-\epsilon_q^{\,1} \text{ for } q=1,2,\ldots,s,\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{2r+1,2r+1+q} + \mathbf{e}_{2r+1+s+q,2r+1};\\
&(3)\ \alpha=\epsilon_p^{\,1}+\epsilon_q^{\,2} \text{ for } p=1,2,\ldots,r, \text{ and } q=1,2,\ldots,s,\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p,2r+1+s+q} - \mathbf{e}_{2r+1+q,r+p};\\
&(4)\ \alpha=-(\epsilon_p^{\,1}+\epsilon_q^{\,2}) \text{ for } p=1,2,\ldots,r,\\
&\qquad \text{and } q=1,2,\ldots,s,\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p+r,2r+1+q} + \mathbf{e}_{2r+1+s+q,p};\\
&(5)\ \alpha=\epsilon_p^{\,1}-\epsilon_q^{\,2} \text{ for } p=1,2,\ldots,r, \text{ and } q=1,2,\ldots,s.\\
&\qquad \text{for which } e_\alpha=\mathbf{e}_{p+r,2r+1+s+q} - \mathbf{e}_{2r+1+q,p};
\end{aligned} \tag{2.13}$$

(6) $\alpha = -(\epsilon_p{}^1 - \epsilon_q{}^2)$ for $p = 1,2,...,r$,

and $q = 1,2,...,s$,

for which $\mathbf{e}_\alpha = \mathbf{e}_{p,2r+1+q} + \mathbf{e}_{2r+1+q,r+p}$.

In the general analysis of the Lie superalgebra chains that follows the roots will be used in the form quoted above. However, to make contact with the special case in which $m = 1$ and $n = 1$ that has already been considered[20] [where the roots of $B(1/2)$ were expressed in terms of linear combinations of the simple roots] and also for completeness, it will be indicated how they can be expressed in terms of the simple roots. The distinguished set of simple roots of $B(r/s)$ may be taken to be the following.

(1) Simple even roots:

$$\alpha_j = \epsilon_j{}^1 - \epsilon_{j+1}{}^1, \text{ for } j = 1,2,...,s-1 \quad (\text{if } s > 1),$$
$$\alpha_{j+s} = \epsilon_j{}^2 - \epsilon_{j+1}{}^2, \text{ for } j = 1,2,...,r-1 \quad (\text{if } r > 1),$$
$$\alpha_{r+s} = \epsilon_r{}^2; \tag{2.14a}$$

(2) simple odd root:

$$\alpha_s = \epsilon_s{}^1 - \epsilon_1{}^2. \tag{2.14b}$$

In terms of these simple roots the expressions for the linear functionals $\epsilon_p{}^1$ and $\epsilon_p{}^2$ of $B(r/s)$ are

$$\epsilon_p{}^1 = \sum_{j=p}^{r+s} \alpha_j;$$

and

$$\epsilon_p{}^2 = \sum_{j=p}^{r} \alpha_{j+s},$$

from which the expressions for the roots in terms of the simple roots follow immediately from (2.12) and (2.13).

III. THE CHAINS osp $(2m+1/2n+2)\supset$osp$(2m+1/2n)\supset$osp$(2m,2n)$

A. osp$(2m+1/2n)$ as a subalgebra of osp$(2m+1/2n+2)$

The detailed structure of the osp$(2m+1/2n+2)$ Lie superalgebra is given by setting $r = m$ and $s = n+1$ in the analysis of Sec. II. We also will assume in the following that $m \geqslant 1$ and $n \geqslant 1$, some other cases being examined in Sec. 5.

The basis of the osp$(2m+1/2n)$ subalgebra of osp$(2m+1/2n+2)$ may be taken to consist of all the basis elements of osp$(2m+1/2n+2)$ whose $(2m+2)$th and $(2m+n+3)$th rows and columns consist entirely of zero matrix elements. [Choosing instead the $(2m+j+1)$th and $(2m+n+j+2)$th rows and columns for $j = 2,3,...,s$ would merely give a conjugate embedding.] The dimensions of the even and odd parts of osp$(2m+1/2n)$ are $m(2m+1)+n(2n+1)$ and $2n(2m+1)$, respectively. In this embedding the $(m+n)$-dimensional Cartan subalgebra of osp$(2m+1/2n)$ is spanned by $\mathbf{h}_j{}^1 \equiv (2.7)$ for $j = 2,3,...,n+1$, and $\mathbf{h}_j{}^2 \equiv (2.8)$ for $j = 1,2,...,m$, that is, it is spanned by h_α for $\alpha = \epsilon_j{}^1$ $(j = 2,3,...,n+1)$ and $\alpha = \epsilon_j{}^2$ $(j = 1,2,...,m)$.

The $2(m^2+n^2)$ even roots of osp$(2m+1/2n)$ [$= B(m/n)$] again fall in general into eight classes:

(1) $\alpha = \epsilon_p{}^1 - \epsilon_q{}^1$ for $p,q = 2,3,...,n+1$ $\quad (p \neq q)$;

(2) $\alpha = \epsilon_p{}^1 + \epsilon_q{}^1$ for $p,q = 2,3,...,n+1$ $\quad (p \leqslant q)$;

(3) $\alpha = -(\epsilon_p{}^1 + \epsilon_q{}^1)$ for $p,q = 2,3,...,n+1$ $\quad (p \leqslant q)$;

(4) $\alpha = \epsilon_p{}^2 - \epsilon_q{}^2$ for $p,q = 1,2,..,m$ $\quad (p \neq q)$; $\quad$ (3.1)

(5) $\alpha = \epsilon_p{}^2 + \epsilon_q{}^2$ for $p,q = 1,2,...,m$ $\quad (p < q)$;

(6) $\alpha = -(\epsilon_p{}^2 + \epsilon_q{}^2)$ for $p,q = 1,2,...,m$ $\quad (p < q)$;

(7) $\alpha = \epsilon_p{}^2$ for $p = 1,2,...,m$;

(8) $\alpha = -\epsilon_p{}^2$ for $p = 1,2,...,m$;

Notice that if $m = 1$, the sets (4)–(6) are empty.

The $2n(2m+1)$ odd roots of osp$(2m+1/2n)$ fall into six classes:

(1) $\alpha = \epsilon_q{}^1$ for $q = 2,3,...,n+1$;

(2) $\alpha = -\epsilon_q{}^1$ for $q = 2,3,...,n+1$;

(3) $\alpha = \epsilon_p{}^1 + \epsilon_q{}^2$ for $p = 1,2,...,m$,

and $q = 2,3,...,n+1$; $\quad$ (3.2)

(4) $\alpha = -(\epsilon_p{}^1 + \epsilon_q{}^2)$ for $p = 1,2,...,m$,

and $q = 2,3,...,n+1$;

(5) $\alpha = \epsilon_p{}^1 - \epsilon_q{}^2$ for $p = 1,2,...,m$,

and $q = 2,3,...,n+1$;

(6) $\alpha = -(\epsilon_p{}^1 - \epsilon_q{}^2)$ for $p = 1,2,...,m$,

and $q = 2,3,...,n+1$.

The results can be summarized by the statement that the linear functional $\epsilon_1{}^1$ does not appear in any root of osp$(2m+1/2n)$.

B. osp$(2m/2n)$ as a subalgebra of osp$(2m+1/2n)$

The basis of the osp$(2m/2n)$ subalgebra of the osp$(2m+1/2n+2)$ superalgebra described in the previous subsection may be taken to consist of all the basis elements of osp$(2m+1/2n)$ whose $(2m+1)$th row and column consist entirely of zero matrix elements. The dimensions of the even and odd parts of osp$(2m/2n)$ are $m(2m-1)+n(2n+1)$ and $4nm$, respectively. [In the notation of Kac[3] osp$(2m/2n) = D(m/n)$ if $m \geqslant 2$, but osp$(2/2n) = C(n+1)$.] In this embedding the $(m+n)$-dimensional Cartan subalgebra of osp$(2m/2n)$ is again spanned by $\mathbf{h}_j{}^1 \equiv (2.7)$ for $j = 2,3,...,n+1$, and $\mathbf{h}_j{}^2 \equiv (2.8)$ for $j = 1,2,...,m$, that is, it is again spanned by h_α for $\alpha = \epsilon_j{}^1$ $(j = 2,3,...,n+1)$ and $\alpha = \epsilon_j{}^2$ $(j = 1,2,...m)$.

The $2(m^2 - m + n^2)$ even roots of osp$(2m/2n)$ fall in general into six classes, which are given by (1)–(6) of (3.1). It should be noticed that if $m = 1$, the sets (4)–(6) are empty.

The $4mn$ odd roots of osp$(2m/2n)$ fall into four classes, which are given by (1)–(4) of (3.2).

C. The complement of osp$(2m/2n)$ in osp$(2m+1/2n)$ and its important role

The complement of osp$(2m/2n)$ in osp$(2m+1/2n)$ will be denoted by osp$(2m/2n)_{\text{comp}}$. It is the $2(m+n)$-dimensional subspace of osp$(2m+1/2n)$ [and also of osp$(2m+1/2n+2)$] that has as its basis the $2m$ even elements e_α with $\alpha = \pm\epsilon_j{}^2$ (for $j = 1,2,...,m$) and the $2n$ odd elements e_α with $\alpha = \pm\epsilon_j{}^1$ (for $j = 2,3,...,n+1$). In the following, $[e_\alpha, e_\beta]$ indicates the anticommutator $[e_\alpha, e_\beta]_+$ if both e_α and e_β are odd, but otherwise represents the commutator $[e_\alpha, e_\beta]_-$.

Proposition 1: If a is any element of $\mathrm{osp}(2m/2n)$ and b is any element of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$, then

$$[a,b]\in \mathrm{osp}(2m/2n)_{\mathrm{comp}}, \tag{3.3}$$

and every element of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$ appears in this way. This can be written more concisely as

$$[\mathrm{osp}(2m/2n),\mathrm{osp}(2m/2n)_{\mathrm{comp}}] = \mathrm{osp}(2m/2n)_{\mathrm{comp}}. \tag{3.4}$$

Proposition 2: If a and b are any two elements of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$, then

$$[a,b]\in \mathrm{osp}(2m/2n), \tag{3.5}$$

and every element of $\mathrm{osp}(2m/2n)$ appears in this way. Again, this can be written more concisely as

$$[\mathrm{osp}(2m/2n)_{\mathrm{comp}},\ \mathrm{osp}(2m/2n)_{\mathrm{comp}}] = \mathrm{osp}(2m/2n).$$

Proof: Both of these results are immediate consequences of a well-known theorem in the theory of simple Lie superalgebras.[3,4,21] This states that if α and β are roots of a simple Lie superalgebra $\mathscr{L}_s$ and if $e_\alpha \in \mathscr{L}_{s\alpha}$ and $e_\beta \in \mathscr{L}_{s\beta}$ then $[e_\alpha, e_\beta] \in \mathscr{L}_{s(\alpha+\beta)}$ if $\alpha+\beta$ is a root of $\mathscr{L}_s$ and $[e_\alpha, e_\beta] = 0$ if $\alpha+\beta$ is not a root of $\mathscr{L}_s$. The propositions are then immediate consequences of the following observations. (i) the sum $\alpha+\beta$ of every root α of $\mathrm{osp}(2m/2n)$ with every root β corresponding to an element e_β of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$ is either not a root of $\mathrm{osp}(2m+1/2n+2)$ or is a root corresponding to an element of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$, and every root corresponding to an element of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$ can be obtained in this way; (ii) the sum $\alpha+\beta$ of every pair of roots α and β corresponding to elements e_α and e_β of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$ is either not a root of $\mathrm{osp}(2m+1/2n+2)$ or is a root of $\mathrm{osp}(2m/2n)$, and every root of $\mathrm{osp}(2m/2n)$ can be obtained in this way.

Both of these propositions concern the structure of $\mathrm{osp}(2m+1/2n)$ alone, and do not depend on the embedding of $\mathrm{osp}(2m+1/2n)$ in $\mathrm{osp}(2m+1/2n+2)$.

IV. THE HEISENBERG SUPERALGEBRA sh($2m/2n$)

A. The embedding of the Heisenberg superalgebra sh($2m/2n$) in osp($2m+1/2n+2$)

The Heisenberg superalgebra $\mathrm{sh}(2m/2n)$ consists [cf. (1.6)] of an identity I, $2n$ other even basis elements $P_{+,k}$ and $P_{-,l}$ $(k,l = 1,...,n)$ and $2m$ odd basis elements $T_{+,i}$ and $T_{-,j}$ $(i,j = 1,...,m)$; the only nonzero commutation and anticommutation relations are assumed to be

$$[P_{-,k},P_{+,l}]_- = 2\omega\delta_{kl}I, \quad [T_{-,i},T_{+,j}]_+ = \delta_{ij}I. \tag{4.1}$$

Such a Heisenberg superalgebra $\mathrm{sh}(2m/2n)$ can be embedded in $\mathrm{osp}(2m+1/2n+2)$ by making the following identifications:

(i) $I = e_\alpha$, with $\alpha = 2\epsilon_1{}^1$;

(ii) $P_{+,k} = (2\omega)^{1/2} e_\alpha$, with $\alpha = \epsilon_1{}^1 + \epsilon_k{}^1$ (for $k = 2,3,...,n+1$); (4.2)

(iii) $P_{-,k} = (2\omega)^{1/2} e_\alpha$, with $\alpha = \epsilon_1{}^1 - \epsilon_k{}^1$ (for $k = 2,3,...,n+1$);

(iv) $T_{+,j} = e_\alpha$, with $\alpha = \epsilon_1{}^1 + \epsilon_j{}^2$ (for $j = 1,2,...,m$);

(v) $T_{-,j} = -e_\alpha$, with $\alpha = \epsilon_1{}^1 - \epsilon_j{}^2$ (for $j = 1,2,...,m$).

Apart from the fairly trivial numerical factors, all the commutation and anticommutation relations of $\mathrm{sh}(2m/2n)$ follow from the theorem on the roots of simple Lie superalgebras mentioned in Sec. III C when taken with the following observations: (1) the even root $2\epsilon_1{}^1$ can be written *both* as the sum of the two even roots $\epsilon_1{}^1 + \epsilon_k{}^1$ and $\epsilon_1{}^1 - \epsilon_k{}^1$ for each of n different values of k (i.e., with $k = 2,3,...,n+1$), *and* also as the sum of the two odd roots $\epsilon_1{}^1 + \epsilon_j{}^2$ and $\epsilon_1{}^1 - \epsilon_j{}^2$ for each of m different values of j (i.e., with $j = 1,2,...,m$); (2) if α is any of these $2m+2n+1$ roots of $\mathrm{osp}(2m+1/2n+2)$, then 2α is *not* a root of $\mathrm{osp}(2m+1/2n+2)$; (3) if α and β are any pair of these $2m+2n+1$ roots of $\mathrm{osp}(2m+1/2n+2)$ [*except* for the $2m+2n$ pairs of (a)], then $\alpha+\beta$ is *not* a root of $\mathrm{osp}(2m+1/2n+2)$.

B. Properties of the embedding of the Heisenberg superalgebra sh($2m/2n$) in osp($2m+1/2n+2$)

Proposition 3: The subspace spanned by the basis elements of $\mathrm{osp}(2m/2n)$ and $\mathrm{sh}(2m/2n)$ together form a *subalgebra* of $\mathrm{osp}(2m+1/2n+2)$ which has the *semidirect sum structure* $[\mathrm{osp}(2m/2n) \oplus \mathrm{sh}(2m/2n)]$.

Proof: This follows from the theorem mentioned in Sec. III C and the fact that the sum $\alpha+\beta$ of every root α of $\mathrm{osp}(2m/2n)$ with every root β corresponding to an element e_β of $\mathrm{sh}(2m/2n)$ is either not a root of $\mathrm{osp}(2m+1/2n+2)$ or is a root corresponding to an element of $\mathrm{sh}(2m/2n)$.

Proposition 4: If a is any of the elements $P_{+,k}, P_{-,l}, T_{+,i}$, or $T_{-,j}$ of $\mathrm{sh}(2m/2n)$, then, with the choice $\gamma = -\epsilon_1{}^1$, the *odd* element e_γ has the property that

$$[a,e_\gamma]\in \mathrm{osp}(2m/2n)_{\mathrm{comp}},$$

and every basis element of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$ appears in this way. This is therefore a one-to-one mapping between the elements of $\mathrm{sh}(2m/2n)$ (apart from its identity) and the elements of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$. As $[P_{\pm,k}, e_\gamma]_-$ are all *odd* and $[T_{\pm,j}, e_\gamma]_+$ are all *even*, this explains the "character reversal" phenomenon observed by Beckers *et al.*[19]

Proof: These results are again consequences of the theorem stated above. This time they depend on the observation that if α is any of the roots associated with $P_{\pm,k}$ or $T_{\pm,j}$ and $\gamma = -\epsilon_1{}^1$, then $e_{\alpha+\gamma}$ is a member of $\mathrm{osp}(2m/2n)_{\mathrm{comp}}$.

V. COMMENTS AND CONCLUSIONS

All the considerations and results of Sec. III and IV have been obtained by assuming $m \geqslant 1$ and $n \geqslant 1$. Here we wish to comment further on some special cases of the chain (1.10), i.e.,

$$\mathrm{osp}(2m+1/2n+2) \supset \mathrm{osp}(2m+1/2n) \supset \mathrm{osp}(2m/2n). \tag{5.1}$$

(a) The case $m = n = 0$ is of no significance as it leaves only osp(1/2) as the nontrivial first member of the triplet (5.1). (Such a superalgebra has already been recognized of physical interest in different applications of quantum physics.[22,23]

(b) The case m arbitrary (but$\geqslant 1$) and $n = 0$ has to be considered as a purely algebraic matter since the integer n is currently associated in the applications with the number of spatial dimensions [e.g., the one of the quantum harmonic oscillator as mentioned in the Introduction; see relations (1.1), (1.3), (1.5), and (1.7)]. Indeed let us recall that the symplectic content $\mathrm{sp}(2n)$ refers to the $2n$ bosonic variables in the context of supersymmetry. However, if $n = 0$ the whole line of argument presented in Sec. III and IV still goes through, the chain (5.1) becoming

$$\mathrm{osp}(2m+1/2) \supset \mathrm{so}(2m+1) \supset \mathrm{so}(2m). \tag{5.2}$$

The only modifications are that the orthosymplectic Lie superalgebras $\mathrm{osp}(2m+1/2n)$ and $\mathrm{osp}(2m/2n)$ just reduce to the Lie algebras $\mathrm{so}(2m+1)$ and $\mathrm{so}(2m)$, respectively, which of course do not have odd roots. Denoting the complement of $\mathrm{so}(2m)$ in $\mathrm{so}(2m+1)$ by $\mathrm{so}(2m)_{\mathrm{comp}}$, Proposition 1 becomes

$$[\mathrm{so}(2m),\ \mathrm{so}(2m)_{\mathrm{comp}}] = \mathrm{so}(2m)_{\mathrm{comp}}, \tag{5.3}$$

and Proposition 2 reduces to

$$[\mathrm{so}(2m)_{\mathrm{comp}},\ \mathrm{so}(2m)_{\mathrm{comp}}] = \mathrm{so}(2m). \tag{5.4}$$

Similarly Proposition 3 now states that the subspace spanned by the basis elements of $\mathrm{so}(2m)$ and $\mathrm{sh}(2m/0)$ together form a subalgebra of $\mathrm{osp}(2m+1/2)$ which has the semidirect sum structure

$$\{\mathrm{so}(2m) \oplus\!\!\!\!\supset \mathrm{sh}(2m/0)\}.$$

Likewise Proposition 4 states that if a is any of the odd elements T_{+j} or T_{-j} of $\mathrm{sh}(2m/0)$, then, with the choice $\gamma = -\epsilon_1{}^1$, the *odd* element e_γ of $\mathrm{osp}(2m+1/2)$ has the property that

$$[a,e_\gamma] \in \mathrm{so}(2m)_{\mathrm{comp}}, \tag{5.5}$$

and every basis element of $\mathrm{so}(2m)_{\mathrm{comp}}$ appears in this way.

(c) The case $m = n = 1$ clearly leads to the chain (1.9) that was summarized in a previous study.[20] This corresponds to the physical context of a one-dimensional harmonic oscillator within the now well-understood correspondence (1.4) that involves 12-dimensional superalgebras. Here the semidirect sum of the superalgebras osp(2/2) and sh(2/2) is related to the superalgebra osp(3/2). The (even) Lie algebras contained in osp(2/2) are sp(2) and so(2), which are generated by (H_B, $C_\pm$) and $H_F \equiv Y$, respectively, where H_B and H_F are the bosonic and fermionic Hamiltonians and the other two generators of sp(2) are associated with dilations and expansions.[13,16] The four other (odd) osp(2/2) generators are the $Q_\pm$ (Ref. 5)—and $S_\pm$ (Ref. 9)—supercharges which are such that[10]

$$[Q_+,Q_-]_+ = H_B + H_F, \quad [S_+,S_-]_+ = H_B - H_F. \tag{5.6}$$

Finally in addition to the (even) Lie algebra $h(2)$ generated by (P_+,P_-,I), where $P_\pm$ can be interpreted as being boson annihilation and creation operators, the Heisenberg superalgebra sh(2/2) contains two more (odd) generators denoted by $T_\pm$ and satisfying

$$[T_+,T_-]_+ = I. \tag{5.7}$$

[in accordance with (4.1)]. The superalgebra osp(3/4) is singled out as the first member of the chain (1.9) for the simple reason that it is the smallest superalgebra containing *both* the superalgebras osp(3/2) and $\mathrm{osp}(2/2) \oplus\!\!\!\!\supset \mathrm{sh}(2/2)$. The corresponding generalization provides the starting point for all the following cases.

(d) The case $m = n \neq 0$ is the direct generalization of the preceding one leading to the chain

$$\mathrm{osp}(2n+1/2n+2) \supset \mathrm{osp}(2n+1/2n) \supset \mathrm{osp}(2n/2n) \tag{5.8}$$

and to the correspondence (1.7). This demonstrates that the "maximal" or largest invariance superalgebra[12] of the n-dimensional harmonic oscillator is the nonsimple superalgebra $\mathrm{osp}(2n/2n) \oplus\!\!\!\!\supset \mathrm{sh}(2n/2n)$, which, by virtue of (1.7), is associated with the simple superalgebra $\mathrm{osp}(2n+1/2n)$. This is a new result that does not appear in previous works.[7,12,17,19] Physically this case applies to a situation which admits the same number ($2n$) of bosonic and fermionic degrees of freedom and which is associated with the standard supersymmetrization procedure "à la Witten."[5] This largest invariance superalgebra has $8n^2 + 4n(+1)$ dimensions consisting of $8n^2$ generators for $\mathrm{osp}(2n/2n)$ and $4n(+1)$ generators for $\mathrm{sh}(2n/2n)$. Let us mention[12] that the $\mathrm{osp}(2n/2n)$ superalgebra is generated (for $k,l = 1,2,\ldots,n$) by the $(2n^2+n)$ operators $(T_{kl},C_{\pm,kl})$ of $\mathrm{sp}(2n)$, by the $(2n^2-n)$ operators $(Y_{kl},Z_{\pm,kl})$ of $\mathrm{so}(2n)$ as well as by the $4n^2$ supercharges $(Q_{\pm,kl},S_{\pm,kl})$, while the Heisenberg superalgebra $\mathrm{sh}(2n/2n)$ contains the $4n$ significant generators $(P_{\pm,k},T_{\pm,k})$. This corresponds to the $(4n^2+2n)$ *odd* operators $(Q_{\pm,kl},S_{\pm,kl},T_{\pm,k})$. Here we see also that with respect to $\mathrm{osp}(2n+1/2n)$, the $\mathrm{sp}(2n)$ content is unchanged and that we get a (new) $\mathrm{so}(2n+1)$ subalgebra generated by $(Y_{kl},Z_{\pm,kl},$ *and* $T'_{\pm,k})$ while the *odd* part now consists of the $(4n^2+2n)$ operators $(Q_{\pm,kl},S_{\pm,kl},P'_{\pm,k})$, where primes refer to the "reversal character" proposed in Ref. 19 and explained in Sec. IV.

(e) The case $m = 1$, n arbitrary (but$\geqslant 1$) leads to the chain

$$\mathrm{osp}(3/2n+2) \supset \mathrm{osp}(3/2n) \supset \mathrm{osp}(2/2n) \tag{5.9}$$

and to the correspondence (1.5), which are significant relations connecting the invariance superalgebra $\mathrm{osp}(2/2n) \oplus\!\!\!\!\supset \mathrm{sh}(2/2n)$ proposed by Beckers *et al.*[19] with the Lie superalgebra $\mathrm{osp}(3/2n)$ proposed by Englefield.[17] In such a case we are restricted to only two fermionic degrees of freedom while the $2n$ bosonic degrees still exist. This corresponds to the supersymmetrization procedures[12,24] which are nonstandard for $n \neq 1$ and their properties have been explored.[19] Inside the $\mathrm{osp}(2/2n)$ superalgebra the $\mathrm{sp}(2n)$ content is here unchanged with respect to case (d) but the preceding $\mathrm{so}(2n)$ algebra reduces to a Lie algebra so(2), which is generated by the only Y operator present.[19] This case is clearly the starting point of the present study leading to the general case $m \geqslant 1$, $n \geqslant 1$ we have already analyzed.

Let us now conclude this paper. We have concentrated attention on the chains (5.1) and their properties in Sec. III and IV as well as on the specific cases discussed in this section. In each of these chains of three orthosymplectic superalgebras, we have pointed out the special significance of their second member as well as the role of the complement of their third member. The first member of each triplet is always the smallest superalgebra containing both the second superalgebra *and* the semidirect sum of the third one with its associated Heisenberg superalgebra. We also observe that these chains are closed in the sense that it is *not* possible to construct an extended chain containing more than three members which possesses the analogous property for *every* consecutive set of three members. As far as the physics of the quantum harmonic oscillator (in arbitrary n spatial dimensions) is concerned, we have seen that cases (c)–(e) above are useful. They give us a good understanding of the corresponding invariance superalgebra, their kinematical and dynamical contents can be explicitly extracted, and the corresponding constants of motion can be identified. In every case it corresponds to a well-defined supersymmetric Hamiltonian readily obtained by extending previous results.[12,19]

ACKNOWLEDGMENTS

The authors would like to thank Dr. J. Van der Jeugt for drawing their attention on the correspondence (1.8) with $m \neq n$ as well as for his interest in the present work.

[1] P. Ramond, Phys. Rev. D **3**, 2415 (1971); A. Neveu and J. Schwarz, Nucl. Phys. B **31**, 86 (1971); D. V. Volkov and V. P. Akulov, Phys. Lett. **46**, 109 (1973); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974); A. Salam and J. Strathdee, *ibid.* **76**, 477 (1974); Phys. Rev. D **11**, 1521 (1975).

[2] P. Fayet and S. Ferrara, Phys. Rep. C **32**, 250 (1977); *Supersymmetry in Physics* (North-Holland, Amsterdam, 1985); M. F. Sohnius, Phys. Rep. C **128**, 39 (1985).

[3] V. G. Kac, Adv. Math **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977); *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978), p. 587; M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146, 155 (1977); V. Rittenberg, *Lecture Notes in Physics*, Vol. 79 (Springer, Berlin, 1977), p. 3.

[4] M. Scheunert, *The Theory of Lie Superalgebras* (Springer, Berlin, 1979).

[5] E. Witten, Nucl. Phys. B **188**, 513 (1981); **202**, 253 (1982).

[6] P. Salomonson and J. Van Holten, Nucl. Phys. **196**, 509 (1982).

[7] M. de Crombrugghe and V. Rittenberg, Ann. Phys. (NY) **151**, 199 (1983).

[8] F. Cooper and B. Freedman, Ann. Phys. (NY) **146**, 262 (1983); D. Lancaster, Nuovo Cimento A **79**, 28 (1984).

[9] S. Fubini and E. Rabinovici, Nucl. Phys. B **245**, 17 (1984).

[10] J. Beckers and V. Hussin, Phys. Lett. A **118**, 319 (1986).

[11] J. Beckers, D. Dehin, and V. Hussin, J. Phys. A **20**, 1137 (1987).

[12] J. Beckers, D. Dehin, and V. Hussin, J. Phys. A **21**, 651 (1988).

[13] U. Niederer, Helv. Phys. Acta **45**, 802 (1972); **46**, 191 (1973).

[14] A. O. Barut, Phys. Rev. **135**, B 839 (1964).

[15] B. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).

[16] C. R. Hagen, Phys. Rev. D **5**, 377 (1972).

[17] M. J. Englefield, J. Phys. A **21**, 1309 (1988).

[18] J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).

[19] J. Beckers, D. Dehin, and V. Hussin, J. Math. Phys. **29**, 1705 (1988).

[20] J. Beckers and J. F. Cornwell, J. Phys. A **22**, 925 (1989).

[21] J. F. Cornwell, *Group Theory in Physics* (Academic, London, 1989), Vol. III.

[22] E. D'Hoker and L. Vinet, Phys. Lett. B **137**, 72 (1984).

[23] A. B. Balantekin, H. A. Schmitt, and B. R. Barett, J. Math. Phys. **29**, 1634 (1988).

[24] J. Gamboa and J. Zanelli, Phys. Lett. B **165**, 91 (1985); A. B. Balantekin, Ann. Phys. (NY) **164**, 277 (1985).

Multiseries Lie groups and asymptotic modules for characterizing and solving integrable models

Marcel Jaulent
Laboratoire de Physique Mathématique,[a] *Université des Sciences et techniques du Languedoc, 34060 Montpellier cedex, France*

Miguel A. Manna
Instituto de Física Teórica, Universidade Estadual Paulista, Rua Pamplona 145, 01405 São Paulo, Brazil

Luis Martínez Alonso
Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense, 28040, Madrid, Spain

(Received 19 October 1988; accepted for publication 22 March 1989)

A multiseries integrable model (MSIM) is defined as a family of compatible flows on an infinite-dimensional Lie group of N-tuples of formal series around N given poles on the Riemann sphere. Broad classes of solutions to a MSIM are characterized through modules over rings of rational functions, called asymptotic modules. Possible ways for constructing asymptotic modules are Riemann–Hilbert and $\bar{\partial}$ problems. When MSIM's are written in terms of the "group coordinates," some of them can be "contracted" into standard integrable models involving a small number of scalar functions only. Simple contractible MSIM's corresponding to one pole, yield the Ablowitz–Kaup–Newell–Segur (AKNS) hierarchy. Two-pole contractible MSIM's are exhibited, which lead to a hierarchy of solvable systems of nonlinear differential equations consisting of $(2+1)$-dimensional evolution equations and of quite strong differential constraints.

I. INTRODUCTION

During the past ten years or so the application of Lie-algebraic methods has clarified and developed essential parts of the theory of integrable systems. These methods lead to simple geometric interpretations of integrable systems and exhibit important algebraic properties.[1–3] Moreover, they provide a link between two basic problems of the theory: namely, to classify integrable models and describe their solutions. The purpose of this paper is to present a Lie-algebraic scheme that allows us to determine new hierarchies of integrable systems and analyze relevant families of their solutions. We use the term integrable because there are natural methods for constructing solutions to these systems. However, we are not concerned here with aspects such a Hamiltonian formalism or constants of motion (see Refs. 1–3). The main points of our approach are as follows.

(1) A general class of integrable models called multiseries integrable models (MSIM's) is introduced. A MSIM is defined as a family of compatible flows on an infinite-dimensional Lie group of N-tuples of formal series with matrix-valued coefficients around N given poles on the Riemann sphere S. These flows arise as a consequence of the presence of a double Lie algebra structure $\mathscr{G} = \mathscr{G}_+ + \mathscr{G}_-$. Here $\mathscr{G}_+$ is isomorphic to a subset of an associative algebra $\mathscr{R}$ of matrix-valued rational functions of a complex variable k with given poles.

(2) When MSIM's are written in terms of the "group coordinates," some of them can be "contracted" into standard integrable models consisting of systems of nonlinear differential equations (NDE's) involving a small number of scalar functions only. In general, reductions of standard integrable models correspond to reductions of MSIM's. Simple contractible MSIM's corresponding to one pole yield the AKNS hierarchy.[4]

Two-pole contractible MSIM's are exhibited that lead to a hierarchy (H) of solvable systems of NDE's consisting of $(2+1)$-dimensional evolution equations and of quite strong differential constraints. Among these systems we mention

$$q_t = \alpha\left(\tfrac{1}{4}\, q_{xxx} - \tfrac{3}{2}\, q^2 q_x\right) + \alpha'\left(-\tfrac{1}{4}\, q'_{xyy} + \tfrac{3}{2}(q')^2 q'_x\right), \quad (1.1a)$$

$$q'_t = \alpha'\left(\tfrac{1}{4}\, q'_{yyy} - \tfrac{3}{2}(q')^2 q'_y\right) + \alpha\left(-\tfrac{1}{4}\, q_{yxx} + \tfrac{3}{2}\, q^2 q_y\right), \quad (1.1b)$$

$$q_y = -q'_x, \quad \frac{q_{xy}}{q} = \frac{q'_{xy}}{q'}, \quad \left(\frac{q_{xy}}{4q}\right)^2 = 1 + \frac{(q_y)^2}{4}, \quad (1.1c)$$

where the unknowns are the complex functions $q(x,y,t)$ and $q'(x,y,t)$, α and α' are given complex numbers, and $q_t, q_x,\ldots$ mean $\partial_t q, \partial_x q,\ldots$. Observe that the system (1.1a)–(1.1c) describes a time evolution in the manifold of solutions to the bidimensional NDE's [(1.1b) and (1.1c)]. On the other hand, since (1.1c) implies $q'_y = \int 4q'\sqrt{1 + \frac{1}{4}(q'_x)^2}\, dx$, i.e., y derivations correspond to some x integrations, the system (1.1a)–(1.1c) may be interpreted formally as an integrodifferential evolution equation in $(1+1)$ dimensions.

Note also that for $\alpha = -4$ and $\alpha' = 0$, (1.1a) reduces to the modified Korteweg-de Vries (MKdV) equation, $q_t + q_{xxx} - 6q^2 q_x = 0$, relative to the variables (x,t).

(3) Certain objects called asymptotically normalized wave functions (NW functions) and asymptotic modules

[a] Unité de Recherche Associée au C.N.R.S. N.UA 040768, Recherche Coopérative sur Programme N.P 080264.

 0022-2488/89/081662-12$02.50

(AM's) play a fundamental role for characterizing broad classes of solutions to MSIM's and as a consequence to their associated standard integrable models [such as (1.1a)–(1.1c)] in the contractible cases. An AM is a module over the previously introduced ring $\mathscr{R}$. Both the NW functions and the elements of AM's are functions of k admitting specific asymptotic expansions (AE's) around the given poles. For a considered MSIM, each NW function determines one solution. On the other hand any AM provides a new NW function from a given NW function, i.e., we thus obtain an iterative procedure for generating solutions and, in particular, for adding solitons. Possible ways for constructing AM's are Riemann–Hilbert problems and $\bar{\partial}$ (DBAR) equations outside of given poles on the Riemann sphere.

Concerning points (1) and (2), we are inspired by the theory of the Kadomtsev–Petviashvili (KP) hierarchy, more specifically, by its formulation as the system of compatible flows,[5,6]

$$\frac{\partial K}{\partial t_r} = (K D^r K^{-1})_+ K - K D^r, \quad r \geqslant 1, \tag{1.2}$$

where K lies on the Volterra group of pseudodifferential operators,

$$K = 1 + \sum_{n=1}^{\infty} a_n(x,t) D^{-n}, \quad t = (t_r), \quad D = \partial_x, \tag{1.3}$$

and the subscript + in (1.2) means the differential operator part. On the other hand, we remark that the notion of loop group used for the AKNS hierarchy[1] is a particular case of that group of formal multiseries considered here.

With respect to point (3), we recall that $\bar{\partial}$ equations were very useful for extending the range of application of the inverse scattering transform method of the solution for (1 + 1) NDE's and (2 + 1) NDE's.[7] Subsequently, $\bar{\partial}$ equations were considered as the starting point for introducing and solving NE's.[8,9(a),10] As a matter of fact, the concept of AM is motivated partly by the analysis of the algebraic structure underlying the use of $\bar{\partial}$ equations in Refs. 9(a) and 10. Further motivations come from other important methods[11] for solving integrable models such as the construction of finite gap solutions in the framework of algebraic geometry and that of rational solutions in the context of the Grassmannian formalism where modules over polynomial rings also occurs.

For these cases the role of the NW functions is played by the Baker functions. In some sense AM's provide the bridge between the Grassmannian formalism and the $\bar{\partial}$ and Zakharov–Shabat dressing methods[12] for integrable systems of the AKNS type. On the other hand, they show the group-theoretical content of these solution methods. However, if one is more interested in the construction of solutions to NDE's than in the group aspects, an economical scheme based on AM's can be restated. On this point we refer to Ref. 9(b), where additional information can be found with regard to the hierarchy (H) containing (1.1a)–(1.1c). Other hierarchies of evolution NDE's, with constraints solvable within the framework of the AM scheme, are investigated in Ref. 9(c) [(2 + 1)-dimensional case] and Ref. 9(d) [$(N+1)$-dimensional case, $N \geqslant 1$]. Genuine (2 + 1)-dimensional equations can also be obtained in the context of AM's [see Ref. 9(e)]. The formulation of the associated Lie-algebraic approach would require the use of pseudodifferential operators in addition to that of formal multiseries [see (1.2) and (1.3) for the KP case]. Finally we notice that the AM scheme can be developed in discrete cases as well: see Ref. 9(f), where an integrable (2 + 1)-dimensional generalization of the Volterra model is derived.

This paper is organized as follows. Section II starts with an abstract introduction to the class of compatible families of flows used in our work. Then we define MSIM's in the one-pole case and we discuss the contractible models corresponding to the AKNS hierarchy. Section III deals with MSIM's in the general N-pole case. Some contractible models are exhibited for $N = 2$. They yield the sinh–Gordon equation and the hierarchy (H), including the system (1.1a)–(1.1c). In Sec. IV we analyze solution methods to MSIM's from the point of view of NW functions and AM's. We exhibit Riemann–Hilbert and $\bar{\partial}$ realizations of AM's. Special attention is devoted to soliton solutions and some explicit examples are worked out in detail. In particular, we give a simple derivation of the Blaschke–Potapov factor[3] of soliton dressing. More general procedures for constructing asymptotic modules, based not only on the Riemann sphere but also on higher-genus Riemann surfaces, will be presented elsewhere,[9(g)] allowing us to characterize wide classes of solutions, including the rational and soliton classes as well as those arising in the finite-zone integration method. There are also two appendices. The first one considers some properties of differential polynomials that are used throughout the paper and the other includes the proof of the fundamental property of AM's.

II. MULTISERIES INTEGRABLE MODELS (MSIM's): PRELIMINARY MATERIAL AND THE ONE-POLE CASE

A. Double Lie algebras and compatible flows on Lie groups

Much of the geometric content of integrable systems is often related with the presence of a double Lie algebra structure; that is to say, a Lie algebra $\mathscr{G}$ that admits a decomposition into a linear direct sum of two Lie subalgebras,[2,3,13,14]

$$\mathscr{G} = \mathscr{G}_+ + \mathscr{G}_-. \tag{2.1}$$

In this subsection we are going to use this structure for defining compatible flows on Lie groups. Given a double Lie algebra $\mathscr{G}$ and $u \in \mathscr{G}$ we will denote by $u_\pm$ the projections of u on $\mathscr{G}_\pm$ associated with the decomposition (2.1), and by π the projection operator,

$$\pi(u) = u_+, \quad (1-\pi)(u) = u_-.$$

Let $\hat{\mathscr{G}}$ be a Lie group with Lie algebra $\mathscr{G}$ and $\hat{\mathscr{G}}_-$ a subgroup with Lie algebra $\mathscr{G}_-$. Given a commutative family of elements $\{c_i\}_1^s$ in $\mathscr{G}$,

$$[c_i, c_j] = 0, \tag{2.2}$$

we consider the following associated family of flows on $\hat{\mathscr{G}}_-$:

$$\partial_{t_i} g = -(g c_i g^{-1})_- g, \quad g \in \hat{\mathscr{G}}_-, \quad i = 1,\dots,s. \tag{2.3}$$

Here, $g c_i g^{-1}$ denotes the image of c_i under the adjoint action of $g \in \hat{\mathscr{G}}_-$ on the Lie algebra $\mathscr{G}$ and $(g c_i g^{-1})_- g$ is the image of $(g c_i g^{-1})_- \in \mathscr{G}_-$ under the right-translation action of $g\in\hat{\mathscr{G}}_-$. Thus the right-hand side of (2.3) determines a well-defined vector field on $\hat{\mathscr{G}}_-$. The remarkable property of these flows is that they commute with each other, i.e., that the quantity

$$Y \doteq (\partial_{t_i}\partial_{t_j}g - \partial_{t_j}\partial_{t_i}g)g^{-1}$$

is zero. To prove this, we use the identities

$$Y = [(\partial_{t_i}g)g^{-1},(\partial_{t_j}g)g^{-1}] + \partial_{t_i}((\partial_{t_j}g)g^{-1}) - \partial_{t_j}((\partial_{t_i}g)g^{-1}),$$

and

$$\partial_{t_i}(g\,c_jg^{-1}) = [(\partial_{t_i}g)g^{-1},g\,c_jg^{-1}],$$

so that (2.3) implies

$$Y = [(q_i)_-,q_j]_- - [(q_j)_-,q_i]_- + [(q_j)_-,(q_i)_-], \tag{2.4}$$

where

$$q_i = g\,c_ig^{-1}. \tag{2.5}$$

Now from (2.2), $[q_i,q_j] = g[c_i,c_j]\,g^{-1} = 0$, and therefore

$$[(q_j)_-,q_i] = -[(q_j)_+,q_i] = -[(q_j)_+,(q_i)_+] - [(q_j)_+,(q_i)_-].$$

Hence

$$[(q_j)_-,q_i]_- = -[(q_j)_+,(q_i)_-]_-$$

since

$[(q_j)_+,(q_i)_+]_- = 0$ ($\mathscr{G}_+$ is a Lie subalgebra).

By inserting this into (2.4) we get

$$Y = [(q_i)_-,(q_j)_-]_- + [(q_j)_-,(q_i)_-].$$

Hence $Y=0$ since

$$[(q_i)_-,(q_j)_-]_- = [(q_i)_-,(q_j)_-]$$

($\mathscr{G}_-$ is a Lie subalgebra).

As a consequence of the compatibility, it is reasonable to consider simultaneous solutions $g(t)$ $[t=(t_1,\dots,t_s)]$ to the family of flows (2.3). These solutions constitute the main elements of our analysis in the specific cases described below where $\mathscr{G}$ is an algebra of formal multiseries.

It is worth noting that $\hat{\mathscr{G}}$, the Lie group with Lie algebra $\mathscr{G}$, plays no role in the above discussion. Indeed, the construction of the flows (2.3) only requires three basic objects: a double Lie algebra $\mathscr{G}$, a Lie group $\hat{\mathscr{G}}_-$ with Lie algebra $\mathscr{G}_-$, such that $\hat{\mathscr{G}}_-$ acts on $\mathscr{G}$ by means of the adjoint action, and some commutative subset $\{c_i\}_1^s$ of $\mathscr{G}$. This is an important fact, since in the applications we have to deal with infinite-dimensional Lie algebras for which $\hat{\mathscr{G}}$ turns out to be a much more complicated object than $\hat{\mathscr{G}}_-$. In the cases relevant to this paper, both structures, $\mathscr{G}$ and $\hat{\mathscr{G}}_-$, are immersed in some associative algebra of formal multiseries that allows us to define the adjoint action of $\hat{\mathscr{G}}_-$ on $\mathscr{G}$ in a natural way.

Finally we notice that it is sometimes useful to consider a family of reduced flows, i.e., a family of flows (2.3) on a subgroup $\hat{\mathscr{G}}'_-$ of $\hat{\mathscr{G}}_-$, which is invariant under certain automorphisms.

B. Definition of MSIM's in the one-pole case

Let $\mathscr{A}$ be the associative algebra of formal series of the form

$$u(k) = \sum_{n=-\infty}^{N} u_nk^n, \quad N\in\mathbb{Z}, \tag{2.6}$$

with $d\times d$ matrix coefficients u_n. The product uv in $\mathscr{A}$ is defined through a term by term series multiplication. The subset $\mathscr{G} = \{u\in\mathscr{A}$ s.t. tr $u_n = 0$ for all $n\}$ is a Lie algebra with the Lie product $[u,v]\doteq uv - vu$. Further, $\mathscr{G}$ admits a double Lie algebra structure determined by the projection operator

$$\pi(u) = \sum_{0\leqslant n\leqslant N} u_nk^n. \tag{2.7}$$

The corresponding Lie subalgebras $\mathscr{G}_+$ and $\mathscr{G}_-$ of $\mathscr{G}$ are given by

$$\mathscr{G}_+ = \{u\in\mathscr{G}/u_n = 0 \text{ for all } n<0\},$$
$$\mathscr{G}_- = \{u\in\mathscr{G}/u_n = 0 \text{ for all } n\geqslant 0\}.$$

Products of exponentials of elements in $\mathscr{G}_-$ generate a Lie group $\hat{\mathscr{G}}_-\subset\mathscr{A}$, whose elements are of the form

$$g(k) = \sum_{n=0}^{\infty} a_nk^{-n}, \quad a_0 = 1, \tag{2.8a}$$

and satisfy the constraint

$$\det g(k) = 1. \tag{2.8b}$$

We will refer to the coefficients $\{a_n\}_1^\infty$ as the coordinates of the group element g. Note that the inversion operation in $\hat{\mathscr{G}}_-$ can be performed as follows:

$$g^{-1} = 1 + \sum_{l=1}^{\infty}(-1)^l(g-1)^l = 1 - a_1k^{-1} + \cdots. \tag{2.9}$$

Thus it is clear that $\hat{\mathscr{G}}_-$ has a well-defined adjoint action on $\mathscr{G}$,

$$\mathscr{G}\to\mathscr{G}, \quad \mathrm{ad}g(u) = gug^{-1}.$$

Given a commutative family, $\{c_i\}_1^s\subset\mathscr{G}_+$, we define the associated MSIM as the family of compatible flows (2.3) on $\hat{\mathscr{G}}_-$. This definition corresponds to the case of a unique pole $k=\infty$ on the Riemann sphere. A reduction of a MSIM is defined as a family of flows (2.3) on a subgroup $\hat{\mathscr{G}}'_-$ of $\hat{\mathscr{G}}_-$, which is invariant under certain automorphisms.

C. Simple contractible MSIM's: The AKNS hierarchy

Suppose now that $d=2$ and consider a commutative family in $\mathscr{G}_+$ of the form

$$c_i(k) = \omega_i(k)\sigma_3, \quad i=1,\dots,s, \tag{2.10}$$

where $\omega_i(k)$ are arbitrary polynomials in k and σ_3 is the Pauli matrix. The associated MSIM is the family of compatible flows,

$$\partial_{t_i}g = -(\omega_i(k)g\sigma_3g^{-1})_-g, \quad g\in\hat{\mathscr{G}}_-. \tag{2.11}$$

Each of these equations can be described in terms of the coordinates $\{a_n\}_1^\infty$ of the group element (2.8a), so that (2.11) constitutes a system of equations with an infinite number of dependent scalar variables. However, because of the compatibility of this system, it is possible to deduce standard integrable models, i.e., differential equations involving a small number of dependent scalar variables. We say that the MSIM is "contractible."

Let us make explicit the case $s=2$, $c_1 = -ik\sigma_3$, $c_2 = i\omega(k)\sigma_3$ for an arbitrary polynomial $\omega(k)$. The associated MSIM consists of the two flows

$$\partial_x g = (ikg\sigma_3 g^{-1})_- g, \tag{2.12a}$$

$$\partial_t g = -(i\omega(k) g\sigma_3 g^{-1})_- g, \tag{2.12b}$$

where we have set $x = t_1$ and $t = t_2$. Let us introduce the following element of $\mathscr{G}$:

$$r = ig\sigma_3 g^{-1}, \tag{2.13}$$

which, according to (2.8) and (2.9), is of the form

$$r = \sum_{n=0}^{\infty} r_n k^{-n}, \quad r_0 = i\sigma_3. \tag{2.14}$$

One proves (see Appendix A) that the matrix elements of the coefficients r_n are polynomials in the matrix elements of $[\sigma_3, a_1]$ and their derivatives, with respect to x. Now, Eq. (2.12b) can be rewritten as

$$(\partial_t g) g^{-1} = (\pi - 1)(\omega(k) r(k)). \tag{2.15}$$

Then if

$$\omega(k) = \sum_{l=0}^{N} \alpha_l k^l,$$

by identifying the coefficients of k^{-1} in Eq. (2.15), we find at once

$$\partial_t a_1 = -\sum_{l=0}^{N} \alpha_l r_{l+1},$$

which implies

$$i\partial_t [\sigma_3, a_1] = -i \sum_{l=0}^{N} \alpha_l [\sigma_3, r_{l+1}]. \tag{2.16}$$

These differential equations are the members of the AKNS hierarchy. They involve the matrix elements of $[\sigma_3, a_1]$ only.

Reductions of AKNS hierarchy equations can be obtained by means of reductions of the MSIM (2.11). For example, the modified KdV hierarchy (MKdV) is characterized as follows. Let $\widehat{\mathscr{G}}'_-$ be the set of elements $g \in \widehat{\mathscr{G}}_-$, verifying

$$\sigma_1 g(-k)\sigma_1 = g(k). \tag{2.17}$$

Clearly $\widehat{\mathscr{G}}'_-$ is a subgroup of $\widehat{\mathscr{G}}_-$ and its Lie algebra $\mathscr{G}'_-$ is given by the elements of $\mathscr{G}_-$ satisfying (2.17). Now if we take $c_i(k) = \omega_i(k)\sigma_3$ with $\omega_i(k)$ being odd polynomials in k, then the right-hand side of (2.11) determines a vector field on $\widehat{\mathscr{G}}'_-$ and consequently we obtain a reduction on $\widehat{\mathscr{G}}'_-$ of the MSIM (2.11). Since $\sigma_1 a_1 \sigma_1 = -a_1$ for the coordinate a_1 of $g \in \widehat{\mathscr{G}}'_-$, these reduced MSIM's impose a compatible constraint to the AKNS equations (2.16), which turns out to yield the members of the MKdV hierarchy (see Appendix A).

III. MULTISERIES INTEGRABLE MODELS (MSIM's): THE GENERAL CASE

A. Formal multiseries

In order to apply the construction of compatible flows of Sec. II A to define general MSIM's, we will use algebras of multiseries as the basic algebraic objects, so it is convenient to introduce some appropriate notation conventions.

Given a positive integer d and N different points $\{k_n\}_1^N$ on the Riemann sphere S with $k_1 = \infty$, let $\mathscr{A}$ be the set of N-tuples of the formal series

$$u = (u_1(k), \ldots, u_N(k)), \tag{3.1a}$$

$$u_1(k) = \sum_{m=-\infty}^{M_1} u_{1m} k^m, \tag{3.1b}$$

$$u_n(k) = \sum_{m=M_n}^{\infty} u_{nm}(k-k_n)^m, \quad n = 2, \ldots, N, \tag{3.1c}$$

where $M_n \in \mathbb{Z}$ and u_{nm} are $d \times d$ matrix coefficients. We point out that all the formal series involved here are of finite order M_n at their corresponding reference points. Therefore, besides the usual notions of sum and multiplication by complex numbers, we can define a product operation in $\mathscr{A}$,

$$uu' = (u_1(k)u_1'(k), \ldots, u_N(k)u_N'(k)),$$

where the products $u_n(k)u_n'(k)$ are understood in the sense of a term by term series multiplication. With these operations $\mathscr{A}$ becomes an associative algebra.

Furthermore, let $\mathscr{R}$ be the associative algebra of $d \times d$ matrix-valued rational functions on S with possible poles at $\{k_n\}_1^N$ only. Given $U \in \mathscr{R}$, let us denote by $u_n(k)$ its corresponding Laurent series at $k = k_n$. Then the map

$$\mathscr{R} \xrightarrow{\tau} \mathscr{A}, \quad \tau(U) = (u_1(k), \ldots, u_N(k)) \tag{3.2}$$

is an injective homomorphism between the associative algebras $\mathscr{R}$ and $\mathscr{A}$.

The following linear map will be particularly important in our discussion:

$$\mathscr{A} \xrightarrow{p} \mathscr{R}, \quad p(u) = \sum_{0 \leqslant m \leqslant M_1} u_{1m} k^m + \sum_{n=2}^{N} \left(\sum_{M_n \leqslant m < 0} u_{nm}(k-k_n)^m \right). \tag{3.3}$$

Here u_{nm} are the coefficients of the formal series $u_n(k)$ that determine $u \in \mathscr{A}$ [see (3.1)]. Observe that $p(u)$ is obtained by adding the principal parts of the series $u_n(k)$ ($n = 1, \ldots, N$) and the constant term of $u_1(k)$. Now consider the composition of τ and p,

$$\mathscr{A} \xrightarrow{\tau} \mathscr{A}, \quad \pi = \tau \circ p. \tag{3.4}$$

It follows at once from the partial fractial decomposition theorem that $p \circ \tau = \mathrm{Id}_{\mathscr{R}}$, so that π is a projection operator on the vector space $\mathscr{A}$ that determines a decomposition of $\mathscr{A}$ into a linear direct sum of two subalgebras,

$$\mathscr{A} = \mathscr{A}_+ + \mathscr{A}_-, \tag{3.5}$$

where $\mathscr{A}_+ = \mathrm{Ran}\, \pi = \tau(\mathscr{R})$ and $\mathscr{A}_- = \mathrm{Ker}\, \pi$ consists of the elements (3.1), such that $M_1 = -1$ and $M_n = 0$ for $n = 2, \ldots, N$.

B. Definition of MSIM's

We are now ready to generalize the definition of MSIM's given in Sec. II B by considering Lie algebras and groups contained in the associative algebra $\mathscr{A}$ of multiseries (3.1). The subset $\mathscr{G} = \{u \in \mathscr{A}$ s.t. tr $u_{nm} = 0$ for all $n,m\}$ is a Lie algebra with the Lie product $[u,v] \doteqdot uv - vu$. Here $\mathscr{G}$ is an invariant subspace of $\mathscr{A}$ under the projection operator π of (3.4), and the corresponding restriction $\mathscr{G} \xrightarrow{\pi} \mathscr{G}$ determines a double Lie algebra structure on $\mathscr{G}$ with $\mathscr{G}_{\pm} = \mathscr{G} \cap \mathscr{A}_{\pm}$.

By means of the exponentials of elements of $\mathscr{G}_-$ one generates a Lie group $\widehat{\mathscr{G}}_-$ that consists of the elements of $\mathscr{A}$,

$$g = (g_1(k),\ldots,g_N(k)),$$

$$g_1(k) = 1 + \sum_{m=1}^{\infty} g_{1m} k^{-m}, \tag{3.6}$$

$$g_n(k) = \sum_{m=0}^{\infty} g_{nm}(k-k_n)^m, \quad n = 2,\ldots,N,$$

such that

$$\det g_n(k) = 1, \quad n = 1,\ldots,N. \tag{3.7}$$

We will refer to the coefficients g_{nm} as the coordinates of $g \in \widehat{\mathscr{G}}_-$. It is easy to see that $\widehat{\mathscr{G}}_-$ has a well-defined adjoint action on $\mathscr{G}$.

Henceforth formal Lie algebras and groups of N-tuples of $d \times d$ matrix-valued series centered at $\{k_n\}_1^N$ will be called multiseries Lie algebras and groups with reference points $\{k_n\}_1^N$.

At this point we can take any commutative family $\{c_i\}_1^s \subset \mathscr{G}_+$ and define the associated MSIM as the family of compatible flows (2.3) on $\widehat{\mathscr{G}}_-$. Equations (2.3) can be described in terms of the coordinates g_{nm} of g so that they constitute a system of equations with an infinite number of dependent scalar variables. However, because of the compatibility of this system, in some cases it is possible to deduce standard integrable models, i.e., differential equations involving a small number of dependent scalar variables. Then we say that the MSIM is contractible. Many integrable models in $1+1$ dimensions can be obtained by the contraction of MSIM's. Among these are the Toda lattice, the sine–Gordon, and the Chiral-field models.

A reduction of a MSIM is defined as a family of flows (2.3) on a subgroup $\widehat{\mathscr{G}}'_-$ of $\widehat{\mathscr{G}}_-$ that is invariant under certain automorphisms. In general, reductions of standard integrable models correspond to reductions of MSIM's.

As an illustration of the above abstract considerations we now analyze the simple case given by $d=2$, $N=2$, and $\{k_1 = \infty, k_2 = 0\}$. Here the elements of $\mathscr{A}$ can be written as $u = (u_1(k), u_2(k))$ with

$$u_1(k) = \sum_{m=-\infty}^{M_1} u_{1m} k^m,$$
$$u_2(k) = \sum_{m=M_2}^{\infty} u_{2m} k^m, \quad M_1, M_2 \in \mathbb{Z}, \tag{3.8}$$

where u_{nm} are 2×2 matrices. In this particular case the map (3.2) takes a very simple form. Indeed since $\mathscr{R}$ is now the set of rational functions on $\mathbb{S}$ with poles at $k_1 = \infty$ and $k_2 = 0$ only, its elements are finite sums of the form $U = \sum_{m=N}^{M} u_m k^m$, with $M, N \in \mathbb{Z}$ $(M \geqslant N)$. Hence

$$\tau(U) = (U(K), U(K)), \quad U \in \mathscr{R}. \tag{3.9}$$

The Lie algebra $\mathscr{G}$ and the Lie group $\widehat{\mathscr{G}}_-$ are readily characterized as subsets of $\mathscr{A}$. We express the elements $g = (g_1(k), g_2(k))$ of $\widehat{\mathscr{G}}_-$ as

$$g_1(k) = 1 + \sum_{m=1}^{\infty} a_m k^{-m}, \quad g_2(k) = \sum_{m=0}^{\infty} b_m k^m. \tag{3.10}$$

In the following subsections we give some examples of contractible MSIM's in the case $d=2$, $N=2$, and $\{k_1 = \infty, k_2 = 0\}$. Some N-pole contractible MSIM's (with $N > 2$) will be investigated elsewhere.

C. A contractible two-pole MSIM: The sinh–Gordon equation

Let us consider the MSIM defined by the following pair of compatible flows on $\widehat{\mathscr{G}}_-$:

$$\partial_x g = -(gcg^{-1})_- g, \quad \partial_y g = -(gc'g^{-1})_- g, \tag{3.11}$$

associated with the following commuting elements in $\widehat{\mathscr{G}}_+$:

$$c = \tau(-ik\sigma_3) = (-ik\sigma_3, -ik\sigma_3), \tag{3.12a}$$

$$c' = \tau(-ik^{-1}\sigma_3) = [-i(\sigma_3/k), -i(\sigma_3/k)]. \tag{3.12b}$$

From the definition (3.3) of the map p we have

$$p(gcg^{-1}) = -ik\sigma_3 + i[\sigma_3, a_1],$$

$$p(gc'g^{-1}) = -(i/k) b_0 \sigma_3 b_0^{-1}.$$

Then, taking into account (3.4) and (3.9), Eq. (3.11) implies that

$$\partial_x g_j = ik[g_j, \sigma_3] + i[\sigma_3, a_1] g_j, \tag{3.13a}$$

$$\partial_y g_j = (i/k) b_0 [b_0^{-1} g_j, \sigma_3], \quad j = 1,2. \tag{3.13b}$$

By substituting the expansions (3.10) into (3.13) and identifying the terms in k^0 and $1/k$ in (3.13a) and (3.13b), respectively, we get

$$\partial_x b_0 = i[\sigma_3, a_1] b_0, \quad \partial_y a_1 = i[\sigma_3, b_0] b_0^{-1}, \tag{3.14}$$

which imply a differential equation for b_0,

$$\partial_y[(\partial_x b_0) b_0^{-1}] = [\sigma_3, b_0 \sigma_3 b_0^{-1}]. \tag{3.15}$$

Furthermore, since $\det g_2(k) = 1$, b_0 satisfies the constraint $\det b_0 = 1$.

Now we consider the reduction of the previous MSIM on the group $\widehat{\mathscr{G}}'_-$ of elements, $g = (g_1, g_2) \in \widehat{\mathscr{G}}_-$, such that g_j $(j = 1,2)$ satisfies (2.17). [This is possible because $-ik\sigma_3$ and $-ik^{-1}\sigma_3$ satisfy (2.17).] Under this assumption we deduce $\sigma_1 b_0 \sigma_1 = b_0$, which together with $\det b_0 = 1$ implies the following form for b_0:

$$b_0 = \begin{pmatrix} \cosh\varphi & \sinh\varphi \\ \sinh\varphi & \cosh\varphi \end{pmatrix}, \tag{3.16}$$

where $\varphi \in \mathbb{C}$. Now from (3.15) one finds at once the sinh–Gordon equation for φ,

$$\partial_{xy}\varphi = -2\sinh(2\varphi). \tag{3.17}$$

D. Contractible two-pole MSIM's: A hierarchy of evolution (2+1) NDE's with constraints

Let us consider the MSIM obtained by adding the following flow to the system (3.11):

$$\partial_t g = -(gc'' g^{-1})_- g, \tag{3.18}$$

where $c'' \in \mathcal{G}_+$ is defined by

$$c'' = \tau(i\omega(k)\sigma_3) = (i\omega(k)\sigma_3, i\omega(k)\sigma_3), \tag{3.19a}$$

with

$$\omega(k) = \sum_{l=-M}^{N} \alpha_l k^l, \quad N,M>0. \tag{3.19b}$$

Now the solutions to (3.11) and (3.18) will depend on three variables (x,y,t). In order to investigate the differential equations satisfied by the coordinates of g we rewrite (3.18) as

$$(\partial_t g)g^{-1} = (\pi - 1)(g\tau(i\omega\sigma_3)g^{-1}). \tag{3.20}$$

It is easy to see that $g\tau(i\omega\sigma_3)g^{-1}$ can be written in the form

$$g\tau(i\omega\sigma_3)g^{-1} = (\omega r, \omega b_0 r' b_0^{-1}), \tag{3.21a}$$

where r and r' are defined by

$$r = g_1 i\sigma_3 g_1^{-1}, \quad r' = g_1' i\sigma_3 g_1'^{-1}, \tag{3.21b}$$

with

$$g_1' \doteq b_0^{-1} g_2 = 1 + \sum_{m=1}^{\infty} a_m' k^m, \quad a_m' \doteq b_0^{-1} b_m. \tag{3.22}$$

Both r and r' are interesting objects since they depend on a few group coordinates only. Indeed, they are of the form

$$r = \sum_{n=0}^{\infty} r_n k^{-n}, \quad r' = \sum_{n=0}^{\infty} r_n' k^n, \quad r_0 = r_0' = i\sigma_3, \tag{3.23}$$

with the coefficients r_n (resp. r_n') being polynomials in the off-diagonal elements of a_1 (resp. of $a_1' = b_0^{-1} b_1$) and their derivatives with respect to the x (resp. y) variable. To prove this we notice that

$$\operatorname{tr} r = \operatorname{tr} r' = 0, \quad \det r = \det r' = 1. \tag{3.24}$$

On the other hand, we have

$$\partial_x r = -ik[\sigma_3, r] + i[[\sigma_3, a_1], r], \tag{3.25a}$$

$$\partial_y r' = -(i/k)[\sigma_3, r'] + i[[\sigma_3, a_1'], r']. \tag{3.25b}$$

Equation (3.25a) is an immediate consequence of Eq. (3.13a) for g_1, whereas Eq. (3.25b) follows from the equation

$$\partial_y g_1' = (i/k)[g_1', \sigma_3] + i[\sigma_3, a_1'] g_1', \tag{3.26}$$

which in turn derives from Eq. (3.13b) for g_2 by taking into account that this latter implies

$$\partial_y b_0 = i b_0 [a_1', \sigma_3]. \tag{3.27}$$

Now, as it is proved in Appendix A, Eqs. (3.23)–(3.25) are all we need to conclude that r_n (resp. r_n') is a polynomial in q, s (resp. q', s') and their x derivatives (resp. y derivatives), where

$$i[\sigma_3, a_1] = -\begin{pmatrix} 0 & q \\ s & 0 \end{pmatrix}, \quad i[\sigma_3, a_1'] = -\begin{pmatrix} 0 & q' \\ s' & 0 \end{pmatrix}. \tag{3.28}$$

Moreover, in view of (3.25) and (3.28), it is clear that r_n' is obtained from r_n by replacing $(\partial_x^m q, \partial_x^m s)$ with $(\partial_y^m q', \partial_y^m s')$ for all $m \geqslant 0$.

We are now ready to write (3.20) in terms of the coordinates of g. However, we are only interested in obtaining differential equations for the first few coordinates. In this sense observe that

$$(\partial_t g)g^{-1} = \left(\frac{\partial_t a_1}{k} + O\left(\frac{1}{k^2}\right), (\partial_t b_0)b_0^{-1} + b_0(\partial_t a_1')b_0^{-1}k + O(k^2)\right).$$

Furthermore, from (3.3) and (3.19)–(3.23) it follows that

$$p(g\tau(i\omega\sigma_3)g^{-1}) = \sum_{n=0}^{N} k^n \left(\sum_{l=n}^{N} \alpha_l r_{l-n}\right) + \sum_{n=1}^{M} \frac{1}{k^n}\left(\sum_{l=n}^{M} \alpha_{-l} b_0 r_{l-n}' b_0^{-1}\right).$$

Then, since $\pi = \tau \circ p$ and taking (3.9) into account by identifying coefficients in (3.20), we get

$$i\partial_t[\sigma_3, a_1] = -i \sum_{l=-1}^{N} \alpha_l [\sigma_3, r_{l+1}] + i \sum_{l=1}^{M} \alpha_{-l} [\sigma_3, b_0 r_{l-1}' b_0^{-1}], \tag{3.29a}$$

$$i\partial_t[\sigma_3, a_1'] = -i \sum_{l=-1}^{M} \alpha_{-l} [\sigma_3, r_{l+1}'] + \sum_{l=1}^{N} \alpha_l [\sigma_3, b_0^{-1} r_{l-1} b_0], \tag{3.29b}$$

$$\partial_t b_0 = \sum_{l=0}^{N} \alpha_l r_l b_0 - \sum_{l=0}^{M} \alpha_{-l} b_0 r_l'. \tag{3.29c}$$

From the properties of r_n and r_n' these differential equations involve the matrix elements of $[\sigma_3, a_1]$, $[\sigma_3, a_1']$, and b_0 only, and contain derivatives with respect to three variables x, y, and t. On the other hand, besides (3.29a)–(3.29c) we have to consider the equations corresponding to the flows (3.11) as well. These later derive from (3.13) and reduce to four additional relations:

$$i[\sigma_3, a_1] = (\partial_x b_0)b_0^{-1}, \quad i[\sigma_3, a_1'] = -b_0^{-1}\partial_y b_0, \tag{3.29d}$$

$$i\partial_y[\sigma_3, a_1] = [\sigma_3, b_0\sigma_3 b_0^{-1}],$$
$$i\partial_x[\sigma_3, a_1'] = [\sigma_3, b_0^{-1}\sigma_3 b_0]. \tag{3.29e}$$

By using (3.29d) we can express r_l and r_l' in terms of b_0 and its derivatives with respect to x and y. In this way (3.29c) becomes an evolution equation for the matrix b_0. The remaining constraints on b_0 follow from (3.29d) and (3.29e) and are resumed by the equation

$$\partial_y((\partial_x b_0)b_0^{-1}) = [\sigma_3, b_0\sigma_3 b_0^{-1}]. \tag{3.30}$$

Observe that Eqs. (3.29a) and (3.29b) are a consequence of (3.29c) and (3.29d). To see this point, it is enough to differentiate (3.29c) with respect to x and y, taking into account (3.29d) and the following equations:

$$b_0^{-1}(\partial_y r)b_0 = -(i/k)[\sigma_3, b_0^{-1} r b_0],$$
$$b_0(\partial_x r')b_0^{-1} = -ik[\sigma_3, b_0 r' b_0^{-1}],$$

which derive from (3.13) and imply

$$b_0^{-1}(\partial_y r_{l+1})b_0 = -i[\sigma_3, b_0^{-1} r_l b_0],$$
$$b_0(\partial_x r'_{l+1})b_0^{-1} = -i[\sigma_3, b_0 r'_l b_0^{-1}].$$

We also remark [see Ref. 9(b)] that by eliminating b_0 in Eqs. (3.29a) and (3.29b) it is possible to obtain a system of four evolution scalar NDE's with some differential constraints. These NDE's involve the four scalar functions q, s, q', and s' defined by (3.28).

As an illustration of the rich structure [Ref. 9(b)] that underlies Eqs. (3.29), we will analyze one of the reductions of the MSIM [(3.11)–(3.18)]. Suppose that $\omega(k)$ is an odd polynomial,

$$\omega(K) = \sum_{l=-M}^{N} \alpha_{2l+1} k^{2l+1}. \tag{3.31}$$

Then $i\omega(k)\sigma_3$ satisfies (2.17) and the three flows (3.11), (3.18) can be defined on the subgroup $\hat{\mathscr{G}}'_-$ of elements $g=(g_1,g_2)\in\hat{\mathscr{G}}_-$, such that g_j verifies (2.17) for $j=1,2$. Let g be a solution on the reduced group; then we have

$$\sigma_1 g_1(-k)\sigma_1 = g_1(k), \quad \sigma_1 g'_1(-k)\sigma_1 = g'_1(k), \tag{3.32}$$

$$\sigma_1 b_0 \sigma_1 = b_0. \tag{3.33}$$

As it is shown in Appendix A, (3.32) implies $q=s$ and $q'=s'$, that is to say

$$i[\sigma_3, a_1] = -q\sigma_1, \quad i[\sigma_3, a'_1] = -q'\sigma_1. \tag{3.34}$$

Furthermore, if we set

$$r_n = \begin{pmatrix} \xi_n & \eta_n \\ \gamma_n & -\xi_n \end{pmatrix}, \quad r'_n = \begin{pmatrix} \xi'_n & \eta'_n \\ \gamma'_n & -\xi'_n \end{pmatrix},$$

then (3.32) means that relations (A12) of Appendix A hold for both r_n and r'_n. Hence

$$r_{2n+1} = \eta_{2n+1}\sigma_1, \quad r'_{2n+1} = \eta'_{2n+1}\sigma_1. \tag{3.35}$$

On the other hand, as we saw in Sec. III C, (3.33) enables us to write b_0 in the form (3.16), which implies

$$\begin{aligned}(\partial_t b_0)b_0^{-1} = \varphi_t\sigma_1, \quad (\partial_x b_0)b_0^{-1} = \varphi_x\sigma_1,\\ b_0^{-1}(\partial_y b_0) = \varphi_y\sigma_1.\end{aligned} \tag{3.36}$$

In this way, one easily finds that Eqs. (3.29c) and (3.29d) become

$$\varphi_t = \sum_{l=0}^{N} \alpha_{2l+1}\eta_{2l+1} - \sum_{l=1}^{M} \alpha_{-2l+1}\eta'_{2l-1}, \tag{3.37}$$

$$q = -\varphi_x, \quad q' = \varphi_y, \tag{3.38}$$

$$q_y = -q'_x = 2\sinh(2\varphi). \tag{3.39}$$

By using (3.38) we can express η_{2l+1} and η'_{2l-1} in terms of φ so that (3.37) represents a hierarchy of evolution NDE's in $(2+1)$ dimensions for the single function φ. However, the function φ is subject to a differential constraint; it derives from (3.38) and (3.39) and takes the form

$$\varphi_{xy} = -2\sinh(2\varphi), \tag{3.40}$$

as it should be expected in view of (3.30) and of the results of Sec. III C. In addition, it is easy to see that differentiation of (3.37) with respect to x and y yields evolution NDE's for q and q' in $(2+1)$ dimensions. Note also that (3.38) and (3.39) determine the differential constraints,

$$q_y = -q'_x, \quad \frac{q_{xy}}{q} = \frac{q'_{xy}}{q'}, \quad \left(\frac{q_{xy}}{4q}\right)^2 = 1 + \frac{(q_y)^2}{4}. \tag{3.41}$$

The first nontrivial example of (3.37) corresponds to $\omega(k) = \alpha k^3 + \alpha' k^{-3}$ and leads to

$$\varphi_t = \alpha(\tfrac14 \varphi_{xxx} - \tfrac12(\varphi_x)^3) + \alpha'(\tfrac14 \varphi_{yyy} - \tfrac12(\varphi_y)^3). \tag{3.42}$$

Differentiation of (3.42), with respect to x and y, gives

$$q_t = \alpha(\tfrac14 q_{xxx} - \tfrac32 q^2 q_x) + \alpha'(-\tfrac14 q'_{xyy} + \tfrac32 (q')^2 q'_x), \tag{3.43a}$$

$$q'_t = \alpha'(\tfrac14 q'_{yyy} - \tfrac32 (q')^2 q'_y) + \alpha(-\tfrac14 q_{yxx} + \tfrac32 q^2 q_y). \tag{3.43b}$$

In addition to these evolution equations, the functions φ and (q,q') satisfy the constraints (3.40) and (3.41), respectively.

IV. ASYMPTOTIC MODULES (AM's) AND SOLUTION METHODS

The MSIM's described in the preceding sections admit natural solution methods based on the construction of particular objects called normalized wave functions and asymptotic modules. In particular, these methods provide solutions to the standard integrable models associated with contractible MSIM's.

A. Normalized wave (NW) functions

Again we will use the multipole structures $\mathscr{A}$, $\mathscr{R}$, $\mathscr{G}$, $\mathscr{G}_+$, $\mathscr{G}_-$, and $\hat{\mathscr{G}}_-$ with reference points $\{k_n\}_1^N$ and the maps τ, p, $\pi = \tau\circ p$ defined in Sec. III. Let us take a commutative family $\{c_i\}_1^s \subset \mathscr{G}_+$. Since $\mathscr{G}_+ = \mathscr{G}\cap\tau(\mathscr{R})$ there is a commutative family $\{C_i\}_1^s \subset \mathscr{R}$ with $\operatorname{tr} C_i = 0$, such that

$$c_i = \tau(C_i). \tag{4.1}$$

We look for solutions to the MSIM (2.3) associated with (4.1). This system can be rewritten as

$$\partial_{t_i} g = \pi(g c_i g^{-1})g - g c_i, \quad g\in\hat{\mathscr{G}}_-. \tag{4.2}$$

We now introduce the following associative functional algebras $\mathscr{A}(D)$. Let D be a subset of the Riemann sphere S such that $\{k_n\}_1^N$ are limit points of D. By $\mathscr{A}(D)$ we will denote the set of $d\times d$ matrix-valued functions $H = H(k)$, defined on D, for which there is an element $h = (h_1(k),\dots,h_N(k))$ in $\mathscr{A}$, such that $H(k)$ admits $h_n(k)$ as its asymptotic expansion (AE) as $k\to k_n$ $(n=1,\dots,N)$. Obviously, $\mathscr{R}\subset\mathscr{A}(D)$ and the map (3.2) admits an extension,

$$\mathscr{A}(D) \xrightarrow{\tau} \mathscr{A}, \quad \tau(H) = (h_1(k),\dots,h_N(k)), \tag{4.3}$$

which is a homomorphism between the associative algebras $\mathscr{A}(D)$ and $\mathscr{A}$. Next we define the projection operator,

$$\mathscr{A}(D) \xrightarrow{\Pi} \mathscr{A}(D), \quad \Pi = p\circ\tau. \tag{4.4}$$

From (3.3)–(3.5) it follows at once that $\Pi_{\mathscr{R}} = Id_{\mathscr{R}}$, $\operatorname{Ran}\Pi = \mathscr{R}$, while $\operatorname{Ker}\Pi = \tau^{-1}(\mathscr{A}_-)$. The maps τ and Π enable us to formulate a version of (4.2) on $\mathscr{A}(D)$. Indeed, consider the system

$$\partial_{t_i} G = \Pi(G C_i G^{-1})G - G C_i, \tag{4.5a}$$

with the conditions valid for any $t = (t_1,\dots,t_s)$,

$$G(t),\quad G^{-1}(t)\in\mathscr{A}(D),\quad \tau(G(t))\in\widehat{\mathscr{G}}_-. \tag{4.5b}$$

Given a solution $G(t)$ of (4.5), since $\tau\circ\Pi=\pi\circ\tau$ and τ is an algebra homomorphism, it follows at once that $g(t)=\tau(G(t))$ is a solution of (4.2). We may still perform a further reformulation of our problem by means of the function

$$F=G\exp\Big(\sum_{i=1}^{s}t_iC_i\Big). \tag{4.6}$$

It is clear that (4.5a) is equivalent to

$$\partial_{t_i}F=\Pi(FC_iF^{-1})F, \tag{4.7}$$

while (4.5b) is verified if F satisfies

$$\det F=1 \tag{4.8}$$

and admits AE's of the form

$$F(k,t)\sim\Big(1+\sum_{m=1}^{\infty}g_{1m}(t)k^{-m}\Big)\exp\Big(\sum_{i=1}^{s}t_iC_i(k)\Big),\quad k\to\infty, \tag{4.9a}$$

$$F(k,t)\sim\Big(\sum_{m=0}^{\infty}g_{nm}(t)(k-k_n)^m\Big)\exp\Big(\sum_{i=1}^{s}t_iC_i(k)\Big),\quad k\to k_n,\quad n=2,\dots,N. \tag{4.9b}$$

A function $F(k,t)$ (such that $k\in D\subset S$ and $\{k_n\}_1^N$ are limit points of D), which satisfies (4.7)–(4.9), will be called an asymptotically normalized wave function (NW function) on D for the MSIM (4.2). To summarize we can state that each NW function $F(k,t)$ determines a solution of (4.2) in the form

$$g(t)=\tau\Big(F(k,t)\exp\Big(-\sum_{i=1}^{s}t_iC_i\Big)\Big),\quad t=(t_1,\dots,t_s). \tag{4.10}$$

We notice that there is an elementary NW function on $S-\{k_n\}_1^N$ given by

$$E(k,t)=\exp\Big(\sum_{i=1}^{s}t_iC_i\Big),$$

which corresponds to the trivial solution $g=1$ to (4.2).

B. Asymptotic modules

A set $\mathscr{W}$ of $d\times d$ matrix-valued functions defined on a subset $D\subset S$ is said to be a (left) $\mathscr{R}$ module if it satisfies

$$H_1+H_2\in\mathscr{W},\quad\text{for all } H_1,H_2\in\mathscr{W}, \tag{4.11a}$$

$$UH\in\mathscr{W},\quad\text{for all } U\in\mathscr{R},\quad H\in\mathscr{W}. \tag{4.11b}$$

We are going to see how some $\mathscr{R}$ modules, called asymptotic modules (AM's), allow us to reproduce NW functions. Let D_0 and D_1 be subsets of S such that $\{k_n\}_1^N$ are limit points for $D_0\cap D_1$. We are given a NW function, $F_0=F_0(k,t)$, on D_0 for the MSIM (4.2) and we want to produce another NW function, $F_1=F_1(k,t)$, on D_1 for (4.2). The simplest case corresponds to $F_0(k,t)=E(k,t)$. The strategy is the following. First it is convenient to look for a function F differing from F_1 by a normalization factor, i.e., condition (4.8) is not required for F_1. Here F is generally defined on a subset D_1' bigger than D_1. On one hand we characterize the behavior of F in D_1' by looking for $F(t)$ in a fixed $\mathscr{R}$ module for all t (isospectrality condition). Furthermore, in order to guarantee the uniqueness of F, we impose that FF_0^{-1} satisfies some regularity condition in some appropriate set D_{01} containing $D_0\cap D_1'$ and admits some asymptotic structure at $\{k_n\}_1^N$.

More precisely, we will say that a set $\mathscr{W}$ of $d\times d$ matrix-valued functions $H(k)$, defined on D_1', is an AM around the NW function $F_0(k,t)$ on D_0 for the MSIM (4.2) if the following conditions hold:

(1) $\mathscr{W}$ is an $\mathscr{R}$ module; (2) for each value of $t=(t_1,\dots,t_s)$, there is a *unique* function $F(t)\in\mathscr{W}$ such that (i) the function $\widehat{F}(t)\doteq F(t)\,F_0^{-1}(t)$ defined for $k\in D_0\cap D_1'$ has a "smooth" extension in D_{01}, and (ii) $\widehat{F}(t)$ belongs to $\mathscr{A}(D_{01})$ with $\Pi(\widehat{F}(t))=1$, i.e., $\widehat{F}(k,t)$ admits AE's of the form

$$\widehat{F}(k,t)\sim 1+\sum_{m=1}^{\infty}\varphi_{1m}(t)k^{-m},\quad k\to\infty, \tag{4.12a}$$

$$\widehat{F}(k,t)\sim\sum_{m=0}^{\infty}\varphi_{nm}(t)(k-k_n)^m,\quad k\to k_n,\quad n=2,\dots,N; \tag{4.12b}$$

and (3) $\det F(k,t)\neq 0$ for $k\in D_1\subset D_1'$.

Then we have the following important property whose proof is given in Appendix B. If $\mathscr{W}$ is an AM around the NW function F_0 on D_0 for the MSIM (4.2), then it turns out that

$$F_1(k,t)=F(k,t)\,[\det F(k,t)]^{-1/d} \tag{4.13}$$

is a NW function on D_1 for (4.2).

By using (4.10) we conclude that starting from the solution

$$g_0(t)=\tau\Big(F_0(k,t)\exp\Big(-\sum_{i=1}^{s}t_iC_i\Big)\Big)$$

of the MSIM (4.2) we can construct the new solution

$$g_1(t)=\tau\Big(F_1(k,t)\exp\Big(-\sum_{i=1}^{s}t_iC_i\Big)\Big).$$

Thus we obtain an iterative procedure for generating solutions to MSIM's and as a consequence to their associated standard integrable models [such as the AKNS hierarchy and the system (1.1a)–(1.1c) identical with (3.43) and (3.41)] in the contractible cases.

Note that the terminology in this paper is slightly different from Ref. 9(b) where we use the ring $\widetilde{\mathscr{R}}$ of $\mathscr{R}$-valued functions of t and we name AM the $\widetilde{\mathscr{R}}$ module $\widetilde{\mathscr{W}}$ of dimension 1 and of basis $F(k,t)$.

C. Construction of asymptotic modules

First we remark that $\mathscr{R}$ modules can be constructed in a natural way by means of Riemann–Hilbert and $\bar\partial$ problems.

Let γ be an oriented curve in $S-\{k_n\}_1^N$ and let $G(k)$ be a $d\times d$ matrix-valued function defined on γ. Let us denote by $\mathscr{W}$ the set of $d\times d$ matrix-valued functions, defined on $D_1'=S-(\gamma\cup\{k_n\}_1^N)$, whose left and right boundary values $H_\pm$ on γ exist and satisfy

$$H_-(k)=H_+(k)G(k). \tag{4.14}$$

Then $\mathscr{W}$ defines an obvious $\mathscr{R}$ module. We recall that such Riemann–Hilbert problems appear in the Zakharov–Shabat dressing method.[12]

Now let us consider a $d\times d$ matrix-valued distribution $R(k)$ with support in $S-\{k_n\}_1^N$. The set $\mathscr{W}$ of $d\times d$ matrix-valued functions $H(k)$, defined on $D_1'=S-\{k_n\}_1^N$, which satisfy the $\bar\partial$ equation

$$\frac{\partial H}{\partial \bar k}(k)=H(k)R(k),\quad k\in S-\{k_n\}_1^N, \tag{4.15a}$$

is also an obvious $\mathscr{R}$ module.

It is known that a Riemann–Hilbert problem can be considered formally as a particular $\bar\partial$ problem. For this reason we will only investigate the construction of AM's associated with $\bar\partial$ problems.

Let us prove that for an appropriate choice of the input function $R(k)$, the $\mathscr{R}$ module $\mathscr{W}$ associated with (4.15a) is an AM around each of the NW functions $F_0(k,t)$ considered in (a) and (b).

(a) Here $F_0(k,t)=E(k,t)$ (the elementary NW function). Then $D_0=D_1'=D_1=D_{01}=S-\{k_n\}_1^N$. The function $\hat F \doteq FF_0^{-1}$ must satisfy the $\bar\partial$ equation

$$\frac{\partial \hat F}{\partial \bar k}(k)=\hat F(k)\hat R(k),\quad k\in S-\{k_n\}_1^N, \tag{4.15b}$$

with $\hat R(k)\doteq F_0(k)R(k)F_0^{-1}(k)$. Here $\hat F$ must also admit AE's of the form (4.12). By applying the generalized Cauchy formula in a way similar to Refs. 9(a), 10(a), (c) one can see that $\hat F$ is a solution of the integral equation

$$(1-J)\hat F=1, \tag{4.15c}$$

where J is the integral operator

$$J\hat F(k)=\frac{1}{2i\pi}\iint_{\mathbb{R}^2}\frac{dq\wedge d\bar q}{q-k}\hat F(q)\hat R(q).$$

With reasonable assumptions on $R(k)$, (4.15c) has a unique solution and det $\hat F(k)$ does not vanish. As a consequence $\mathscr{W}$ is an AM around F_0. Note that if tr $R(k)=0$, then we also have tr $\hat R(k)=0$ so that formula (4.15b) implies $(\partial/\partial\bar k)\det\hat F=0$. By using (4.12) we find $\det F=\det\hat F=1$ and the formula (4.13) becomes merely $F_1=F$. Note also that the new NW function (4.13) can be computed easily when $R(k)$ is a linear combination of delta functions. The corresponding new solution of the MSIM (4.2) will be called, in a wider sense, a multisoliton solution since it yields in some cases a multisoliton solution to a standard integrable model.

(b) Here $F_0(k,t)$ satisfies a $\bar\partial$ equation,

$$\frac{\partial F_0}{\partial \bar k}(k)=F_0(k)R_0(k),\quad k\in S-\{k_n\}_1^N.$$

Then $D_0=D_1=D_1'=D_{01}=S-\{k_n\}_1^N$. The function $\hat F\doteq FF_0^{-1}$ must satisfy the $\bar\partial$ equation (4.15b) with $\hat R=F_0(R-R_0)F_0^{-1}$ and admit AE's of the form (4.12). Similar to (a), $\hat F$ is a solution of the integral equation (4.15c) and, with reasonable assumptions on $R(k)$ and $R_0(k)$, we conclude that $\mathscr{W}$ is an AM around F_0. If tr $R_0(k)$ $=$ tr $R(k)=0$, the formula (4.13) becomes $F_1=F$.

D. Solitons

There are other ways than using Riemann–Hilbert and $\bar\partial$ problems for constructing AM's. For example, we will now construct some AM's that are particularly appropriate for analyzing soliton solutions.

Let F_0 be a NW function analytic on a dense open set D_0 of S, where $S-D_0$ is made up of isolated points including the $\{k_n\}_1^N$ and of some curves. Let k_0 and k_0' be two different complex numbers in D_0 and let M and N be two subspaces of $\mathbb{C}^d$, such that $\mathbb{C}^d=M+N$. Denote by $\mathscr{W}$ the set of $d\times d$ matrix-valued functions $H(k)$ analytic on $D_1'=D_0-\{k_0\}$ with, at most, a single pole at $k=k_0$ and verifying (a) the coefficient R of $(k-k_0)^{-1}$ in the Laurent expansion of $H(k)$ at $k=k_0$ satisfies

$$R(M)=\{0\}; \tag{4.16}$$

and (b) the value S of $H(k)$ at k_0' satisfies

$$S(N)=\{0\}. \tag{4.17}$$

It is clear that $\mathscr{W}$ is an $\mathscr{R}$ module. To show that it is an AM around F_0, let us look for functions $F(t)\in\mathscr{W}$ such that $\hat F=FF_0^{-1}$ has a continuous extension in $D_{01}\doteq S-\{k_0\}$ and admits AE's of the form (4.12). It is easy to see that $\hat F$ must be, in fact, analytic in $S-\{k_0\}$ (note that $\det F_0=1$). It follows that F can be written as

$$F(k,t)=[1+A(t)/(k-k_0)]F_0(k,t). \tag{4.18}$$

Thus conditions (a) and (b) become

$$A(t)F_0(k_0,t)(M)=\{0\}, \tag{4.19a}$$

$$(k_0'-k_0+A(t))F_0(k_0',t)(N)=\{0\}. \tag{4.19b}$$

If we assume that $\mathbb{C}^d$ can be decomposed into a direct sum of the subspaces $F_0(k_0,t)(M)$ and $F_0(k_0',t)(N)$ for all t, then we can determine projection operators $P(t)$ on $\mathbb{C}^d$ from the conditions

$$\operatorname{Ker}P(t)=F_0(k_0,t)(M),$$

$$\operatorname{Ran}P(t)=F_0(k_0',t)(N).$$

In this way (4.19) can be rewritten in the simpler form,

$$A(1-P)=0,\quad (k_0'-k_0+A)P=0, \tag{4.20}$$

which immediately gives $A=(k_0-k_0')P$. Therefore there is a unique function $F(t)$ satisfying the required conditions and $\mathscr{W}$ is an AM around F_0. By using (4.13) we get the new NW function on $D_1=D_0-\{k_0,k_0'\}$,

$$F_1(k,t)=[(k-k_0)/(k-k_0')]^{d'/d}\times(1-[(k_0'-k_0)/(k-k_0)]P(t))F_0(k,t), \tag{4.21}$$

where $d'=\dim N$. The new NW function differs from the old one just by the presence of a Blaschke–Potapov factor.[3] The interpretation of this result is that the new solution of the MSIM (4.2) has an additional soliton. This is in agreement with Refs. 3 and 12 for the standard integrable models.

As an example we consider the MSIM defined by the system of compatible flows (3.11) and (3.18) and choose for F_0 the elementary NW function on $D_0=\mathbb{C}-\{0\}$:

$$E(k,x,y,t)=\exp(-i(kx+y/k-\omega(k)t)\sigma_3), \tag{4.22}$$

while M and N are defined as

$$M = \mathrm{lin}\{(\begin{smallmatrix}1\\-1\end{smallmatrix})\}, \quad N = \mathrm{lin}\{(\begin{smallmatrix}1\\1\end{smallmatrix})\}.$$

Then we get

$$P(x,y,t) = \frac{1}{2\cosh\rho_+}\begin{pmatrix} e^{\rho_+} & e^{\rho_-} \\ e^{-\rho_-} & e^{-\rho_+}\end{pmatrix},$$

where

$$\rho_\pm = \pm i[(k_0 \mp k_0')x + [1/k_0 \mp 1/k_0']y - (\omega(k_0) \mp \omega(k_0'))t].$$

The corresponding solution of the hierarchy (3.29) is characterized by

$$a_1 = (k_0 - k_0')(P - \tfrac{1}{2}), \quad a_1' = (1/k_0k_0')a_1,$$
$$b_0 = (k_0/k_0')^{1/2}(1 + [(k_0' - k_0)/k_0]P).$$

Clearly this solution represents a plane soliton.

E. Reductions

AM's are also suitable for constructing solutions to the MSIM (4.2), subject to lie on a reduced subgroup $\hat{\mathscr{G}}'_-$ of $\hat{\mathscr{G}}_-$. To illustrate this fact we consider the system (3.11)–(3.18) with the reduction (2.17). Let F_0 be a NW function, verifying

$$\sigma_1 F_0(-k,t)\sigma_1 = F_0(k,t) \tag{4.23}$$

and let $\mathscr{W}$ be an AM around F_0 such that $\sigma_1 H(-k)\sigma_1$ belongs to $\mathscr{W}$ for all $H \in \mathscr{W}$ (the subsets D, D_1', D_{01} are supposed to be symmetric with respect to 0). Under these assumptions it is not difficult to see that condition (2) for AM's implies that (4.23) holds for the new NW function F_1. From (4.10) we conclude that the solution $g_1(t)$, associated with F_1, lies on the reduced subgroup.

As an example, again let us take the elementary NW function (4.22) with $\omega(k)$ being an odd polynomial in order to satisfy condition (4.23). Let k_0 and k_0' be two different complex numbers in $D_0 = \mathbb{C} - \{0\}$ and let M and N be two subspaces such that $\mathbb{C}^2 = M \oplus N$. We define $\mathscr{W}$ as the set of 2×2 matrix functions $H(k)$, analytic on $D_1' = \mathbb{C} - \{0, \pm k_0\}$, with at most single poles at $k = \pm k_0$, and verifying that (a) the coefficients $R_\pm$ of $(k \mp k_0)^{-1}$ in the Laurent expansions of $H(k)$ at $k = \pm k_0$ satisfy

$$R_+(M) = \{0\}, \quad R_-\sigma_1(M) = \{0\},$$

and (b) the values $S_\pm$ of $H(k)$ at $k = \pm k_0'$ satisfy

$$S_+(N) = \{0\}, \quad S_-\sigma_1(N) = \{0\}.$$

It follows that $\mathscr{W}$ is an AM around $F_0 = E$ such that $\sigma_1 H(-k)\sigma_1$ belongs to $\mathscr{W}$ for all $H \in \mathscr{W}$. Thus if we define

$$M = \mathrm{lin}\{(\begin{smallmatrix}1\\0\end{smallmatrix})\}, \quad N = \mathrm{lin}\{(\begin{smallmatrix}1\\1\end{smallmatrix})\},$$

the new NW function F_1 on $D_1' = \mathbb{C} - \{0, \pm k_0, \pm k_0'\}$ turns out to be

$$F_1(k,x,y,t) = \left(\frac{k^2 - k_0^2}{k^2 - k_0'^2}\right)^{1/2}\left(1 + \frac{k_0' - k_0}{k + k_0}P_2\right) \times\left(1 + \frac{k_0 - k_0'}{k - k_0}P_1\right)E,$$

where

$$P_1 = \begin{pmatrix} 0 & e^{2\rho} \\ 0 & 1\end{pmatrix},$$

$$P_2 = \frac{1}{\gamma}\begin{pmatrix} (1 + \alpha e^{4\rho})(1+\beta) & -(1+\alpha e^{4\rho})\beta e^{2\rho} \\ (1+\alpha)(1+\beta)e^{2\rho} & -(1+\alpha)\beta e^{4\rho}\end{pmatrix},$$

with

$$\alpha = (k_0' - k_0)/(k_0 + k_0'), \quad \beta = (k_0' - k_0)/2k_0,$$
$$\rho = -ik_0'x - i(y/k_0') + i\omega(k_0')t,$$
$$\gamma = 1 + \beta + (\alpha - \beta)\,e^{4\rho}.$$

It is not difficult to compute the corresponding solution to the hierarchy (3.37). It is given by

$$\sinh(2\varphi) = 2i\sinh(2\rho + \theta)/\cosh^2(2\rho + \theta),$$

where $\theta = \log(i\beta/i + \beta)$.

ACKNOWLEDGMENTS

M. M. and L. M. A. wish to thank Professor P. C. Sabatier and the Laboratoire de Physique Mathématique de Montpellier for their warm hospitality.

APPENDIX A: STRUCTURE OF THE COEFFICIENTS r_n

In this appendix we give a simple argument that proves that the coefficients r_n of the formal series r in (2.14) are differential polynomials in the off-diagonal elements of a_1. Moreover, we study some properties of the reduction (2.17).

We begin by obtaining a differential equation for r. From (2.12a) and taking into account that r commutes with kr, we have

$$\partial_x r = [(kr)_-, r] = -[(kr)_+, r].$$

Now

$$(kr)_+ = (ikg\sigma_3 g^{-1})_+ = ik\sigma_3 - i[\sigma_3, a_1].$$

Therefore

$$\partial_x r = -ik[\sigma_3, r] + i[[\sigma_3, a_1], r]. \tag{A1}$$

Next we introduce some notation,

$$i[\sigma_3, a_1] = -\begin{pmatrix} 0 & q \\ s & 0\end{pmatrix}, \tag{A2}$$

$$r = \begin{pmatrix} \xi & \eta \\ \gamma & -\xi\end{pmatrix}. \tag{A3}$$

Observe that according to (2.13), tr $r = 0$. Note also that the matrix elements q and s in (A2) are proportional to the off-diagonal elements of a_1 $[q = -2i(a_1)_{12}, s = 2i(a_1)_{21}]$. From (2.14) we deduce that the matrix elements of r are of the form

$$\xi = \sum_{n=0}^{\infty} \xi_n k^{-n}, \quad \eta = \sum_{n=0}^{\infty} \eta_n k^{-n},$$
$$\gamma = \sum_{n=0}^{\infty} \gamma_n k^{-n}, \quad \xi_0 = i, \quad \eta_0 = \gamma_0 = 0, \tag{A4}$$

so that (A1)–(A4) imply

$$\partial_x \xi_n = s\eta_n - q\gamma_n, \tag{A5a}$$
$$2i\eta_{n+1} = -\partial_x\eta_n + 2q\xi_n, \tag{A5b}$$
$$2i\gamma_{n+1} = \partial_x\gamma_n + 2s\xi_n. \tag{A5c}$$

With the only information given by (A5) we cannot guarantee that the coefficients of the formal series ξ, η, and γ are differential polynomials in q and s. Indeed, solving (A5a) would require boundary conditions on the x dependence of ξ_n, which are absent from our analysis. Nevertheless from (2.13) it is clear that r satisfies the constraint $\det r = 1$, so that

$$\xi^2 + \gamma\eta = -1. \tag{A6}$$

This relation and (A4) imply

$$2\xi_{n+1} = -\sum_{l=1}^{n} (\xi_l \xi_{n+1-l} + \gamma_l \eta_{n+1-l}). \tag{A7}$$

Now it is obvious that (A5b), (A5c), and (A7) form a system of recursion relations that enable us to express the unknowns ξ_n, η_n, and γ_n as polynomials in q, s, and their derivatives, with respect to x. For the sake of completeness, we list some of these polynomials:

$$\begin{aligned}
&\eta_1 = q, \quad \gamma_1 = s,\\
&\eta_2 = (i/2)q_x, \quad \gamma_2 = -(i/2)s_x,\\
&\eta_3 = -\tfrac{1}{4} q_{xx} + \tfrac{1}{2} q^2 s, \quad \gamma_3 = -\tfrac{1}{4} s_{xx} + \tfrac{1}{2} q s^2,\\
&\eta_4 = -(i/8)q_{xxx} + (3i/4)qsq_x,\\
&\gamma_4 = (i/8)s_{xxx} - (3i/4)qss_x.
\end{aligned} \tag{A8}$$

The AKNS equations (2.16) can thus be written in the form

$$\partial_t q = 2i \sum_{l=0}^{N} \alpha_l \eta_{l+1}, \quad \partial_t s = -2i \sum_{l=0}^{N} \alpha_l \gamma_{l+1}. \tag{A9}$$

Let us assume now that the constraint (2.17) is satisfied. Then $\sigma_1 a_1 \sigma_1 = -a_1$ and therefore

$$q = s. \tag{A10}$$

On the other hand, (2.13) implies $\sigma_1 r(-k)\sigma_1 = -r(k)$ and consequently

$$\sigma_1 r_{2n} \sigma_1 = -r_{2n}, \quad \sigma_1 r_{2n+1} \sigma_1 = r_{2n+1}, \tag{A11}$$

or equivalently,

$$\gamma_{2n} = -\eta_{2n}, \quad \xi_{2n+1} = 0, \quad \gamma_{2n+1} = \eta_{2n+1}. \tag{A12}$$

In this way, Eqs. (A9) reduce to a single equation, provided $\alpha_l = 0$ for even l. For example, if $\omega(k) = \alpha_3 k^3$ we get the MKDV equation for q.

APPENDIX B: CONSTRUCTION OF NW FUNCTIONS FROM ASYMPTOTIC MODULES

Here we prove that the function $F_1(k,t)$ of (4.13) defines a NW function for the MSIM (4.2). To this end let us consider the functions $\partial_{t_i} F$. Since $\mathscr{W}$ is an $\mathscr{R}$ module and $\mathbb{C}\subset\mathscr{R}$, it is clear that $\mathscr{W}$ is also a complex vector space, so that under mild assumptions $\partial_{t_i} F \in W$. The function F_0 satisfies Eq. (4.7), so that

$$(\partial_{t_i} F) F_0^{-1} = \partial_{t_i} \hat{F} + \hat{F} U_{0i}, \tag{B1}$$

where

$$U_{0i} = \Pi(F_0 C_i F_0^{-1}) \in \mathscr{R}.$$

On the other hand, by virtue of conditions (2) and (3) for AM's, we have

$$\hat{F}, \hat{F}^{-1} \quad \text{and} \quad \hat{F}U_{0i} - \Pi(\hat{F}U_{0i}\hat{F}^{-1})\hat{F} \in \mathscr{A}(D_{01}),$$

$$\hat{F}U_{0i} - \Pi(\hat{F}U_{0i}\hat{F}^{-1})\hat{F} = [(1-\Pi)(\hat{F}U_{0i}\hat{F}^{-1})]\hat{F} \in \operatorname{Ker}\Pi. \tag{B2}$$

Further, we assume that $\partial_{t_i}\hat{F} \in \mathscr{A}(D_{01})$ and that $\partial_{t_i}\hat{F}$ can be asymptotically expanded through a term by term differentiation of the series (4.12). Hence

$$\partial_{t_i}\hat{F} \in \operatorname{Ker}\Pi. \tag{B3}$$

Next, we consider the function

$$F' \doteq \partial_{t_i} F - \Pi(\hat{F}U_{0i}\hat{F}^{-1})F.$$

We already know that $\partial_{t_i} F \in \mathscr{W}$. Furthermore, since $\mathscr{W}$ is an $\mathscr{R}$ module and $\operatorname{Ran}\Pi = \mathscr{R}$ then $\Pi(\hat{F}U_{0i}\hat{F}^{-1})F \in \mathscr{W}$. Therefore $F' \in \mathscr{W}$. Now from (B1)–(B3) it follows that

$$F' F_0^{-1} = \partial_{t_i}\hat{F} + \hat{F}U_{0i} - \Pi(\hat{F}U_{0i}\hat{F}^{-1})\hat{F}$$

belongs to $\mathscr{A}(D_{01})$ and, more precisely, is an element of $\operatorname{Ker}\Pi$. As a consequence $F + F' \in \mathscr{W}$, $(F+F')F_0^{-1} \in \mathscr{A}(D_{01})$ and $\Pi((F+F')F_0^{-1}) = 1$. Therefore from condition (2) for AM's we deduce that $F + F' = F$. Hence $F' = 0$ and

$$\partial_{t_i} F = \Pi(\hat{F}U_{0i}\hat{F}^{-1})F.$$

In addition, we note that (4.12) implies

$$\hat{F}(\operatorname{Ker}\Pi)\hat{F}^{-1} \subset \operatorname{Ker}\Pi,$$

so that

$$\begin{aligned}
\Pi(\hat{F}U_{0i}\hat{F}^{-1}) &= \Pi(\hat{F}F_0 C_i F_0^{-1}\hat{F}^{-1})\\
&\quad - \Pi(\hat{F}[(1-\Pi)(F_0 C_i F_0^{-1})]\hat{F}^{-1})\\
&= \Pi(F C_i F^{-1}),
\end{aligned}$$

and the differential equations for F become

$$\partial_{t_i} F = \Pi(F C_i F^{-1})F. \tag{B4}$$

Finally, as $\operatorname{tr}\Pi(FC_iF^{-1}) = 0$, Eq. (B4) implies that $\partial_{t_i} \det F = 0$, while conditions (2) and (3) for AM's lead to AE's of the form

$$\begin{aligned}
&\det F(k,t) \sim 1 + \sum_{m=1}^{\infty} d_{1m} k^{-m}, \qquad k \to \infty,\\
&\det F(k,t) \sim \sum_{m=0}^{\infty} d_{nm}(k - k_n)^m, \qquad k \to k_n,\\
&\qquad n = 2,\ldots,N,
\end{aligned} \tag{B5}$$

with $d_{n0} \neq 0$. In this way, from (B4), (B5), (4.12), and (4.9) for F_0, it readily follows that (4.13) is a NW function on D_1 for (4.2).

[1]A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).

[2]A. G. Reyman, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **131**, 118 (1983), translated in J. Sov. Math. **30**, 2319 (1985).

[3]L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, Berlin, 1987).

[4]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 294 (1974).

[5]M. Mulase, Adv. Math. **54**, 57 (1984).

[6]G. Segal and G. Wilson, Publ. Math. de l'IHES **61**, 1 (1985).

[7]R. Beals and R. R. Coifman, Séminaire Goulaouic–Meyer–Schwartz 1980–81, exp. 22, Ecole Polytechnique, Palaiseau; Physica D **18**, 242 (1986); M. J. Ablowitz, D. Bar Yaacov, and A. S. Fokas, Stud. Appl. Math. **69**, 135 (1983).

[8]V. E. Zakharov and S. V. Manakov, Funct. Anal. Appl. **19**, 89 (1985).

[9]M. Jaulent, M. Manna, and L. Martinez Alonso, (a) Inverse Prob. **4**, 123 (1988); (b) J. Phys. A **21**, L 1019 (1988); (c) Phys. Lett. A **132**, 414

(1988); (d) J. Phys. A **22**, L 13 (1989); (e) Phys. Lett. A **135**, 438 (1989); (f) J. Phys. A **21**, L719 (1988); (g) "Asymptotic modules for solving integrable models," to appear in Inverse Prob.
[10]M. Jaulent and M. Manna, (a) *Inverse Problems with Interdisciplinary Applications*, edited by P. C. Sabatier (Academic, New York, 1987); pp. 429–444; (b) J. Math. Phys. **28**, 2338 (1987); (c) Phys. Lett. A **117**, 62 (1986); (d) Inverse Prob. **2**, L35 (1986); (e) **3**, L13 (1987); (f) Europhys. Lett. **2**, 891 (1986).
[11]I. M. Krichever, Russ. Math. Surveys **32**, 185 (1977); G. Wilson, Philos. Trans. R. Soc. London Ser. A **314**, 393 (1985).
[12]V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. **13**, 166 (1979).
[13]A. G. Reyman and M. A. Semenov-Tian-Shansky, Invent. Math. **54**, 81 (1979); **63**, 423 (1981).
[14]V. G. Drinfeld, Sov. Math. Dokl. **27**, 68 (1982).

Exceptional Lie algebras in terms of fermionic creation–annihilation operators

A. Sciarrino[a)]
CRM, Universite de Montreal, C. P. 6128, Succ. A, Montreal, Quebec H3C 3J7, Canada

(Received 29 December 1988; accepted for publication 12 April 1989)

From the embedding $E_8 \supset D_4 \times D_4$, defining fermionic operators which transform, respectively, as the vectorial and spinorial representations of D_4, the algebra of E_8 is constructed in terms of a bilinear in the fermionic operators. Then, by embeddings, the adjoint and fundamental representations of the other exceptional algebras are explicitly written.

I. INTRODUCTION

String theories have focused new, deep attention to the exceptional algebras. The renewed interest started with the remark that an anomaly-free superstring theory requires as a gauge group $E_8 \times E_8$ or $SO(32)/Z_2$.[1] It became still deeper with the observed relation between space-time supersymmetry and exceptional groups in four-dimensional string theory.[2,3]

However, to our knowledge, no realization of the exceptional algebras has been given in terms of a bilinear in fermionic creation–annihilation operators as for classical algebras.[4] For G_2, such a realization can be obtained by its embedding in SO(7). This kind of formulation, besides the advantages of any explicit construction of the algebra, has revealed itself to be very useful for classical algebras as it allows an immediate connection with the second quantization language and it provides a useful tool to build up representations.

The purpose of this paper is to propose such a formulation for the exceptional algebras. There are two key points: (i) the embedding

$$E_8 \supset D_4 \times D_4', \tag{1.1}$$

with the corresponding decomposition of the 248 (adjoint) representation

$$248 = (28,1) + (1,28') + (8_v,8_v') + (8_s,8_s') + (8_c,8_c') \tag{1.2}$$

where we have denoted with $28, 8_v, 8_s, 8_c$, respectively, the adjoint, the vectorial, and the two spinorial representations of D_4, in Dynkin notation,[5] respectively, (0100), (1000), (0001), (0010); (ii) the definition of fermionic operators that transform, respectively, as the 8_v, 8_s, and 8_c, and the use of the triality property.

In Sec. II, we recall the construction of the D_4 algebra in terms of fermionic creation and annihilation operators, which transform as the 8_v and introduce fermionic operators transforming as the 8_s and 8_c. In Sec. III, the exceptional algebra E_8 is explicitly written using the decomposition (1.2). In Secs. IV and V, by the embeddings $E_{i+1} \supset E_i$, the adjoint and fundamental representations of E_7 and E_6 are explicitly written.

The realization of the algebra F_4 (G_2) is obtained by the folding of the algebra E_6 (D_4), and is presented in Sec. VI, (Sec. VII), together with the fundamental representation.

II. D_4 AND ITS FUNDAMENTAL REPRESENTATIONS

In this section, we recall the well-known construction of D_4 in terms of a bilinear in fermionic operators and introduce fermionic operators that transform as the spinorial representations. One starts with four fermionic creation (annihilation) operators a_i^+ (a_i), which satisfy

$$i,j = 1,2,3,4,$$
$$\{a_i,a_j\} = \{a_i^+,a_j^\dagger\} = 0, \tag{2.1}$$
$$\{a_i^\dagger,a_j\} = \delta_{ij}.$$

The 28 generators of D_4 are given by ($i,j = 1,2,3,4$)

$$h_i = a_i^\dagger a_i, \tag{2.2a}$$

$$a_i^+ a_j^+,\ a_j a_i,\ a_i^+ a_j,\ a_j^+ a_i \quad (i<j). \tag{2.2b}$$

The generators of Eq. (2.2b) correspond to the roots $(e_i + e_j)$, $-(e_i + e_j)$, $(e_i - e_j)$, and $(e_j - e_i)$. The simple positive generators (i.e., the generators corresponding to simple positive roots) are

$$a_1^+ a_2,\ a_2^+ a_3,\ a_3^+ a_4,\ a_3^+ a_4^+. \tag{2.3}$$

The generators of the Cartan subalgebra (in the usual Cartan–Weyl basis) are

$$H_i = h_i - h_{i+1} \quad (1 = 1,2,3),$$
$$H_4 = h_3 + h_4. \tag{2.4}$$

The commutation relation can be easily worked out by Eq. (2.1) using

$$[AB,CD] = A\{B,C\}D - C\{A,D\}B + \{A,C\}DB - AC\{B,D\}. \tag{2.5}$$

They are ($i,j,k,l = 1,2,3,4$)

$$[h_i, h_j] = 0,$$
$$[h_i,a_j^+ a_k^+] = (\delta_{ij} + \delta_{ik})a_j^+ a_k^+,$$
$$[h_i,a_j^+ a_k] = (\delta_{ij} - \delta_{ik})a_j^+ a_k,$$
$$[a_i^+ a_j,a_k^+ a_l] = \delta_{jk}a_i^+ a_l - \delta_{il}a_k^+ a_j, \tag{2.6}$$
$$[a_i^+ a_j^+,a_k^+ a_l] = \delta_{il}a_j^+ a_k^+ - \delta_{jl}a_i^+ a_k^+,$$
$$[a_i^+ a_j^+,a_j a_i] = h_i + h_j,$$
$$[a_i^+ a_j,a_j^+ a_i] = h_i - h_j.$$

[a)] On leave of absence from Dipartimento di Scienze Fisiche, Universitá di Napoli, Italy.

 0022-2488/89/081674-05$02.50

The other commutation relations can be worked out by Hermitian conjugation.

The eight operators $a_i^\dagger, a_i$ transform as the 8_v representation of D_4, as can be easily seen from Eqs. (2.1) and (2.2). In particular, we have

$$[h_i, a_j^+] = \delta_{ij} a_j^+ \quad [h_i, a_j] = -\delta_{ij} a_j. \tag{2.7}$$

The highest weight (h.w.) is a_1^+.

To have a realization of the spinorial representations 8_s and 8_c in terms of operators, we introduce 16 fermionic operators labeled by a quadruple of numbers (b_1, b_2, b_3, b_4), where b_i $(i = 1,2,3,4)$ belongs to $\{-\frac{1}{2}, \frac{1}{2}\}$. We divide this set into two (eight-dimensional) subsets, denoted by α and β, that are specified, respectively, by an even (odd) number of b_i equal to $\frac{1}{2}$, and define a Hermitian operation:

$$(F^+_{b_1,b_2,b_3,b_4})^+ = F_{-b_1,-b_2,-b_3,-b_4}. \tag{2.8}$$

We can naturally identify the subset $\{\alpha\}$ with the 8_s and the subset $\{\beta\}$ with 8_c, as we shall see. We denote each element of these subsets by lower indices, written in increasing order, assuming a value between 1 and 4, and identifying the numbers b_i equal to $\frac{1}{2}$ in α^+, β^+.

So we have, e.g.,

$$\begin{aligned}
&\alpha_{1/2\ 1/2\ 1/2\ 1/2} = \alpha^+_{1234},\\
&\alpha_{-1/2\ -1/2\ -1/2\ -1/2} = \alpha_{1234},\\
&\alpha_{1/2\ -1/2\ 1/2\ -1/2} = \alpha^+_{13},\\
&\beta_{1/2\ 1/2\ 1/2\ -1/2} = \beta^+_{123},\\
&\beta_{1/2\ -1/2\ -1/2\ -1/2} = \beta^+_1.
\end{aligned} \tag{2.9}$$

These operators satisfy

$$\begin{aligned}
&\{\alpha^+, \alpha^+\} = \{\alpha, \alpha\} = \{\beta^+, \beta^+\} = \{\beta, \beta\} = 0,\\
&\{\alpha^+_{1234}, \alpha_{1234}\} = 1 \quad \{\alpha^+_{ij}, \alpha_{kl}\} = \delta_{ik}\delta_{jl},\\
&\{\beta_i^+, \beta_j\} = \delta_{ij}.
\end{aligned} \tag{2.10}$$

As we have

$$\begin{aligned}
&8_s \otimes 8_s = 35 \oplus 28 \oplus 1,\\
&8_c \otimes 8_c = 35' \oplus 28 \oplus 1.
\end{aligned} \tag{2.11}$$

We can also realize the algebra D_4 in terms of a bilinear of the form $\alpha^+\alpha$ or $\beta^+\beta$. Using Eq. (2.10), it is immediate to write a bilinear in α or β that satisfies the analogous relations of Eq. (2.6). We can make the following identifications:

$$\begin{aligned}
&a_i^+ a_j^+ \simeq \alpha^+_{1234}\alpha^+_{ij} \simeq \beta^+_{ijk}\beta^+_{ijl} \quad (k<l),\\
&a_i^+ a_j \simeq \alpha^+_{ik}\alpha^+_{il} \simeq \beta_i^+ \beta_j \quad (i<j, k<l),\\
&\alpha^+_{1234}\alpha_{1234} \simeq \frac{1}{2}\sum_{i=1}^{4} h_i,\\
&\alpha^+_{ij}\alpha_{ij} \simeq \frac{1}{2}\Big(h_i + h_j - \sum_{k\neq i\neq j} h_k\Big),\\
&\beta_i^+\beta_i \simeq \frac{1}{2}\Big(h_i - \sum_{k\neq i} h_k\Big).
\end{aligned} \tag{2.12}$$

It is convenient to introduce an overabundant notation, with the following relations (no sum over repeated indices):

$$\begin{aligned}
&\alpha_{ij} = \epsilon_{ijkl}\alpha^+_{kl},\\
&\beta_i = \epsilon_{ijkl}\beta^+_{jkl},
\end{aligned} \tag{2.13}$$

$$\begin{aligned}
&\alpha^+_{ij} = -\alpha^+_{ji} \quad (i<j),\\
&\beta^+_{ijk} = \eta\beta^+_{P(ijk)} \quad (i<j<k),
\end{aligned} \tag{2.14}$$

where η is the parity of the permutation P acting on the labels ijk. The a^+, α^+, β^+ are not independent due to the relation (triality property)

$$\begin{aligned}
&8_s \otimes 8_c = 56 \oplus 8_v,\\
&8_v \otimes 8_s = 56' \oplus 8_c,\\
&8_v \otimes 8_c = 56'' \oplus 8_s.
\end{aligned} \tag{2.15}$$

From consistency conditions on the transformation properties of a^+, α^+, β^+ under the algebra of D_4 written in the three forms a^+a, $\alpha^+\alpha$, and $\beta^+\beta$, we deduce the following $(j<k<l,\ i,j,k,l = 1,2,3,4)$:

$$\begin{aligned}
&[a_i^+, \alpha^+_{1234}] = 0, \quad [a_i, \alpha^+_{1234}] = \beta_i, \quad [a_i^+, \alpha^+_{jk}] = \beta^+_{ijk},\\
&[a_i, \alpha^+_{jk}] = \delta_{ij}\beta^+_k - \delta_{jk}\beta^+_j, \quad [a_i^+, \beta_j^+] = \alpha^+_{ij},\\
&[a_i^+, \beta_j] = \delta_{ij}\alpha^+_{1234}, \quad [a_i^+, \beta^+_{jkl}] = \alpha_{ijkl},\\
&[a_j, \beta^+_{jkl}] = \alpha^+_{kl}, \quad [\beta_i^+, \alpha^+_{jk}] = \epsilon_{ijkl}a_l,\\
&[\beta_i, \alpha^+_{jk}] = \delta_{ij}a_k^+ - \delta_{ik}a_j^+, \quad [\beta_i^+, \alpha_{1234}] = 0,\\
&[\beta_i, \alpha_{1234}] = \alpha_i
\end{aligned} \tag{2.16}$$

and the ones obtained by Hermitian conjugation. If two equal indices appear in α^+ or β^+ on the rhs of Eq. (2.16), then the commutation has to be read as vanishing. From Eq. (2.16) and the identity

$$[[A,B],C] = [A,[B,C]] + [[A,C],B], \tag{2.17}$$

one can compute the action of the algebra D_4 (in any of its realizations) on a^+, α^+, β^+, and the commutation relations that will be needed in the following sections. In particular, we have

$$\begin{aligned}
&[h_i, \alpha^+_{1234}] = \tfrac{1}{2}\alpha^+_{1234},\\
&[h_i, \alpha^+_{jk}] = (\delta_{ij} + \delta_{jk} - \tfrac{1}{2})\alpha^+_{jk},\\
&[h_i, \beta_j^+] = (\delta_{ij} - \tfrac{1}{2})\beta_j^+,\\
&[h_i, \beta^+_{jkl}] = (\delta_{ij} + \delta_{ik} + \delta_{il} - \tfrac{1}{2})\beta^+_{jkl}.
\end{aligned} \tag{2.18}$$

The h.w. of the representation 8_s (8_c) is α^+_{1234} (β^+_{123}).

For the needs of Secs. IV and V, it is convenient to study in some detail the following embedding in D_4,

$$D_4 \supset D_2 + D_2' \simeq A_{1,R} + A_{1,L} + A'_{1,R} + A'_{1,L}. \tag{2.19}$$

The adjoint and fundamental representations of D_4 decompose as follows:

$$\begin{aligned}
28 = {}&(3,1,1,1) + (1,3,1,1) + (1,1,3,1)\\
&+ (1,1,1,3) + (2,2,2,2)
\end{aligned} \tag{2.20}$$

in terms of a bilinear on a^+a,

$$a_1^+ a_2^+,\ a_2 a_1,\ h_1 + h_2, \tag{2.21a}$$

$$a_1^+ a_2,\ a_2^+ a_1,\ h_1 - h_2, \tag{2.21b}$$

$$a_3^+ a_4^+,\ a_4 a_3,\ h_3 + h_4, \tag{2.21c}$$

$$a_3^+ a_4,\ a_4^+ a_3,\ h_3 - h_4, \tag{2.21d}$$

$$a_1^+ a_3^+,\ a_3^+ a_2,\ a_2^+ a_3^+,\ a_1^+ a_4,\ a_1^+ a_4^+,$$
$$a_3^+ a_1,\ a_4 a_2,\ a_4^+ a_2,$$
$$a_2^+ a_4,\ a_2^+ a_4^+,\ a_1^+ a_3,\ a_4 a_1,\ a_4^+ a_1,$$
$$a_3 a_2,\ a_2^+ a_3,\ a_3 a_1, \tag{2.21e}$$
$$8_v = (2,2,1,1) + (1,1,2,2), \tag{2.22}$$
$$a_1^+,\ a_2,\ a_2^+,\ a_1, \tag{2.23a}$$
$$a_3^+,\ a_4,\ a_4^+,\ a_3, \tag{2.23b}$$
$$8_s = (1,2,1,2) + (2,1,2,1), \tag{2.24}$$
$$\alpha_{13}^+,\ \alpha_{23}^+,\ \alpha_{14}^+,\ \alpha_{24}^+, \tag{2.25a}$$
$$\alpha_{1234}^+,\ \alpha_{12},\ \alpha_{34},\ \alpha_{1234}, \tag{2.25b}$$
$$8_c = (1,2,2,1) + (2,1,1,2), \tag{2.26}$$
$$\beta_{134}^+,\ \beta_{234}^+,\ \beta_1^+,\ \beta_2^+, \tag{2.27a}$$
$$\beta_{123}^+,\ \beta_3^+,\ \beta_{124}^+,\ \beta_4^+. \tag{2.27b}$$

The states in Eqs. (2.21e), (2.23), (2.25), and (2.27) have been written (up to a factor) following the convention of applying the lowering operator $(J'_{-,L})^{K_1}(J'_{-,R})^{K_2}(J_{-,L})^{K_3}(J_{-,R})^{K_4}$ to the h.w. (first state on the left), where $K_1K_2K_3K_4$ is a naturally ordered quadruple of numbers ($K_i = 0,1$).

III. REALIZATION OF E_8

Now using the embedding (1.1) and the results of Sec. II, it is an easy task to write the algebra of E_8. We introduce for D_4' independent sets of fermionic operators a,α,β, which will be labeled by indices running from 5 to 8 and which satisfy Eq. (2.1), and among themselves, analogous relations of Eq. (2.16).

Explicitly we have (with slightly abusive, but simplified notation) ($i,j,k = 1,2,3,4$, $m,n,p = 5,6,7,8$, $i<j<k$, $m<n<p$)

$$a_i^+ a_j^+,\ a_i^+ a_j,\ h_i,\ \text{and H.c.}, \tag{3.1a}$$
$$a_m^+ a_n^+,\ a_m^+ a_n,\ h_m,\ \text{and H.c.}, \tag{3.1b}$$
$$a_i^+ a_m^+,\ a_i^+ a_m,\ \text{and H.c.}, \tag{3.1c}$$
$$\alpha_{1234}^+ \alpha_{mn}^+,\ \alpha_{ij}^+ \alpha_{5678}^+,\ \text{and H.c.},$$
$$\alpha_{1234}^+ \alpha_{5678}^+,\ \alpha_{1234}^+ \alpha_{5678},\ \text{and H.c.}, \tag{3.1d}$$
$$\alpha_{ij}^+ \alpha_{mn}^+,$$
$$\beta_{ijk}^+ \beta_{mnp}^+,\ \beta_{ijk}^+ \beta_m^+,\ \beta_i^+ \beta_{mnp}^+,\ \beta_i^+ \beta_m^+. \tag{3.1e}$$

Each set of operators in Eq. (3.1) corresponds to a term on the rhs of Eq. (1.2).

The simple generators are, with reference to the labeling of Fig. 1 ($r = 3,4,...,8$),

$$E_1 = \beta_1^+ \beta_8^+,\quad E_2 = a_1^+ a_2^+,\quad E_r = a_{r-1}^+ a_{r-2}, \tag{3.2a}$$

and the corresponding Cartan generators are

$$H_1 = \frac{1}{2}\Big(h_1 + h_8 - \sum_{s=2}^{7} h_s\Big),$$
$$H_2 = h_1 + h_2,\quad H_r = h_{r-1} - h_{r-2}. \tag{3.2b}$$

Let us remark that generators bilinear in $\{a\}$ correspond to roots of the forms $\pm e_i, \pm e_j$ ($i,j = 1,...,8$) (i.e., the roots of $D_8 \subset E_8$); generators bilinear in $\{\alpha\},\{\beta\}$ correspond to roots of the form

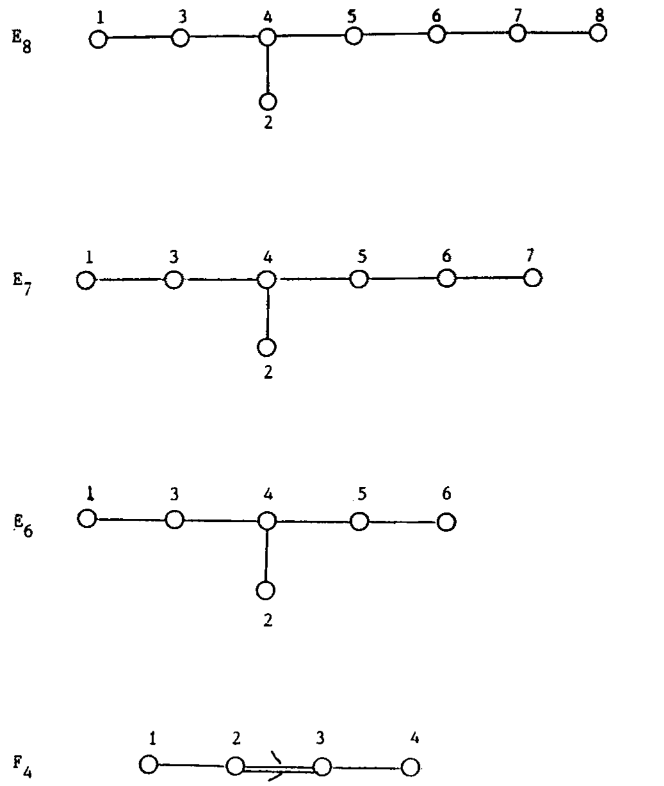


FIG. 1. Dynkin diagrams for the exceptional algebras with the labeling of simple positive roots.

$$\frac{1}{2}\sum_{i=1}^{\infty} e_i(-1)^{\nu(i)},\quad \sum \nu(i)\ \text{even},$$

which form the 128 representation of D_8.

IV. REALIZATION OF E_7

To obtain a realization of E_7 in terms of a bilinear in creation and annihilation operators, we consider the maximal embedding[5,6]

$$E_8 \supset E_7 + A_1 \tag{4.1}$$

and the corresponding decomposition

$$248 = (133,1) + (1,3) + (56,2). \tag{4.2}$$

We identify the algebra A_1 in Eq. (4.1) with the algebra

$$a_7^+ a_8^+,\ a_8 a_7,\ h_7 + h_8. \tag{4.3}$$

We make this choice in order to obtain the simple positive generators of E_7 as a subset of the simple positive generators of E_8.

We can identify easily the terms of the adjoint (dim 133) and fundamental (dim 56) representations of E_7 as the terms which are, respectively, singlets and doublets under Eq. (4.3) in Eq. (3.1). We obtain explicitly for the adjoint the generators of Eq. (3.1a) and ($i,j,k = 1,2,3,4$, $m,n = 5,6$, $p = 7,8$)

$$a_m^+ a_n^+, a_m^+ a_n, \text{ and H.c.},$$
$$h_m, h_n,$$
$$a_7^+ a_8, a_8^+ a_7, h_7 - h_8,$$
$$a_i^+ a_m^+, a_i^+ a_m, \text{ and H.c.}, \quad (4.4a)$$
$$\alpha_{1234}^+ \alpha_{mp}^+, \alpha_{1234}\alpha_{mp}^+, \alpha_{ij}^+ \alpha_{mp}^+, \quad (4.4b)$$
$$\beta_{ijk}^+ \beta_{56p}^+, \beta_{ijk}^+ \beta_p^+, \beta_i^+ \beta_{56p}^+, \beta_i^+ \beta_p^+, \quad (4.4c)$$

and for the fundamental

$$a_i^+ a_p^+, a_i^+ a_p, \text{ and H.c.},$$
$$a_m^+ a_p^+, a_m^+ a_p, \text{ and H.c.}, \quad (4.5a)$$
$$\alpha_{1234}^+ \alpha_{5678}^+, \alpha_{1234}^+ \alpha_{56}, \text{ and H.c.},$$
$$\alpha_{1234}\alpha_{5678}^+, \alpha_{1234}\alpha_{56}, \text{ and H.c.}, \quad (4.5b)$$
$$\alpha_{ij}^+ \alpha_{5678}^+, \alpha_{ij}^+ \alpha_{5678}, \alpha_{ij}^+ \alpha_{56}, \alpha_{ij}^+ \alpha_{78},$$
$$\beta_{ijk}^+ \beta_{m78}^+, \beta_{ijk}^+ \beta_m^+, \beta_i^+ \beta_{m78}^+, \beta_i^+ \beta_m^+. \quad (4.5c)$$

To write Eqs. (4.4) and (4.5), we used Eqs. (2.21e), (2.23), (2.25), and (2.27) with a four-units shift in the indices.

The simple generators are given by Eq. (3.2), omitting the value $r=8$.

V. REALIZATION OF E_6

The construction of E_6 is most easily obtained by the maximal embedding

$$E_7 \supset E_6 \times U(1) \quad (5.1)$$

and the corresponding decomposition of the adjoint representation of E_7,

$$133 = (78,0) + (1,0) + (27,-1) + (\overline{27},1). \quad (5.2)$$

The generator of U(1) in Eq. (5.1) is given by

$$\tfrac{1}{2}(2h_6 - h_7 + h_8). \quad (5.3)$$

By looking to the transformation properties of Eqs. (3.1a) and (4.4) under Eq. (5.3), we can identify the 78, 27, and $\overline{27}$.

The 78 is explicitly given by Eq. (3.1a) and

$$h_5, 2h_6 + h_7 - h_8,$$
$$a_i^+ a_5^+, a_i^+ a_5, \text{ and H.c.},$$
$$\alpha_{1234}^+ \alpha_{67}^+, \alpha_{1234}\alpha_{67}^+, \alpha_{ij}^+ \alpha_{67}^+, \quad (5.4)$$
$$\alpha_{1234}^+ \alpha_{58}^+, \alpha_{1234}\alpha_{58}^+, \alpha_{ij}^+ \alpha_{58}^+,$$
$$\beta_{ijk}^+ \beta_{567}^+, \beta_{ijk}^+ \beta_8^+, \beta_i^+ \beta_{567}^+, \beta_i^+ \beta_8^+.$$

The 27 is given by

$$a_5^+ a_6, a_5 a_6, a_7^+ a_8,$$
$$a_i^+ a_6, a_i a_6, \quad (5.5)$$
$$\alpha_{1234}^+ \alpha_{57}^+, \alpha_{1234}\alpha_{57}^+, \alpha_{ij}^+ \alpha_{57}^+,$$
$$\beta_{ijk}^+ \beta_7^+, \beta_i^+ \beta_7^+.$$

The simple generators are given by Eq. (3.2) omitting the indices $r=7,8$.

VI. REALIZATION OF F_4

To build up F_4, we use its maximal embedding in E_6. Moreover, we use the property that the Dynkin diagram of F_4 (see Fig. 1) can be obtained by that of E_6 by the following "folding"[7] associated to the outer automorphism τ in E_6 defined by

$$\tau(\alpha_i) = \alpha_i \quad (i=2,4), \qquad \tau(\alpha_3) = \alpha_5, \quad \tau(\alpha_1) = \alpha_6. \quad (6.1)$$

Labeling by a prime the simple (positive) roots of F_4, we have

$$\alpha_1' = \alpha_2, \quad \alpha_2' = \alpha_4,$$
$$\alpha_3' = \tfrac{1}{2}(\alpha_3 + \tau(\alpha_3)) = \tfrac{1}{2}(\alpha_3 + \alpha_5), \quad (6.2)$$
$$\alpha_4' = \tfrac{1}{2}(\alpha_1 + \tau(\alpha_1)) = \tfrac{1}{2}(\alpha_1 + \alpha_6).$$

So, for the simple positive and Cartan generators we obtain (omitting the prime)

$$E_1 = a_1^+ a_2^+, \quad E_2 = a_3^+ a_2,$$
$$E_3 = a_2^+ a_1 + a_4^+ a_3, \quad E_4 = \beta_1^+ \beta_8^+ + a_5^+ a_4,$$
$$H_1 = h_1 + h_2, \quad H_2 = h_3 - h_2, \quad (6.3)$$
$$H_3 = h_2 + h_4 - h_1 - h_3,$$
$$H_4 = \frac{1}{2}\Big(h_1 + h_8 + 2h_5 - 2h_4 - \sum_{m=2}^{7} h_m\Big).$$

The remaining 20 positive generators are given by

$$a_1^+ a_3^+, a_3^+ a_1 - a_4^+ a_2, a_3 a_5^+ - \beta_2^+ \beta_8^+,$$
$$a_2^+ a_3^+ + a_1^+ a_4^+, a_2 a_5^+ + \beta_3^+ \beta_8^+, a_1^+ a_5^+ - \beta_{123}^+ \beta_8^+,$$
$$a_4^+ a_1, a_2^+ a_4^+, a_1 a_5^+ - \beta_4^+ \beta_8^+, a_3^+ a_4^+,$$
$$a_2^+ a_5^+ - \beta_{124}^+ \beta_8^+, \alpha_{1234}\alpha_{58}^+, a_3^+ a_5^+ + \beta_{134}^+ \beta_8^+, \quad (6.4)$$
$$\alpha_{12}^+ \alpha_{58}^+, a_1^+ a_2 + \alpha_{13}^+ \alpha_{58}^+, \alpha_{13}^+ \alpha_{58}^+,$$
$$\alpha_{23}^+ \alpha_{58}^+ + \alpha_{14}^+ \alpha_{58}^+, \alpha_{24}^+ \alpha_{58}^+, \alpha_{34}^+ \alpha_{58}^+, \alpha_{1234}^+ \alpha_{58}^+.$$

The negative generators are found by Hermitian conjugation.

In order to write explicitly the fundamental (26 dim) representation of F_4, let us remark that in the embedding,

$$E_6 \supset F_4. \quad (6.5)$$

The 78 and 27 representations of E_6 decompose as

$$78 = 52 + 26, \quad (6.6)$$
$$27 = 26 + 1. \quad (6.7)$$

It is easy to identify the singlet of F_4 in the 27 given by Eq. (5.5); it is

$$\alpha_{14}^+ \alpha_{57}^+ - \alpha_{23}^+ \alpha_{57}^+ + \alpha_5 \alpha_6. \quad (6.8)$$

So the 26 is given by Eq. (5.5), neglecting the term (6.8).

VII. REALIZATION OF G_2

Finally, for completeness, we also write a realization of the (14 dim) algebra G_2. We can obtain it by a procedure analogous to the one followed in Sec. VI, starting from D_4 and considering the folding associated to the third-order automorphism of the Dynkin diagram of D_4. We obtain for the simple positive and Cartan generators ($1 \equiv$ long root, $2 \equiv$ short root)

$$E_1 = a_2^+ a_3, \quad E_2 = a_1^+ a_2 + a_3^+ a_4 + a_3^+ a_4^+, \quad H_1 = h_2 - h_3, \quad H_2 = h_1 - h_2 + 2h_3. \tag{7.1}$$

The remaining four positive generators are

$$a_1^+ a_3 - a_2^+ a_4 - a_2^+ a_4^+, a_1^+ a_4 + a_1^+ a_4^+ + a_2^+ a_3^+, a_1^+ a_3^+, a_1^+ a_2^+. \tag{7.2}$$

In the embedding,

$$D_4 \supset G_2. \tag{7.3}$$

The adjoint and fundamental vectorial representations of D_4 decompose as

$$28 = 14 + 7 + 7, \tag{7.4}$$

$$8 = 7 + 1. \tag{7.5}$$

Thus we can realize the 7 dim fundamental representation in three different forms, the h.w. being, respectively,

$$a_1^+ a_4 - a_1^+ a_4^+, a_2^+ a_3^+ - a_1^+ a_4^+, a_1^+. \tag{7.6}$$

VIII. CONCLUSIONS

We have presented a realization of the exceptional algebras and of their fundamental representations in terms of a bilinear in two independent sets of three (not mutually independent) octuples of fermionic operators, transforming as the three fundamental representations of D_4. The triality property has played an essential role in this construction.

Let us remark that we have obtained for the fundamental representations of exceptional algebras (except G_2) expressions that are bilinear in fermionic fields.

From the obtained results, one could discuss the embedding of classical algebras, in the same spirit of Ref. 6. Just to give an example, we discuss the physically relevant embedding

$$E_7 \supset D_6 \times A_1 \supset D_5 \times A_1 \times U(1). \tag{8.1}$$

The 133 decomposes as

$$\begin{aligned} 133 &= (66,1) + (1,3) + (32,2) \\ &= (45,1,0) + (1,1,0) + (10,1,1) + (10,1,-1) \\ &\quad + (1,3,0) + (16,2,-\tfrac{1}{2}) + (16,2,\tfrac{1}{2}), \end{aligned} \tag{8.2}$$

and we obtain $(g,h,i = 1,2,3,4)$ $(j,k = 1,2,3,4,5)$ $(p = 7,8)$

$$(45,1,0) \equiv a_j^+ a_k^+, a_j^+ a_k, h_j, \text{ and H.c.}, \tag{8.3a}$$

$$(1,1,0) \equiv h_6, \tag{8.3b}$$

$$(10,1,1) \equiv a_j^+ a_6^+, a_j a_6^+, \tag{8.3c}$$

$$(10,1,-1) \equiv a_j^+ a_6, a_j a_6, \tag{8.3d}$$

$$(1,3,0) \equiv a_7^+ a_8, a_8^+ a_7, h_7 - h_8, \tag{8.3e}$$

$$(16,2,-\tfrac{1}{2}) \equiv \alpha_{1234}^+ \alpha_{5p}^+, \alpha_{1234} \alpha_{5p}^+, \alpha_{hi}^+ \alpha_{5p}^+, \beta_{ghi}^+ \beta_p^+, \beta_i^+ \beta_p^+, \tag{8.3f}$$

$$(\overline{16},2,\tfrac{1}{2}) \equiv \alpha_{1234}^+ \alpha_{6p}^+, \alpha_{1234} \alpha_{6p}^+, \alpha_{hi}^+ \alpha_{6p}^+, \beta_{ghi}^+ \beta_{56p}^+, \beta_i^+ \beta_{56p}^+. \tag{8.3g}$$

The (66,1) is formed by Eqs. (8.3a)–(8.3d) and (32,2) by Eqs. (8.3f) and (8.3g).

In this way, we have obtained [Eqs. (8.3f) and (8.3g)] an expression for the spinorial representation of D_5. However, one could generalize the idea of fermionic operators of the type introduced in Sec. II for D_4 to describe fermionic representations of any orthogonal B or D algebra, which could be interesting in several contexts, e.g., to obtain a realization of the exceptional superalgebra F(4).

ACKNOWLEDGMENTS

I would like to thank P. Winternitz and the CRM for their hospitality and J. Patera for useful discussions.

[1] M. B. Green and J. H. Schwarz, Phys. Lett. B **149**, 112 (1984).

[2] W. Lerche, D. Lüst, and A. N. Schellekens, Nucl. Phys. B **287**, 477 (1987).

[3] J. Lauer, D. Lüst, and S. Theisen, CERN Report No. CERN-TH 5211/88, where reference to previous works can be found.

[4] B. O. Wybourne, *Classical Groups for Physicists* (Wiley–Interscience, New York, 1971).

[5] R. Slansky, Phys. Rev. **79**, 1 (1981).

[6] F. Buccella, A. Della Selva, and A. Sciarrino, J. Math. Phys. **30**, 585 (1989).

[7] P. Goddard, W. Nahm, D. Olive, and A. Schwimmer, Commun. Math. Phys. **107**, 179 (1986).

The constraints of potentials and the finite-dimensional integrable systems

Yunbo Zeng and Yishen Li
Department of Mathematics, University of Science and Technology of China, Hefei, .
People's Republic of China

(Received 6 September 1988; accepted for publication 8 March 1989)

Restricting potential to the space spanned by the eigenvectors of the recursion operator leads to a natural constraint of potential and a finite-dimensional integrable Hamiltonian system. The general method for proving the consistency of the two systems stemming from the Lax pair and obtaining the constants of the motion for the Hamiltonian system is illustrated by the classical Boussinesq and AKNS hierarchies. By using gauge transformation, similar results for the Jaulent–Miodek and Kaup–Newell hierarchies are presented.

I. INTRODUCTION

It is known that many finite-dimensional integrable Hamiltonian systems can be obtained by restricting infinite-dimensional integrable Hamiltonian systems to finite-dimensional invariant submanifolds of their phase space (see, for example, Refs. 1–5). Imposing some boundary condition, some finite-dimensional invariant submanifolds of the phase space can be obtained from the variational approach. Consider an integrable evolution equation

$$u_t = K(u)\,, \tag{1.1}$$

where u is supposed to tend to zero rapidly as $|x|$ tends to ∞ or is periodic in x. Let $F_0(u)$, ..., $F_N(u)$ be differentiable conserved functionals for the flow (1.1) and denote the gradient of F_j with respect to the L_2 scalar product by $G_{F_j}(u)$; then it is shown by Gardner *et al.*[6] and Lax[7] that solutions of

$$G_{F_0}(u) - \sum_{j=1}^{N} G_{F_j}(u) = 0 \tag{1.2}$$

form an invariant set for the flow (1.1). If the eigenvalue λ of the eigenvalue problem in the Lax pair associated with (1.1) is a conserved functional under the flow (1.1), from (1.2) we can obtain that

$$G_{F_0}(u) - \sum_{j=1}^{N} G_{\lambda_j} = 0 \tag{1.3}$$

determines an invariant submanifold corresponding to (1.1). We can conclude that the two systems obtained from the Lax pair under the constraint condition (1.3) are consistent. For example, consider the Korteweg–deVries (KdV) hierarchy

$$u_t = DL^n u, \quad n = 1,2,\dots\,, \tag{1.4}$$

where L is called the recursion operator defined by

$$L = -\frac{1}{4}D^2 - u + \frac{1}{2}\int u_x\,dx\,.$$

The eigenvalue λ of the Schrödinger eigenvalue problem

$$\phi_{xx} + u\phi = \lambda\phi \tag{1.5}$$

is a conserved functional under the flows (1.4) and $G_\lambda = \phi^2$ (see Refs. 6 and 7). Take $F_0(u)$ to be the first conserved functional for (1.4), that is, $F_0(u) = \int u\,dx$; from (1.3) and (1.5) we obtain that

$$\sum_{j=1}^{N} \phi_j^2 = 1\,,$$

$$u = \sum_{j=1}^{N} (\lambda_j\phi_j^2 + \phi_{jx}^2)\,,$$

which leads to the well-known Neumann system (see Refs. 1 and 2). If we take $F_0(u) = \frac{1}{2}\int u^2\,dx$, the second conserved functional for (1.4), we have that

$$u = \sum_{j=1}^{N} \phi_j^2 \tag{1.6}$$

is an invariant submanifold under the whole hierarchy of flows (1.4). This implies that the system of Neumann type obtained from (1.5) and the system obtained from the time part of the Lax pair under the constraint condition (1.6), respectively, are consistent.

However, if u is required neither to tend to zero as $|x|$ tends to ∞ nor to be periodic in x, we cannot start from (1.2). The constraint condition (1.3) and the consistency of the two systems mentioned above cannot follow from the variational approach either. In this general case, we want to show that the system of Neumann type can be obtained in a straightforward way by restricting the integrable evolution equations to an invariant subspace of the recursion operator. Based on the work by Cao[8] we propose a general method to prove the consistency of the two systems without any boundary condition and to present the constants of the motion for the system of Neumann type.

In Sec. II the main idea is illustrated by the hierarchy of the classical Boussinesq equations.[9] Restricting the potential to the linear space spanned by the eigenvectors of the recursion operator leads to a natural constraint condition on the potential and ensures this space to be left invariant by the recursion operator. This allows us to prove the consistency of the two systems for the whole hierarchy and to construct the constants of the motion for a system of Neumann type. We believe that these constants are in involution and the system is an integrable Hamiltonian system. Also, the solution of this system is shown to satisfy a certain stationary equation in the hierarchy. In Sec. III the system of Neumann type is shown to possess the Painlevé property,[10] which implies that the system is integrable. Then the method is applied to AKNS hierarchy[10] in Sec. IV.

In Sec. V, we want to point out that using the gauge

 0022-2488/89/081679-11$02.50

transformation, which transforms an eigenvalue problem to another one, similar results for the later hierarchy can be deduced from those for the former hierarchy; the transform relation between these two systems of Neumann type can be established easily. For example, similar results for the Jaulent–Miodek[11] and Kaup–Newell[12] hierarchies are obtained in this way.

II. THE GENERAL METHOD

We now illustrate the main idea by the classical Boussinesq hierarchy[9]

$$u_t = L^n u_x = \begin{pmatrix} Q_{nx} \\ R_{nx} \end{pmatrix}, \quad u = \begin{pmatrix} q \\ r \end{pmatrix}, \tag{2.1}$$

with

$$L = \begin{pmatrix} \frac{1}{2}r & -\frac{1}{4}D^2 + q + \frac{1}{2}q_x D^{-1} \\ 1 & \frac{1}{2}(r + r_x D^{-1}) \end{pmatrix},$$

$$D = \frac{d}{dx}, \quad D^{-1}D = DD^{-1} = 1 .$$

It is known that Q_n and R_n are polynomials of q, r, and their derivatives. Equation (2.1) is associated with the eigenvalue problem

$$\phi_{xx} = M\phi, \quad M = -\zeta^2 + q - \tfrac{1}{4}r^2 + \zeta r \tag{2.2}$$

and the evolution equation of ϕ,

$$\phi_t = -\tfrac{1}{2}B_x\phi + B\phi_x , \tag{2.3}$$

where

$$B = \sum_{k=0}^{n} b_k \zeta^{n-k}, \tag{2.4}$$

with the definition

$$b_0 = 1, \quad b_k = \frac{1}{2}R_{k-1}, \quad \begin{pmatrix} Q_{kx} \\ R_{kx} \end{pmatrix} = L^k \begin{pmatrix} q_x \\ r_x \end{pmatrix}. \tag{2.5}$$

The solvability condition of (2.2) and (2.3) reads as

$$\phi_{xxt} - \phi_{txx} = [M_t + \tfrac{1}{2}B_{xxx} - 2B_x M - BM_x]\phi = 0 . \tag{2.6}$$

Inserting (2.4) and (2.5) into (2.6) gives

$$\begin{aligned} M_t &+ \tfrac{1}{2}B_{xxx} - 2B_x M - BM_x \\ &= q_t - Q_{nx} + (\zeta - \tfrac{1}{2}r)(r_t - R_{nx}) \end{aligned} \tag{2.7}$$

which, together with (2.6), obtains (2.1). Indeed, (2.1) can be deduced from (2.2) and (2.3) without requiring any boundary condition at infinity for q and r. Here we define the integral constant of D^{-1} appearing in L to be zero. Thus, for example, the first equation in the hierarchy (2.1) reads as

$$\begin{pmatrix} q \\ r \end{pmatrix}_t = \begin{pmatrix} Q_{1x} \\ R_{1x} \end{pmatrix}, \quad \begin{pmatrix} Q_1 \\ R_1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{4}r_{xx} + qr \\ q + \frac{1}{2}r^2 \end{pmatrix}.$$

First we need to find the "eigenvector" for L. Define

$$L_0 = \begin{pmatrix} \frac{1}{2}r & -\frac{1}{4}D^2 + q + \frac{1}{2}q_x I_0 \\ 1 & \frac{1}{2}r + \frac{1}{2}r_x I_0 \end{pmatrix}, \quad I_0 = \int_{x_0}^{x} \cdot dy ,$$

where x_0 is a fixed arbitrary constant. It is easy to check that if ϕ satisfies (2.2), then

$$L_0\Psi_x = \zeta\Psi_x + eu_x , \tag{2.8}$$

where

$$\Psi = \begin{pmatrix} (\zeta - \frac{1}{2}r)\phi^2 \\ \phi^2 \end{pmatrix}, \quad e = -\tfrac{1}{2}\phi^2|_{x=x_0} .$$

We now consider the following system intead of (2.2):

$$\phi_{jxx} = M_j\phi_j, \quad M_j = -\zeta_j^2 + q - \tfrac{1}{4}r^2 + \zeta_j r, \quad j = 1,...,N, \tag{2.9}$$

where $\zeta_j \neq \zeta_l$ when $j \neq 1$. If $\Phi = (\phi_1,...,\phi_N)$ satisfies (2.9), from (2.8) one obtains

$$L_0\Psi_{jx} = \zeta_j\Psi_{jx} + e_j u_x , \tag{2.10a}$$

where

$$\Psi_j = \begin{pmatrix} (\zeta_j - \frac{1}{2}r)\phi_j^2 \\ \phi_j^2 \end{pmatrix}, \quad e_j = -\tfrac{1}{2}\phi_j^2|_{x=x_0} . \tag{2.10b}$$

Let $\widetilde{H}$ denote the linear subspace of the domain of L_0 spanned by $\{\Psi_{1x},...,\Psi_{Nx}\}$. If we restrict u_x to $\widetilde{H}$, that is,

$$u = \sum_{l=1}^{N} \alpha_l \Psi_l \equiv f(\Phi), \tag{2.11}$$

then it is clear that $\widetilde{H}$ is left invariant by L_0. This property allows us to prove all the results for whole hierarchy (2.1).

Under the constraint condition $u = f(\Phi)$, we have the following systems from (2.2) and (2.3):

$$\phi_{jxx} = \overline{M}_j\phi_j, \quad \overline{M}_j = M_j|_A, \quad j = 1,...,N \tag{2.12}$$

and

$$\phi_{jt} = -\tfrac{1}{2}\overline{B}_{jx}\phi_j + \overline{B}_j\phi_{jx}, \quad j = 1,...,N, \tag{2.13}$$

$$\overline{B}_j = \sum_{k=0}^{n} b_k \zeta_j^{n-k}\Big|_A ,$$

where the subscript A means to substitute $u = f(\Phi)$ into the expression. Since (2.12) and (2.13) can be changed into similar systems by setting $\tilde{\phi}_j = \sqrt{\alpha_j}\phi_j$, $\widetilde{\Psi}_j = \alpha_j\Psi_j$, and $u = \Sigma_{j=1}^{N}\widetilde{\Psi}_j$, one can take $\alpha_l = 1$ without loss of generality. Then (2.11) becomes

$$u = f(\Phi) = \sum_{j=1}^{N} \Psi_j, \quad f(\Phi) = \begin{pmatrix} f_1(\Phi) \\ f_2(\Phi) \end{pmatrix}, \tag{2.14a}$$

which is equivalent to

$$\begin{aligned} r &= f_2(\Phi) = \sum_{l=1}^{N} \phi_l^2 , \\ q &= f_1(\Phi) = \sum_{l=1}^{N} \zeta_l\phi_l^2 - \frac{1}{2}\Big(\sum_{l=1}^{N} \phi_l^2\Big)^2 , \end{aligned} \tag{2.14b}$$

and (2.12) reads as

$$\phi_{jxx} = \Big[-\zeta_j^2 + \sum_{l=1}^{N} \zeta_l\phi_l^2 - \frac{3}{4}\Big(\sum_{l=1}^{N} \phi_l^2\Big)^2 + \zeta_j \sum_{l=1}^{N} \phi_l^2\Big]\phi_j ,$$

$j = 1,...,N$.

We will show that under the constraint condition (2.14), (2.12) and (2.13) are naturally consistent. This ensures that the submanifold of the phase space

$$\overline{H} = \Big\{u = \sum_{j=1}^{N} \Psi_j \,\Big|\, \Phi \text{ satisfies } (2.12)\Big\}$$

is left invariant by all the flows in the hierarchy (2.1). Since system (2.12) is obtained by restricting the infinite-dimensional integrable Hamiltonian system to the finite-dimensional invariant submanifold of its phase space, it is supposed[5] to be an integrable system. Indeed, we will show in Sec. III that system (2.12) possesses the Painlevé property.

Following the procedure applied to the KdV equation,[8] we now propose a general method to prove the consistency of (2.12) and (2.13) and give the constants of the motion for (2.12).

Lemma 2.1: Let Φ satisfy (2.13), $F=(F_1,\ldots,F_N)^T$, and

$$F_j=\phi_{jxx}-M_j|_A\phi_j\,. \tag{2.15}$$

Then

$$F_{jt}=\overline{B}_jF_{jx}+\tfrac{3}{2}\overline{B}_{jx}F_j-\phi_j[q_t-Q_{nx}+(\zeta_j-\tfrac{1}{2}r)(r_t-R_{nx})]|_A\,. \tag{2.16}$$

Proof: By (2.13) and (2.15), it is found that

$$\begin{aligned}F_{jt}&=(\phi_{jxxt}-M_{jt}\phi_j-M_j\phi_{jt})|_A\\&=[-\tfrac{1}{2}B_{jxxx}\phi_j+\tfrac{3}{2}B_{jx}\phi_{jxx}+B_j\phi_{jxxx}-M_{jt}\phi_j\\&\quad+\tfrac{1}{2}B_{jx}M_j\phi_j-M_jB_j\phi_{jx}]|_A\\&=\overline{B}_jF_{jx}+\tfrac{3}{2}\overline{B}_{jx}F_j-\phi_j[M_{jt}+\tfrac{1}{2}B_{jxxx}\\&\quad-2B_{jx}M_j-B_jM_{jx}]|_A\,,\end{aligned}$$

which leads to (2.16) by using (2.7).

Theorem 2.1: Let $\Phi(x,t)$ be a solution to (2.13) and $\Phi(x,0)$ be a solution to (2.12). Then $\Phi(x,t)$satisfies (2.12) and $u=f(\Phi)$ satisfies (2.1).

Proof: Using (2.15), a direct calculation gives

$$L_0\Psi_{jx}=\zeta_j\Psi_{jx}+e_ju_x+G_j\,,$$

where

$$G_j=\begin{pmatrix}G_{1j}\\G_{2j}\end{pmatrix}=\begin{pmatrix}-\tfrac{1}{2}F_{jx}\phi_j-\tfrac{3}{2}F_j\phi_{jx}\\0\end{pmatrix},\quad e_j=-\tfrac{1}{2}\phi_j^2|_{x=x_0}\,.$$

Then it is easy to show by induction that

$$\begin{aligned}L_0u_x|_A&=\sum_{j=1}^N\zeta_j\Psi_{jx}+\alpha_1\sum_{j=1}^N\Psi_{jx}+G^{(1)},\\L_0^ku_x|_A&=\sum_{j=1}^N\Psi_{jx}\sum_{m=0}^k\alpha_m\zeta_j^{k-m}+G^{(k)},\end{aligned} \tag{2.17}$$

where

$$\alpha_0=1,\quad \alpha_m=\sum_{l=0}^{m-1}\alpha_l\delta_{m-l}\,, \tag{2.18a}$$

$$\delta_0=1,\quad \delta_l=\sum_{j=1}^N\zeta_j^{l-1}e_j\,, \tag{2.18b}$$

$$G^{(k)}=\sum_{j=1}^NG_j\sum_{m=0}^{k-1}\alpha_m\zeta_j^{k-m-1}+L_0G^{(k-1)}\,. \tag{2.18c}$$

Using the identity

$$L\begin{pmatrix}Q_{kx}\\R_{kx}\end{pmatrix}=L_0\begin{pmatrix}Q_{kx}\\R_{kx}\end{pmatrix}+\beta_{k+1}\begin{pmatrix}q_x\\r_x\end{pmatrix},\quad \beta_{k+1}=\frac{1}{2}R_k\Big|_{x=x_0}, \tag{2.19}$$

we have

$$\begin{aligned}Lu_x&=L_0u_x+\beta_1u_x,\quad \beta_1=\tfrac{1}{2}r|_{x=x_0}\,,\\L^ku_x&=\sum_{m=0}^k\beta_mL_0^{k-m}u_x\,,\end{aligned} \tag{2.20}$$

where

$$\beta_0=1,\quad \beta_m=\tfrac{1}{2}R_{m-1}|_{x=x_0}\,.$$

By (2.17) and (2.20), one can deduce that

$$\begin{aligned}L^ku_x|_A&=\sum_{l=0}^k\beta_{k-l}L_0^lu_x\Big|_A\\&=\Big[\sum_{l=0}^k\beta_{k-l}\sum_{j=1}^N\Psi_{jx}\sum_{m=0}^l\alpha_m\zeta_j^{l-m}\\&\quad+\sum_{l=0}^k\beta_{k-l}G^{(l)}\Big]\Big|_A\\&=\sum_{j=1}^N\Psi_{jx}\sum_{l=0}^kC_l\zeta_j^{k-l}+\overline{G}^{(k)}\,,\end{aligned} \tag{2.21a}$$

where

$$C_0=1,\quad C_l=\sum_{m=0}^l\alpha_m\beta_{l-m}\Big|_A,\quad \overline{G}^{(k)}=\sum_{l=0}^k\beta_{k-l}G^{(l)}\,. \tag{2.21b}$$

Equation (2.21a) gives

$$R_{kx}|_A=\sum_{j=1}^N(\phi_j^2)_x\sum_{l=0}^kC_l\zeta_j^{k-l}+\overline{G}_2^{(k)}\,, \tag{2.22a}$$

$$Q_{kx}|_A=\sum_{j=1}^N\Big[\Big(\zeta_j-\frac{1}{2}\sum_{l=1}^N\phi_l^2\Big)\phi_j^2\Big]_x\sum_{l=0}^kC_l\zeta_j^{k-l}+\overline{G}_1^{(k)}\,, \tag{2.22b}$$

which lead to

$$\begin{aligned}R_k|_A&=\sum_{j=1}^N\phi_j^2\sum_{l=0}^kC_l\zeta_j^{k-l}+I_0\overline{G}_2^{(k)}+\widetilde{C}_{k+1}\,,\\Q_k|_A&=\sum_{j=1}^N\phi_j^2\sum_{l=0}^kC_l\zeta_j^{k-l}\Big(\zeta_j-\frac{1}{2}\sum_{m=1}^N\phi_m^2\Big)\\&\quad+I_0\overline{G}_1^{(k)}+\widetilde{C}_{k+2}\,.\end{aligned}$$

By (2.10b), (2.18), (2.19), and (2.21b), it is found that

$$\begin{aligned}\widetilde{C}_{k+1}&=\Big[R_k\Big|_A-\sum_{j=1}^N\phi_j^2\sum_{l=0}^kC_l\zeta_j^{k-l}\Big]\Big|_{x=x_0}\\&=2\beta_{k+1}+2\sum_{l=0}^kC_l\delta_{k-l+1}=2\beta_{k+1}+2\sum_{l=0}^k\sum_{m=0}^l\alpha_m\beta_{l-m}\delta_{k-l+1}\\&=2\beta_{k+1}+2\sum_{l=0}^k\beta_{k-l}\sum_{m=0}^l\alpha_m\delta_{l-m+1}=2\beta_{k+1}+2\sum_{l=0}^k\alpha_{l+1}\beta_{k-l}\\&=2C_{k+1}\,.\end{aligned}$$

In the same way, we have $\widetilde{C}_{k+2}=2C_{k+2}$ by using $Q_{kx}=R_{k+1,x}-\frac{1}{2}(rR_k)_x$. Hence

$$R_k|_A=2b_{k+1}=\sum_{j=1}^{N}\phi_j^2\sum_{l=0}^{k}C_l\zeta_j^{k-l}+2C_{k+1}+I_0\overline{G}_2^{(k)}, \tag{2.23a}$$

$$Q_k|_A=\sum_{j=1}^{N}\phi_j^2\sum_{l=0}^{k}C_l\zeta_j^{k-l}\Big(\zeta_j-\frac{1}{2}\sum_{m=1}^{N}\phi_m^2\Big)+2C_{k+2}+I_0\overline{G}_1^{(k)}. \tag{2.23b}$$

From (2.13), (2.14a), and (2.23a) we obtain

$$\begin{aligned}
r_t|_A&=2\sum_{m=1}^{N}\phi_m\sum_{k=1}^{n}\zeta_m^{n-k}\Big(-\frac{1}{2}b_{kx}\phi_m+b_k\phi_{mx}\Big)+\sum_{m=1}^{N}(\phi_m^2)_x\zeta_m^n\\
&=\sum_{m=1}^{N}\sum_{k=1}^{n}\zeta_m^{n-k}\phi_m\Big[-\frac{1}{2}\phi_m\sum_{j=1}^{N}(\phi_j^2)_x\sum_{l=0}^{k-1}C_l\zeta_j^{k-l-1}-\frac{1}{2}\phi_m\overline{G}_2^{(k-1)}\\
&\quad+\phi_{mx}\sum_{j=1}^{N}\phi_j^2\sum_{l=0}^{k-1}C_l\zeta_j^{k-l-1}+2C_k\phi_{mx}+\phi_{mx}I_0\overline{G}_2^{(k-1)}\Big]+\sum_{m=1}^{N}\zeta_m^n(\phi_m^2)_x\\
&=\sum_{m=1}^{N}\sum_{j=1}^{N}\phi_m\phi_{mx}\phi_j^2\sum_{k=1}^{n}\sum_{l=0}^{k-1}C_l(\zeta_m^{n-k}\zeta_j^{k-l-1}-\zeta_j^{n-k}\zeta_m^{k-l-1})\\
&\quad+\sum_{m=1}^{N}(\phi_m^2)_x\zeta_m^n+\sum_{m=1}^{N}(\phi_m^2)_x\sum_{k=1}^{n}C_k\zeta_m^{n-k}+\overline{E}_2^{(n)}\\
&=\sum_{m=1}^{N}(\phi_m^2)_x\sum_{k=0}^{n}C_k\zeta_m^{n-k}+\overline{E}_2^{(n)},
\end{aligned}$$

with

$$\overline{E}_2^{(n)}=\sum_{m=1}^{N}\sum_{k=1}^{n}\zeta_m^{n-k}\phi_m\Big[-\frac{1}{2}\phi_m\overline{G}_2^{(k-1)}+\phi_{mx}I_0\overline{G}_2^{(k-1)}\Big],$$

where we use the identity

$$\sum_{k=1}^{n}\sum_{l=0}^{k-1}C_l(\zeta_m^{n-k}\zeta_j^{k-l-1}-\zeta_j^{n-k}\zeta_m^{k-l-1})=0. \tag{2.24}$$

Hence

$$(r_t-R_{nx})|_A=\overline{E}_2^{(n)}-\overline{G}_2^{(n)}\equiv E_2^{(n)}. \tag{2.25}$$

Similarly, notice that

$$\sum_{k=1}^{n}\sum_{l=0}^{k-1}C_l(\zeta_m^{n-k+1}\zeta_j^{k-l-1}-\zeta_j^{n-k+1}\zeta_m^{k-l-1})=\sum_{l=0}^{n}C_l(\zeta_m^{n-l}-\zeta_j^{n-l}) \tag{2.26}$$

and we have

$$\begin{aligned}
2\sum_{m=1}^{N}\zeta_m\phi_m\phi_{mt}&=\sum_{m=1}^{N}\sum_{j=1}^{N}\phi_m\phi_{mx}\phi_j^2\sum_{k=1}^{n}\sum_{l=0}^{k-1}C_l(\zeta_m^{n-k+1}\zeta_j^{k-l-1}-\zeta_j^{n-k+1}\zeta_m^{k-l-1})\\
&\quad+\sum_{m=1}^{N}(\phi_m^2)_x\sum_{k=0}^{n}C_k\zeta_m^{n-k+1}+\overline{E}_1^{(n)}\\
&=\sum_{m=1}^{N}\sum_{j=1}^{N}\phi_m\phi_{mx}\phi_j^2\sum_{l=0}^{n}C_l(\zeta_m^{n-l}-\zeta_j^{n-l})+\sum_{m=1}^{N}(\phi_m^2)_x\sum_{k=0}^{n}C_k\zeta_m^{n-k+1}+\overline{E}_1^{(n)}\\
&=r\sum_{m=1}^{N}\phi_m\phi_{mx}\sum_{l=0}^{n}C_l\zeta_m^{n-l}-\frac{1}{2}r_x\sum_{j=1}^{N}\phi_j^2\sum_{l=0}^{n}C_l\zeta_j^{n-l}\\
&\quad+\sum_{m=1}^{N}(\phi_m^2)_x\sum_{k=0}^{n}C_k\zeta_m^{n-k+1}+\overline{E}_1^{(n)},
\end{aligned} \tag{2.27}$$

where

$$\overline{E}_1^{(n)}=\sum_{m=1}^{N}\sum_{k=1}^{n}\phi_m\zeta_m^{n-k+1}\Big[-\frac{1}{2}\phi_m\overline{G}_2^{(k-1)}+\phi_{mx}I_0\overline{G}_2^{(k-1)}\Big].$$

Thus

$$
\begin{aligned}
(q_t - Q_{nx})|_A &= 2\sum_{m=1}^{N}\zeta_m\phi_m\phi_{mt} - 2r\sum_{m=1}^{N}\phi_m\phi_{mt} - Q_{nx}|_A \\
&= \sum_{m=1}^{N}(\phi_m^2)_x\sum_{k=0}^{n}C_k\zeta_m^{n-k+1} + \frac{1}{2}r\sum_{m=1}^{N}(\phi_m^2)_x\sum_{l=0}^{n}C_l\zeta_m^{n-l} - \frac{1}{2}r_x\sum_{j=1}^{N}\phi_j^2\sum_{l=0}^{n}C_l\zeta_j^{n-l} \\
&\quad + \overline{E}_1^{(n)} - r\sum_{m=1}^{N}(\phi_m^2)_x\sum_{k=0}^{n}C_k\zeta_m^{n-k} - r\overline{E}_2^{(n)} - \sum_{j=1}^{N}\left[\left(\zeta_j - \frac{1}{2}r\right)\phi_j^2\right]_x\sum_{l=0}^{n}C_l\zeta_j^{n-l} - \overline{G}_1^{(n)} \\
&= \overline{E}_1^{(n)} - \sum_{j=1}^{N}\phi_j^2\overline{E}_2^{(n)} - \overline{G}_1^{(n)} \equiv E_1^{(n)}. \qquad (2.28)
\end{aligned}
$$

Observe that each term of $E_1^{(n)}$ and $E_2^{(n)}$ contains at least one component of $F, F_x, \ldots$ and

$$
\left.\frac{\partial^m E^{(n)}}{\partial x^m}\right|_{F\equiv 0} \equiv 0, \quad E^{(n)} = \begin{pmatrix} E_1^{(n)} \\ E_2^{(n)} \end{pmatrix}, \quad m = 0,1,\ldots. \qquad (2.29)
$$

Using (2.25) and (2.28), (2.16) becomes

$$
\begin{aligned}
F_{jt} &= \overline{B}_j F_{jx} + \frac{3}{2}\overline{B}_{jx}F_j \\
&\quad - \phi_j\left[E_1^{(n)} + \left(\zeta_j - \frac{1}{2}\sum_{l=1}^{N}\phi_l^2\right)E_2^{(n)}\right]. \qquad (2.30)
\end{aligned}
$$

Since $\Phi(x,0)$ satisfies (2.12), we have

$$F(x,0) \equiv 0$$

and

$$\left.\frac{\partial^m F}{\partial x^m}\right|_{t=0} \equiv 0, \quad m = 0,1,\ldots$$

which, together with (2.29) and (2.30), yields

$$\left.\frac{\partial^{m+1} F}{\partial x^m\,\partial t}\right|_{t=0} \equiv 0, \quad m = 0,1,\ldots.$$

It is easy to show by induction that

$$\left.\frac{\partial^{m+l} F}{\partial x^m\,\partial t^l}\right|_{t=0} \equiv 0, \quad m,l = 0,1,\ldots,$$

which implies

$$F(x,t) \equiv 0.$$

Then we complete the proof by using (2.15), (2.25), and (2.28).

Remark: Theorem 2.1 shows that the submanifold of the phase space

$$\overline{H} = \{u = f(\Phi)\,|\,\Phi \text{ satisfies } (2.12)\}$$

is left invariant by all the flows in the hierarchy (2.1).

Lemma 2.2: The constants of the motion for (2.12) have the following recursion relations:

$$
C_{k+1} = \frac{1}{2}R_k\Big|_A - \frac{1}{2}\sum_{j=1}^{N}\phi_j^2\sum_{l=0}^{k}C_l\zeta_j^{k-l}, \quad k = 0,1,\ldots, \qquad (2.31a)
$$

or

$$
\begin{aligned}
C_{k+2} &= \frac{1}{2}Q_k\Big|_A \\
&\quad - \frac{1}{2}\sum_{j=1}^{N}\phi_j^2\sum_{l=0}^{k}C_l\zeta_j^{k-l}\left(\zeta_j - \frac{1}{2}\sum_{m=1}^{N}\phi_m^2\right). \qquad (2.31b)
\end{aligned}
$$

Proof: If Φ satisfies (2.12), it is found from (2.21) by taking $F\equiv 0$ that

$$
L^k u_x|_A = \sum_{j=1}^{N}\Psi_{jx}\sum_{l=0}^{k}C_l\zeta_j^{k-l}, \qquad (2.32)
$$

where

$$
\begin{aligned}
&C_0 = 1, \quad C_l = \frac{1}{2}\sum_{m=0}^{l}\alpha_m R_{l-m-1}\Big|_{A,x=x_0}, \\
&\alpha_0 = 1, \quad \alpha_m = -\frac{1}{2}\sum_{p=0}^{m-1}\alpha_p\sum_{j=1}^{N}\zeta_j^{m-p-1}\phi_j^2\Big|_{x=x_0}.
\end{aligned}
$$

However, (2.32) implies that the constants C_l's only depend on $\zeta_1,\ldots,\zeta_N$, Φ and have nothing to do with x_0. In other words, C_l takes same value for different x_0. Thus C_l's are the constants of the motion for (2.12). We have the following formulas for C_l:

$$
\begin{aligned}
&C_0 = 1, \quad C_l = \frac{1}{2}\sum_{m=0}^{l}\alpha_m R_{l-m-1}\Big|_A, \\
&\alpha_0 = 1, \quad \alpha_m = -\frac{1}{2}\sum_{p=0}^{m-1}\alpha_p\sum_{j=1}^{N}\zeta_j^{m-p-1}\phi_j^2.
\end{aligned}
$$

Then using (2.23) immediately gives (2.31).

In particular, calculating (2.31) gives

$$
\begin{aligned}
C_1 &= \left(\frac{1}{2}r - \frac{1}{2}\sum_{j=1}^{N}\phi_j^2\right)\Big|_A = 0 \\
C_2 &= \frac{1}{2}\left[q - \sum_{j=1}^{N}\zeta_j\phi_j^2 + \frac{1}{2}\left(\sum_{j=1}^{N}\phi_j^2\right)^2\right. \\
&\quad \left.+ \frac{1}{2}r\left(r - \sum_{j=1}^{N}\phi_j^2\right)\right]\Big|_A = 0, \\
C_3 &= -\frac{1}{4}\left[\sum_{j=1}^{N}\zeta_j^2\phi_j^2 + \sum_{j=1}^{N}\phi_{jx}^2\right. \\
&\quad \left.- \sum_{j=1}^{N}\phi_j^2\sum_{m=1}^{N}\zeta_m\phi_m^2 + \frac{1}{4}\left(\sum_{j=1}^{N}\phi_j^2\right)^3\right].
\end{aligned}
$$

Theorem 2.2: If Φ is a solution to (2.12), then $u = f(\Phi)$ satisfies a certain higher order stationary classical Boussinesq equation

$$
L^N u_x + \sum_{k=0}^{N-1}d_k L^k u_x = 0, \qquad (2.33)
$$

where the d_k's are some constants determined by $\zeta_1,\ldots,\zeta_N$, $C_1,\ldots,C_N$.

Proof: Set

$$p(\zeta) = (\zeta-\zeta_1)\cdots(\zeta-\zeta_N) = \zeta^N + \sum_{k=1}^N g_k \zeta^{N-k}; \tag{2.34a}$$

then

$$p(\zeta_j) = \zeta_j^N + \sum_{k=1}^N g_k \zeta_j^{N-k} = 0. \tag{2.34b}$$

Using (2.32) gives

$$\sum_{k=0}^N d_k L^k u_x = \sum_{j=1}^N \Psi_{jx} \sum_{k=0}^N d_k \sum_{m=0}^k C_m \zeta_j^{k-m}$$
$$= \sum_{j=1}^N \Psi_{jx} \sum_{k=0}^N \zeta_j^{N-k} \sum_{m=0}^k C_m d_{N-k+m}. \tag{2.35}$$

Taking $d_N = 1$ and

$$\sum_{m=0}^k C_m d_{N-k+m} = g_k, \quad k=1,\ldots,N,$$

or

$$d_{N-k} = g_k - \sum_{m=1}^k C_m d_{N-k+m}, \quad k=1,\ldots,N,$$

we then complete the proof by using (2.34) and (2.35).

III. THE INTEGRABILITY OF SYSTEM (2.12)

Introducing the canonical variables

$$q_m = \phi_m, \quad p_m = q_{mx},$$

system (2.12), namely,

$$\phi_{jxx} = -\zeta_j^2 \phi_j + \phi_j \sum_{l=1}^N \zeta_l \phi_l^2$$
$$-\frac{3}{4}\phi_j\Big(\sum_{l=1}^N \phi_l^2\Big)^2 + \zeta_j \phi_j \sum_{l=1}^N \phi_l^2, \; j=1,\ldots,N, \tag{3.1}$$

can be written in the canonical form

$$\frac{dp_m}{dx} = -\frac{\delta H}{\delta q_m}, \quad \frac{dq_m}{dx} = \frac{\delta H}{\delta p_m}, \tag{3.2}$$

where the Hamiltonian function H is given by

$$H = \frac{1}{2}\sum_{l=1}^N p_l^2 + \frac{1}{2}\sum_{l=1}^N \zeta_l^2 q_l^2 - \frac{1}{2}\sum_{j=1}^N q_j^2 \sum_{l=1}^N \zeta_l q_l^2$$
$$+\frac{1}{8}\Big(\sum_{l=1}^N q_l^2\Big)^3 = -2C_3.$$

Lemma 2.2 shows that

$$C_{k+2} = \frac{1}{2} Q_k \Big|_A$$
$$-\frac{1}{2}\sum_{j=1}^N \phi_j^2 \sum_{l=0}^k C_l \zeta_j^{k-l}\Big(\zeta_j - \frac{1}{2}\sum_{m=1}^N \phi_m^2\Big),$$
$$k = 1,2,\ldots \tag{3.3}$$

are the constants of the motion for the Hamiltonian system (3.2). However, Theorem 2.2 implies that when $k > N$, Q_k depends on $Q_1,\ldots,Q_N$. Thus we have the following proposition.

Proposition: $C_3,\ldots,C_{N+2}$ given by (3.3) are the N-independent constants of the motion for (3.2).

We now show that system (2.12), namely (3.1), possesses the Painlevé property. Notice that

$$\phi_j\phi_{jxx} = \frac{1}{2}(\phi_j^2)_{xx} - (\phi_{jx})^2$$
$$= \frac{1}{2}(\phi_j^2)_{xx} - \frac{1}{4}\frac{1}{\phi_j^2}[(\phi_j^2)_x]^2$$

if we call $\phi_j^2 = \psi_j$, then (3.1) is equivalent to

$$2\psi_j\psi_{jxx} - \psi_{jx}^2 = \Big[-4\zeta_j^2 + 4\sum_{l=1}^N \zeta_l\psi_l - 3\Big(\sum_{l=1}^N \psi_l\Big)^2$$
$$+ 4\zeta_j \sum_{l=1}^N \psi_l\Big]\psi_j^2, \quad j=1,\ldots,N. \tag{3.4}$$

Assume that

$$\psi_j = \frac{1}{(x-x_0)^\alpha}\sum_{k=0}^\infty a_{jk}(x-x_0)^k, \quad j=1,\ldots,N \tag{3.5}$$

satisfies system (3.4). A leading order analysis shows that $\alpha = 1$. Substituting (3.5) into (3.4), it is found that a_{j0} satisfies

$$\Big(\sum_{l=1}^N a_{l0}\Big)^2 = -1 \tag{3.6}$$

and the recursion relations for a_{jk} are

$$2a_{j0}\Big[k(k-2)a_{jk} + 3ia_{j0}\sum_{l=1}^N a_{lk}\Big] = f_{jk}, \quad j=1,\ldots,N, \tag{3.7}$$

$$f_{jk} = \sum_{m=1}^{k-1}(m-1)(k-3m+3)a_{jm}a_{j,k-m} - 4\zeta_j^2\sum_{m=0}^{k-2} a_{jm}a_{j,k-m-2}$$
$$+4\sum_{m=0}^{k-1}\sum_{p=0}^m a_{jp}a_{j,m-p}\cdot\sum_{l=1}^N(\zeta_l+\zeta_j)a_{l,k-m-1}$$
$$+3\sum_{p=1}^{k-1}a_{jp}a_{j,k-p} - 3\sum_{m=1}^{k-1}\sum_{p=0}^m\sum_{l=1}^N a_{lp}\sum_{n=1}^N a_{n,m-p}\sum_{q=0}^{k-m}a_{jq}a_{j,k-m-q} - 3a_{j0}^2\sum_{p=1}^{k-1}\sum_{l=1}^N a_{lp}\sum_{n=1}^N a_{n,k-p}.$$

If we call

$$A^{(k)} = (A_{jl}^{(k)}), \quad A_{jl}^{(k)} = 2a_{j0}[k(k-2)\delta_{jl} + 3ia_{j0}],$$
$$a^{(k)} = (a_{1k},\ldots,a_{Nk})^T, \quad f^{(k)} = (f_{1k},\ldots,f_{Nk})^T,$$

then (3.7) can be written as

$$A^{(k)}a^{(k)} = f^{(k)} . \tag{3.8}$$

A direct calculation gives

$$\det A^{(k)} = 2^N a_{10} \cdots a_{N0} (k+1) k^{N-1} (k-2)^{N-1} (k-3) . \tag{3.9}$$

We note that the recursion relations (3.8) are not defined at $k = -1, \underbrace{0,\ldots,0}_{N-1}, \underbrace{2,\ldots,2}_{N-1}, 3$. These values of k are called resonances and correspond to points where arbitrary constants are introduced into the expansion (3.5).[10] At each such resonance consistency demands that the rhs of (3.8) satisfies a compatibility condition—thereby ensuring the indeterminacy of the corresponding a_{jk}.

From (3.7) we find that

$$k=0, \quad \Big(\sum_{l=1}^{N} a_{l0}\Big)^2 = -1 , \tag{3.10}$$

$$k=1, \quad a_{j1} = a_{j0} \sum_{l=1}^{N} \zeta_l a_{l0} - 2i\zeta_j a_{j0}, \quad j = 1,\ldots,N, \tag{3.11a}$$

$$\sum_{l=1}^{N} a_{l1} = -i \sum_{l=1}^{N} \zeta_l a_{l0} , \tag{3.11b}$$

$$\sum_{l=1}^{N} \zeta_l a_{l1} = \Big(\sum_{l=1}^{N} \zeta_l a_{l0}\Big)^2 - 2i \sum_{l=1}^{N} \zeta_l^2 a_{l0} . \tag{3.11c}$$

By using (3.10) and (3.11), one obtains from (3.7) for $k=2$ that

$$-3 \sum_{l=1}^{N} a_{l2} = 3i\Big(\sum_{l=1}^{N} \zeta_l a_{l0}\Big)^2 + 4 \sum_{l=1}^{N} \zeta_l^2 a_{l0} . \tag{3.12}$$

Equations (3.10) and (3.12) show that $a^{(0)}$ and $a^{(2)}$ only need to satisfy one equation, respectively. Therefore, $N-1$ components of $a^{(0)}$ and $a^{(2)}$ are undetermined constants, respectively. For $k=3$, calculating (3.7) and using (3.11) give

$$\begin{aligned} 3a_{j3} + 3ia_{j0} \sum_{l=1}^{N} a_{l3} = \sum_{l=1}^{N} \Big[& 2a_{j0}\zeta_l a_{l2} + 4a_{j1}\zeta_l a_{l1} - 2a_{j2}\zeta_l a_{l0} + 4i\zeta_j a_{j1}\zeta_l a_{l0} - 6ia_{j1}a_{l2} + 3a_{j2}\zeta_l a_{l0} \\ & - 3a_{j0}a_{l1} \sum_{m=1}^{N} a_{m2} + 2\zeta_j a_{j0} a_{l2} + 4\zeta_j a_{j1} a_{l1} + 2a_{j1}\zeta_l a_{l0} \sum_{m=1}^{N} \zeta_m a_{m0} \Big] - 2i\zeta_j a_{j2} \\ = f_{j3}, \quad & j = 1,\ldots,N. \end{aligned} \tag{3.13}$$

By (3.10), (3.11), and (3.12), it is found that

$$\sum_{j=1}^{N} \Big(3a_{j3} + 3ia_{j0} \sum_{l=1}^{N} a_{l3}\Big) = 0 ,$$

$$\sum_{j=1}^{N} f_{j3} = -2 \sum_{l=1}^{N} \zeta_l a_{l0} \Big[3 \sum_{j=1}^{N} a_{j2} + 3i\Big(\sum_{j=1}^{N} \zeta_j a_{j0}\Big)^2 + 4 \sum_{j=1}^{N} \zeta_j^2 a_{j0} \Big] = 0$$

which, together with (3.9), indicate that one component of $a^{(3)}$ is an undetermined constant. Thus there are a total of $2N$ undetermined constants. (The resonance at $k = -1$ corresponds to the arbitrariness of x_0 itself.) Therefore, system (3.4) possesses the Painlevé property.[10] This suggests that systems (3.1) and (3.4) are integrable Hamiltonian systems.

IV. APPLICATION OF THE METHOD

The general method proposed in Sec. II can be applied to other hierarchies, for example, to the AKNS hierarchy[10]:

$$u_t = 2iJL^n u, \quad u = \begin{pmatrix} r \\ q \end{pmatrix}, \quad J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{4.1}$$

with

$$L = \frac{1}{2i} \begin{pmatrix} D - 2rD^{-1}q & 2rD^{-1}r \\ -2qD^{-1}q & -D + 2qD^{-1}r \end{pmatrix},$$

$$D = \frac{d}{dx}, \quad D^{-1}D = DD^{-1} = 1 .$$

Here we define the integral constant of D^{-1} appearing in L to be zero and do not require any boundary condition at infinity for u. The eigenvalue problem associated with (4.1) is the AKNS problem

$$\phi_x = M\phi, \quad M = \begin{pmatrix} -i\zeta & q(x,t) \\ r(x,t) & i\zeta \end{pmatrix}, \quad \phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}. \tag{4.2}$$

The time evolution equation for ϕ is

$$\phi_t = N\phi, \quad N = \begin{pmatrix} A & B \\ C & -A \end{pmatrix}, \tag{4.3}$$

where

$$A = \sum_{k=0}^{n} a_k \zeta^{n-k}, \quad B = \sum_{k=1}^{n} b_k \zeta^{n-k}, \quad C = \sum_{k=1}^{n} c_k \zeta^{n-k},$$

$$\begin{pmatrix} c_k \\ b_k \end{pmatrix} = L^{k-1} u , \tag{4.4a}$$

$$a_0 = -i, \quad a_k = D^{-1}(qc_k - rb_k) . \tag{4.4b}$$

It is known that a_k, b_k, and c_k are polynomials of q, r, and their derivatives.

The solvability condition of (4.2) and (4.3) reads as

$$\phi_{xt} - \phi_{tx} = \{M_t - N_x + [M,N]\}\phi = 0 , \tag{4.5}$$

where $[M,N] = MN - NM$. Substituting (4.4) into (4.5) gives

$$M_t - N_x + [M,N] = \begin{pmatrix} 0 & q_t + 2ib_{n+1} \\ r_t - 2ic_{n+1} & 0 \end{pmatrix}, \tag{4.6}$$

which leads to (4.1).

Consider

$$\Phi_{jx} = M_j \Phi_j, \quad M_j = \begin{pmatrix} -i\zeta_j & q \\ r & i\zeta_j \end{pmatrix}, \quad \Phi_j = \begin{pmatrix} \phi_{1j} \\ \phi_{2j} \end{pmatrix},$$
$$j = 1,...,N, \tag{4.7}$$

where $\zeta_k \neq \zeta_l$ when $k \neq 1$.

If we call

$$\Phi = (\phi_{11},...,\phi_{1N};\phi_{21},...,\phi_{2N})^T, \quad \Psi_j = \begin{pmatrix} \phi_{2j}^2 \\ -\phi_{1j}^2 \end{pmatrix},$$
$$\Psi = (\phi_{21}^2,...,\phi_{2N}^2; -\phi_{11}^2,..., -\phi_{1N}^2)^T,$$
$$L_0 = \frac{1}{2i}\begin{pmatrix} D - 2rI_0 q & 2rI_0 r \\ -2qI_0 q & -D + 2qI_0 r \end{pmatrix}, \quad I_0 = \int_{x_0}^{x} \cdot dy,$$

it is easy to check that if Φ satisfies (4.7), then

$$L_0 \Psi_j = \zeta_j \Psi_j + e_j u, \tag{4.8a}$$

where

$$e_j = -i\phi_{1j}\phi_{2j}|_{x=x_0}. \tag{4.8b}$$

Let $\widetilde{H}$ denote the linear subspace of the domain of the L_0 spanned by $\{\Psi_1,...,\Psi_N\}$. If we restrict u to $\widetilde{H}$, that is,

$$u = f(\Phi) = \sum_{j=1}^{N} \alpha_j \Psi_j, \tag{4.9}$$

then it is clear that $\widetilde{H}$ is left invariant by L_0. This property plays an important role in the proof of all the results for the whole hierarchy.

Under the constraint condition (4.9), we have the following systems from (4.2) and (4.3):

$$\Phi_{jx} = \overline{M}_j \Phi_j, \quad \overline{M}_j = M_j|_A, \quad j = 1,...,N \tag{4.10}$$

and

$$\Phi_{jt} = \overline{N}_j \Phi_j, \quad \overline{N}_j = N|_{A,\zeta=\zeta_j}, \quad j = 1,...,N, \tag{4.11}$$

where the subscript $A,\zeta = \zeta_j$ means to insert $u = f(\Phi)$ into the expression, and to take $\zeta = \zeta_j$.

Without loss of generality, one can consider the following constraint condition instead of (4.9):

$$u = \sum_{j=1}^{N} \Psi_j = f(\Psi) \tag{4.12a}$$

or

$$r = \sum_{j=1}^{N} \phi_{2j}^2, \quad q = -\sum_{j=1}^{N} \phi_{1j}^2. \tag{4.12b}$$

Then (4.10) reads as

$$\phi_{1jx} = -i\zeta_j \phi_{1j} - \sum_{l=1}^{N} \phi_{1l}^2 \phi_{2j},$$

$$\phi_{2jx} = \sum_{l=1}^{N} \phi_{2l}^2 \phi_{1j} + i\zeta_j \phi_{2j}, \quad j = 1,...,N. \tag{4.13}$$

We will see later that the submanifold of the phase space

$$\overline{H} = \left\{ u = \sum_{j=1}^{N} \Psi_j \,\middle|\, \Phi \text{ satisfies } (4.10) \right\}$$

is left invariant by all the flows in the hierarchy (4.1). This suggests that (4.10) is an integrable Hamiltonian system. Indeed, (4.10) can be reduced to a completely integrable Hamiltonian system called the confocal involutive system.[8]

In order to prove that systems (4.10) and (4.11) are naturally consistent, we now assume that $\Phi(x,t)$ satisfies (4.11) and set

$$F_j = \Phi_{jx} - \overline{M}_j \Phi_j, \quad F_j = \begin{pmatrix} F_{1j} \\ F_{2j} \end{pmatrix}, \quad j = 1,...,N, \tag{4.14}$$
$$F = (F_{11},...,F_{1N};F_{21},...,F_{2N})^T.$$

Lemma 4.1: Let $\Phi(x,t)$ be a solution to (4.11). Then

$$F_{jt} = \overline{N}_j F_j - \begin{pmatrix} 0 & q_t + 2ib_{n+1} \\ r_t - 2ic_{n+1} & 0 \end{pmatrix} \Phi_j \Big|_A. \tag{4.15}$$

Proof:

$$F_{jt} = \Phi_{jxt} - \overline{M}_{jt}\Phi_j - \overline{M}_j \Phi_{jt}$$
$$= (N_{jx}\Phi_j + N_j M_j \Phi_j + N_j F_j - M_{jt}\Phi_j - M_j N_j \Phi_j)|_A$$
$$= \overline{N}_j F_j - (M_{jt} - N_{jx} + [M_j,N_j])\Phi_j|_A,$$

which completes the proof by using (4.6).

Cao[8] has shown that the following theorem holds for the first three equations in the hierarchy (4.1) by direct calculation. We now use the general method to prove Theorem 4.1 for the whole hierarchy.

Theorem 4.1: Suppose that $\Phi(x,t)$ satisfies (4.11) and $\Phi(x,0)$ satisfies (4.10). Then $\Phi(x,t)$ satisfies (4.10) and $u = \Sigma_{j=1}^{N} \Psi_j$ is a solution to (4.1).

Proof: Using (4.14), a direct calculation gives

$$L_0 \Psi_j = \zeta_j \Psi_j + e_j u + G_j, \tag{4.16a}$$

where

$$e_j = -i\phi_{1j}\phi_{2j}|_{x=x_0},$$

$$G_j = -iuI_0(F_{1j}\phi_{2j} + F_{2j}\phi_{1j}) - i\begin{pmatrix} \phi_{2j}F_{2j} \\ \phi_{1j}F_{1j} \end{pmatrix}. \tag{4.16b}$$

Hence

$$L_0 u|_A = \sum_{j=1}^{N} \zeta_j \Psi_j + \alpha_1 \sum_{j=1}^{N} \Psi_j + G^{(1)}, \tag{4.17}$$

$$L_0^k u|_A = \sum_{j=1}^{N} \Psi_j \sum_{m=0}^{k} \alpha_m \zeta_j^{k-m} + G^{(k)}, \tag{4.18}$$

where

$$\alpha_0 = 1, \quad \alpha_m = \sum_{l=0}^{m-1} \delta_{m-l}\alpha_l, \tag{4.19}$$

$$\delta_0 = 1, \quad \delta_l = -i\sum_{j=1}^{N} \zeta_j^{l-1}\phi_{1j}\phi_{2j}\Big|_{x=x_0},$$

$$G^{(k)} = \sum_{j=1}^{N} G_j \sum_{m=0}^{k-1} \alpha_m \zeta_j^{k-m-1} + L_0 G^{(k-1)}. \tag{4.20}$$

By (4.4), it is found that

$$L\begin{pmatrix} c_k \\ b_k \end{pmatrix} = L_0 \begin{pmatrix} c_k \\ b_k \end{pmatrix} + \beta_k \begin{pmatrix} r \\ q \end{pmatrix}, \tag{4.21a}$$

where

$$\beta_k = ia_k|_{x=x_0} = [iD^{-1}(qc_k - rb_k)]|_{x=x_0}. \quad (4.21b)$$

Using (4.21) obtains

$$L^k u = \sum_{l=0}^{k} \beta_{k-l} L_0^l u, \quad (4.22)$$

with $\beta_0 = 1$. By (4.18) and (4.22), one can deduce that

$$L^k u|_A = \sum_{j=1}^{N} \Psi_j \sum_{l=0}^{k} h_l \zeta_j^{k-l}\Big|_A + \overline{G}^{(k)}, \quad (4.23a)$$

where

$$h_0 = 1, \quad h_l = \sum_{m=0}^{l} \alpha_m \beta_{l-m}, \quad (4.23b)$$

$$\overline{G}^{(k)} = \sum_{l=0}^{k} \beta_{k-l} G^{(l)}.$$

From (4.4) and (4.23) we have

$$\begin{pmatrix} c_k \\ b_k \end{pmatrix}\Big|_A = L^{k-1}u\Big|_A = \sum_{j=1}^{N} \Psi_j \sum_{l=0}^{k-1} h_l \zeta_j^{k-l-1}\Big|_A + \overline{G}^{(k-1)} \quad (4.24a)$$

and

$$2ra_k = 2rD^{-1}(qc_k - rb_k) = -2ic_{k+1} + c_{kx},$$
$$2qa_k = 2qD^{-1}(qc_k - rb_k) = -2ib_{k+1} - b_{kx}. \quad (4.24b)$$

Hence

$$\begin{aligned} rr_t|_A &= 2r \sum_{j=1}^{N} \phi_{2j}\phi_{2jt} \\ &= \sum_{j=1}^{N}\sum_{k=1}^{n} \phi_{2j}\zeta_j^{n-k}(2rc_k\phi_{1j} - 2ra_k\phi_{2j})\Big|_A + 2ir\sum_{j=1}^{N}\zeta_j^n\phi_{2j}^2 \\ &= \sum_{j=1}^{N}\sum_{k=1}^{n} \phi_{2j}\zeta_j^{n-k}\Big[2r\phi_{1j}\sum_{m=1}^{N}\phi_{2m}^2\sum_{l=0}^{k-1} h_l\zeta_m^{k-l-1} + 2r\phi_{1j}\overline{G}_1^{(k-1)} \\ &\quad + 2i\phi_{2j}\sum_{m=1}^{N}\phi_{2m}^2\sum_{l=0}^{k} h_l\zeta_m^{k-l} + 2i\phi_{2j}\overline{G}_1^{(k)} \\ &\quad - 2\phi_{2j}\sum_{m=1}^{N}\sum_{l=0}^{k-1} h_l\phi_{2m}(r\phi_{1m} + i\zeta_m\phi_{2m}) - \phi_{2j}\overline{G}_{1x}^{(k-1)}\Big]\Big|_A + 2ir\sum_{j=1}^{N}\zeta_j^n\phi_{2j}^2. \end{aligned} \quad (4.25)$$

Using (2.24), (4.25) becomes

$$rr_t|_A = 2ir\sum_{j=1}^{N}\phi_{2j}^2\sum_{k=0}^{n} h_k\zeta_j^{n-k} + r\overline{H}_1^{(n)}, \quad (4.26)$$

where

$$\overline{H}_1^{(n)} = \sum_{j=1}^{N}\sum_{k=1}^{n}\phi_{2j}\zeta_j^{n-k}\Big[2\phi_{1j}\overline{G}_1^{(k-1)} + \frac{2i}{r}\phi_{2j}\overline{G}_1^{(k)} - \frac{1}{r}\phi_{2j}\overline{G}_{1x}^{(k-1)}\Big].$$

Hence

$$(r_t - 2ic_{n+1})|_A = \overline{H}_1^{(n)} - 2i\overline{G}_1^{(n)} \equiv H_1^{(n)}. \quad (4.27)$$

In the same way, one can obtain

$$\begin{aligned}(q_t + 2ib_{n+1})|_A = \sum_{j=1}^{N}\sum_{k=1}^{n}\phi_{1j}\zeta_j^{n-k}\Big[&-2\phi_{2j}\overline{G}_2^{(k-1)} \\ &+ \frac{2i}{q}\phi_{1j}\overline{G}_2^{(k)} + \frac{1}{q}\phi_{1j}\overline{G}_{2x}^{(k-1)}\Big] \\ &- 2i\overline{G}_2^{(n)} \equiv H_2^{(n)}. \end{aligned} \quad (4.28)$$

Then the final argument in the proof of Theorem 2.1 can be used to complete the proof.

Remark: The submanifold of the phase space $\overline{H}$ is left invariant by all the flows in hierarchy (4.1).

Lemma 4.2: The constants of the motion h_l's for (4.10) satisfy the recursion relations

$$h_l = i\sum_{m=0}^{l}\alpha_m a_{l-m}\Big|_A, \quad 1 = 1,2,\ldots, \quad (4.29)$$

$$\alpha_0 = 1, \quad \alpha_m = -i\sum_{j=1}^{N}\phi_{1j}\phi_{2j}\sum_{p=0}^{m-1}\alpha_p\zeta_j^{m-p-1}.$$

Proof: If Φ satisfies (4.10), from (4.23), we obtain by taking $F_j \equiv 0$,

$$L^k u|_A = \sum_{j=1}^{N}\Psi_j\sum_{l=0}^{k} h_l\zeta_j^{k-l}, \quad (4.30)$$

which implies that the constant h_l, defined by (4.19), (4.21b), and (4.23b), is independent of x_0. Thus h_l's are the constants of the motion for (4.10) and satisfy (4.29).

Calculating (4.29) gives

$$h_1 = i\sum_{j=1}^{N}\phi_{1j}\phi_{2j},$$

$$h_2 = -\frac{1}{2}\sum_{j=1}^{N}\phi_{1j}^2\sum_{m=1}^{N}\phi_{2m}^2 - i\sum_{j=1}^{N}\zeta_j\phi_{1j}\phi_{2j} - \Big(\sum_{j=1}^{N}\phi_{1j}\phi_{2j}\Big)^2.$$

In the same way as we did for Theorem 2.2, we have the following theorem.

Theorem 4.2: If Φ is a solution to (4.10), then $u = \sum_{j=1}^{N}\Psi_j$ satisfies a certain higher order stationary AKNS equation

$$L^N u + \sum_{k=0}^{N-1} d_k L^k u = 0, \quad (4.31)$$

where the constants d_k's are determined by $\zeta_1,\ldots,\zeta_N$,

$h_1,\ldots,h_N$.

Equation (4.10) can be written in Hamiltonian form:

$$\phi_{1jx} = -\frac{\delta H}{\delta \phi_{2j}}, \quad \phi_{2jx} = \frac{\delta H}{\delta \phi_{1j}}, \tag{4.32}$$

where the Hamiltonian function H is

$$H = \frac{1}{2}\sum_{j=1}^{N}\phi_{1j}^2 \sum_{m=1}^{N}\phi_{2m}^2 + i\sum_{j=1}^{N}\zeta_j\phi_{1j}\phi_{2j} = -h_2 + h_1^2 .$$

Lemma 4.2 and Theorem 4.2 lead to the following corollary.

Corollary: The N-independent constants of the motion for (4.10) or (4.32) are $h_1,\ldots,h_N$ defined by (4.29).

V. THE APPLICATION OF BÄCKLUND AND GAUGE TRANSFORMATION

(i) The Jaulent–Miodek hierarchy[11]

$$\tilde{u}_t = \tilde{L}^n\tilde{u}_x, \quad \tilde{u} = \begin{pmatrix}\tilde{q}\\ \tilde{r}\end{pmatrix}, \tag{5.1}$$

with

$$\tilde{L} = \begin{pmatrix}0 & -\frac{1}{4}D^2 + \tilde{q} + \frac{1}{2}\tilde{q}_x D^{-1}\\ 1 & \tilde{r} + \frac{1}{2}\tilde{r}_x D^{-1}\end{pmatrix}$$

is associated with the following eigenvalue problem:

$$\phi_{xx} = \tilde{M}\phi, \quad \tilde{M} = -\zeta^2 + \tilde{q} + \zeta\tilde{r}. \tag{5.2}$$

It is known[13] that there exists the Bäcklund transformation between u and $\tilde{u}$, which is a solution to (2.1) and (5.1), respectively:

$$\tilde{q} = q - \tfrac{1}{4}r^2, \qquad \tilde{r} = r. \tag{5.3}$$

Under the transformation (5.3), (2.2) is transformed to (5.2) and vice versa. Using (2.14b) and (5.3) immediately gives the natural constraint for $\tilde{u}$, $\tilde{u} = \tilde{f}(\Phi)$, namely,

$$\tilde{r} = \sum_{j=1}^{N}\phi_j^2 \equiv \tilde{f}_2(\Phi), \qquad \tilde{q} = \sum_{j=1}^{N}\zeta_j\phi_j^2 - \frac{3}{4}\left(\sum_{j=1}^{N}\phi_j^2\right)^2 \equiv \tilde{f}_1(\Phi). \tag{5.4}$$

Under the constraint condition (5.4), the finite-dimensional integrable Hamiltonian system deduced from (5.2) is the same as (2.12), that is,

$$\phi_{jxx} = \left[-\zeta_j^2 + \sum_{l=1}^{N}\zeta_l\phi_l^2 - \frac{3}{4}\left(\sum_{l=1}^{N}\phi_l^2\right)^2 + \zeta_j\sum_{l=1}^{N}\phi_l^2\right]\phi_j, \quad j = 1,\ldots,N. \tag{5.5}$$

Using the Bäcklund transformation (5.3), we can gain similar results for the Jaulent–Miodek hierarchy from ones for the classical Boussinesq hierarchy.

Theorem 5.1: Let $\Phi(x,t)$ be a solution to (2.13) and $\Phi(x,0)$ be a solution to (5.5). Then $\Phi(x,t)$ satisfies (5.5) and $\tilde{u} = \tilde{f}(\Phi)$ satisfies (5.1).

Theorem 5.2: Let Φ be a solution to (5.5). Then $\tilde{u} = \tilde{f}(\Phi)$ satisfies a certain higher order stationary Jaulent–Miodek equation

$$\tilde{L}^N\tilde{u}_x + \sum_{k=0}^{N-1} d_k\tilde{L}^{N-k}\tilde{u}_x = 0, \tag{5.6}$$

where the d_k's are determined by $\zeta_1,\ldots,\zeta_N$ and the constants of the motion for (5.5) or (2.12).

(ii) Consider the Kaup–Newell eigenvalue problem[12]

$$\psi_{1x} = -i\xi^2\psi_1 + \xi q_1\psi_2, \qquad \psi_{2x} = \xi r_1\psi_1 + i\xi^2\psi_2. \tag{5.7}$$

By the transformation

$$\psi_1 = (1/\xi)\tilde{\phi}_1, \qquad \psi_2 = \tilde{\phi}_2, \qquad \xi^2 = \zeta, \tag{5.8}$$

(5.7) is equivalent to

$$\tilde{\phi}_{1x} = -i\zeta\tilde{\phi}_1 + \zeta q_1\tilde{\phi}_2, \qquad \tilde{\phi}_{2x} = r_1\tilde{\phi}_1 + i\zeta\tilde{\phi}_2, \tag{5.9}$$

which is associated with the hierarchy

$$\tilde{u}_t = D\tilde{L}^n\tilde{u}, \quad \tilde{u} = \begin{pmatrix}r_1\\ q_1\end{pmatrix}, \tag{5.10}$$

where

$$\tilde{L} = \frac{1}{2i}\begin{pmatrix}D + ir_1D^{-1}q_1D & ir_1D^{-1}r_1D\\ iq_1D^{-1}q_1D & -D + iq_1D^{-1}r_1D\end{pmatrix}.$$

Following the same procedure as applied to the AKNS hierarchy, we find that the natural constraint for $\tilde{u}$ is

$$\tilde{u} = \tilde{f}(\tilde{\Phi}) = \sum_{j=1}^{N}\begin{pmatrix}\tilde{\phi}_{2j}^2\\ -(1/\zeta_j)\tilde{\phi}_{1j}^2\end{pmatrix}. \tag{5.11}$$

Under the constraint condition (5.11), one obtains, from (5.9),

$$\tilde{\phi}_{1jx} = -i\zeta_j\tilde{\phi}_{1j} - \left(\sum_{l=1}^{N}\frac{1}{\zeta_l}\tilde{\phi}_{1l}^2\right)\zeta_j\tilde{\phi}_{2j}, \qquad \tilde{\phi}_{2jx} = \left(\sum_{l=1}^{N}\tilde{\phi}_{2l}^2\right)\tilde{\phi}_{1j} + i\zeta_j\tilde{\phi}_{2j}, \quad j = 1,\ldots,N, \tag{5.12}$$

which can be shown to possess the Painlevé property. Thus (5.12) is supposed to be an integrable system deduced from (5.9) in a natural way.

In general, it is not easy to find the transform relation between two integrable systems. However, the gauge transformation, which transforms one eigenvalue problem to another, provides a simple way to obtain the transform relation between two systems deduced from these two eigenvalue problems, respectively.

The system (4.10), namely,

$$\phi_{1jx} = -i\zeta_j\phi_{1j} - \sum_{l=1}^{N}\phi_{1l}^2\phi_{2j}, \qquad \phi_{2jx} = \sum_{l=1}^{N}\phi_{2l}^2\phi_{1j} + i\zeta_j\phi_{2j}, \quad j = 1,\ldots,N, \tag{5.13}$$

is an integrable Hamiltonian system. It is known[14] that there exists the following gauge transformation between (4.2) and (5.9):

$$\phi_1 = \lambda\tilde{\phi}_1 + \tfrac{1}{2}iq_1\lambda\tilde{\phi}_2, \qquad \phi_2 = \lambda^{-1}\tilde{\phi}_2 \tag{5.14a}$$

and

$$q = \lambda^2(\tfrac{1}{4}q_1^2r_1 + \tfrac{1}{2}iq_{1x}),$$

$$r=\lambda^{-2}r_1, \tag{5.14b}$$

where

$$\lambda=\exp\left(\int_{x_0}^{x}\frac{1}{2}iq_1r_1\,dy\right).$$

Equations (5.14) transform (5.9) to (4.2). Using the system deduced from (5.14a),

$$\phi_{1j}=\lambda\tilde{\phi}_{1j}+\tfrac{1}{2}iq_1\lambda\tilde{\phi}_{2j},$$
$$\phi_{2j}=\lambda^{-1}\tilde{\phi}_{2j},\quad j=1,\ldots,N, \tag{5.15}$$

it is found from (5.14b) and (5.11) that

$$r=\lambda^{-2}\sum_{j=1}^{N}\tilde{\phi}_{2j}^2=\sum_{j=1}^{N}\phi_{2j}^2,$$

$$q=\lambda^2\left[\frac{1}{4}q_1^2\sum_{j=1}^{N}\tilde{\phi}_{2j}^2-\frac{1}{2}i\sum_{j=1}^{N}(2\zeta_j^{-1}\tilde{\phi}_{1j}\tilde{\phi}_{1jx})\right]$$
$$=-\lambda^2\sum_{j=1}^{N}\left(\tilde{\phi}_{1j}+\frac{1}{2}iq_1\tilde{\phi}_{2j}\right)^2=-\sum_{j=1}^{N}\phi_{1j}^2.$$

Hence the formula

$$\phi_{1j}=\lambda\tilde{\phi}_{1j}-\frac{1}{2}i\lambda\sum_{l=1}^{N}\frac{1}{\zeta_l}\tilde{\phi}_{1l}^2\tilde{\phi}_{2j},$$
$$\phi_{2j}=\lambda^{-1}\tilde{\phi}_{2j},\quad j=1,\ldots,N,$$

with

$$\lambda=\exp\left[-\int_{x_0}^{x}\frac{1}{2}i\sum_{l=1}^{N}\frac{1}{\zeta_l}\tilde{\phi}_{1l}^2\sum_{m=1}^{N}\tilde{\phi}_{2m}^2\,dy\right]$$

gives the transform relation between (5.12) and (5.13). Since (5.13) is equivalent to the well-known confocal involutive system,[8] (5.12) can also be reduced to the confocal involutive system by using (5.14b) and (5.15).

ACKNOWLEDGMENTS

The authors wish to thank the referee for pointing out that formula (1.3) can be obtained from (1.2).

This work was supported by the National Natural Science Foundation of China.

[1] J. Moser, in *Proceedings of the 1983 Beijing Symposium on Differential Geometry and Differential Equations*, edited by S. T. Liao (Science, Beijing, China), pp. 157–229.

[2] H. P. Mckean, Jr., Springer Lecture Notes in Mathematics, Vol. 755.

[3] D. V. Choodnovsky and G. V. Choodnovsky, Nuovo Cimento B **40**, 339 (1977).

[4] H. Airault, H. P. Mckean, Jr., and J. Moser, Commun. Pure Appl. Math. **30**, 95 (1977).

[5] H. Flaschka, in *Proceedings of RIMS Symposium on Nonlinear Integrable Systems–Classical Theory and Quantum Theory*, Kyoto, 1981, edited by M. Jimbo and T. Miwa (World Science, Singapore, 1983), pp. 219–240.

[6] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).

[7] P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975).

[8] C. W. Cao, (a) Henan Science **5**, 1 (1987); (b) (To be published).

[9] D. J. Kaup, Prog. Theor. Phys. **54**, 396 (1975).

[10] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981)

[11] M. Jaulent and I. Miodek, Lett. Math. Phys. **1**, 243 (1976).

[12] D. J. Kaup and A. C. Newell, J. Math. Phys. **19**, 798 (1978).

[13] M. Jaulent and I. Miodek, Lett. Nuovo Cimento **20**, 655 (1977).

[14] Y. S. Li, D. Y. Chen, and Y. B. Zeng, in *Proceedings of the 1983 Beijing Symposium on Differential Geometry and Differential Equations*, edited by S. T. Liao (Science, Beijing, China), pp. 359–368.

Proof of the family index theorem via a path integral

Miao Li[a)]
ICRA–International Center for Relativistic Astrophysics, Rome, Italy and Dipartimento di Fisica, Università "La Sapienza," 00185 Rome, Italy

(Received 19 May 1988; accepted for publication 2 February 1989)

The integral formula of the family index theorem and its equivariant version are derived by use of a path integral, analogous to such a proof of the Atiyah–Singer index theorem, but without supersymmetry. Some formulas explicitly presented here should be useful in applications of the family index theorem to physics.

I. INTRODUCTION

The family index theorem of Atiyah and Singer[1] was proved again by Bismut,[2] using Quillen's superconnection formalism[3] and the estimation of probability method. This proof is closely related to the physicist's proof of the Atiyah–Singer index theorem.[4–6] It is expected that, if we adopt some appropriate Hamiltonian related to the superconnection, the family index theorem and its equivariant version can also be proved by path integral representation instead of by estimation of other methods. Besides the proof being of interest to physicists, the explicit exposition of formulas will be useful in some physical applications of the family index theorem, especially in the study of determinant bundles in which both mathematicians and physicists are most interested recently.[7–9] In this paper, we demonstrate the path integral proof of the formula of the family index theorem, assuming that the formulation of this theorem in superconnection formalism is known. The application of our proof will be considered in a separate paper. I would like to point out that although the integral formula of family index is well-known, the one of equivariant family index seems to be new, to the best of my knowledge.

To begin with, let us review some necessary preliminaries presented in Ref. 2 upon which our proof is based. Suppose X is a fibered manifold $\pi: X \to B$ with fiber space M, an even-dimensional spin manifold. Define a vector bundle over X with fiber $T_x M$, which is the tangent space along M at point x. This bundle is denoted as TM and the tangent bundle of X can be decomposed in some way as $TX = TM \oplus T^H X$, where $T^H X$ is the horizontal part of TX. Given a metric g_M on TM, one then has a related metric on $\Delta(TM)$, the spin vector bundle corresponding to TM. This metric enables us to define the Dirac operator D. If in addition there is a vector bundle E and a connection A on it, we define the Dirac operator coupled to E via A, namely, the Dirac operator acting on $\Delta(TM) \otimes E$.

The "family index" is defined as the Chern character of index bundle $\operatorname{Ind} D = \operatorname{Ker} D - \operatorname{Coker} D$. The latter is well-defined, since M is even dimensional so γ_5 can be defined to split a spinor into parts of positive and negative chirality.

Suppose B is endowed with a metric g_B and its lift is to be one of $T^H M$. TX then has a metric $g = g_M \oplus g_B$. ∇^L is the Levi-Civita connection of metric $g_M \oplus g_B$. Connection ∇^M of TM is defined by $\nabla^M V = P_M \nabla^L V$, where $V \in TM$ and P_M is the projection operator from TX to TM. Further, let ∇^B denote both the Levi-Civita connection of g_B and its lift as one of $T^H X$. Note that $\nabla = \nabla^M \oplus \nabla^B$ is a connection of TX and differs from ∇^L by a tensor S: $\nabla^L = \nabla + S$. Let $\overline{\nabla} = dy^\alpha \nabla_{f_\alpha}$, where dy^α, f_α are bases of $T_x^H M$ and its dual space, respectively. Then $D + \overline{\nabla}$ is a Quillen's superconnction[3] that sends an even element into an odd one and vice-versa. Let

$$E = i\gamma_a \left[\tfrac{1}{4}\langle S(e_a)e_b, f_\alpha\rangle i\gamma_b \, dy^\alpha + \tfrac{1}{4}\langle S(e_a) f_\alpha, f_\beta\rangle dy^\alpha \, dy^\beta - \tfrac{1}{2}\langle S(f_\alpha)e_a, f_\beta\rangle dy^\alpha \, dy^\beta \right],$$

where e_a is a orthonormal basis of $T_x M$, and γ matrices satisfy $\{\gamma_a, \gamma_b\} = 2\delta_{ab}$. $D + \overline{\nabla} + E$ is also a superconnection.

The following result is proved in Ref. 2: Let $I = (D + \overline{\nabla} + E)^2$ and I^β be such an operator corresponding to metric $g_M/\beta \oplus g_B$, then $\operatorname{str} \exp(-\frac{1}{2} I^\beta) = \operatorname{tr} \gamma_5 \exp(-\frac{1}{2} I^\beta)$ represents $\overline{\operatorname{Ch}}(\operatorname{Ind} D)$, where $\overline{\operatorname{Ch}}(R) = \exp(-R/2)$ if R is the curvature form. The proof is analogous to the proof of a similar theorem in the original paper of Quillen.[3]

Based on this theorem, we prove the formula of family index in Sec. III. Section II is devoted to a detailed discussion of geometric data, which will be used in Sec. III and is useful in some applications of the family index theorem. In Sec. IV, we prove the equivariant version of the family index theorem. We shall point out that our proof is acceptable only to physicists, and replaces only the probability proof of Bismut,[2] which is rigorous but not familiar to physicists.

II. GEOMETRIC DATA

For an open set of fibered space X, let there be a trivialization, $\{x^i, y^\alpha\}$, y^α are the coordinates of base space B, and x^i are the coordinates of fiber M. The tangent space of X at x is $T_x X = \{\partial_i, \partial_\alpha\}$ and cotangent space is $T_x^* X = \{dx^i, dy^\alpha\}$. It is easy to see that the bundle TM defined in the last section has a basis of trivialization $T_x M = \{\partial_i\}$. Suppose TX is decomposed in a way that $TX = TM \oplus T^H X$. $T^H X$ can be written in general as $T_x^H X = \{\partial_\alpha + f^i_\alpha \partial_i\}$, where f^i_α are some functions of $\{x^i, y^\alpha\}$. Having $T_x^H X$ well-defined means that, under the coordinate transformation of $\bar{x}^i = \bar{x}^i(x, y)$, f^i_α transforms as

$$\bar{f}^i_\alpha = \frac{\partial \bar{x}^i}{\partial x^j} f^j_\alpha - \frac{\partial \bar{x}^i}{\partial y^\alpha}. \qquad (2.1)$$

Therefore f^i_α can not always be taken as zero. Also, it is the transformations $\bar{x}^i$ that determine the structure of fibered

[a)] On leave from the Center of Astrophysics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China.

 0022-2488/89/081690-05$02.50

space X. Note that f^i_α may be viewed in some sense as a "connection." Bundle T^HX is the lift of bundle TB on X, and the lift ∂^H_α of ∂_α is $\partial_\alpha + f^i_\alpha \partial_i$. The dual of TX can be decomposed as $T^*X = T^*M \oplus T^{*H}X$ such that T^*M and $T^{*H}X$ are orthogonal to TM and T^HX, respectively. Locally, $T^*_xX = \{dx^i - f^i_\alpha\, dy^\alpha\} \oplus \{dy^\alpha\}$. Let metrics g_M and g_B be defined as in the last section so $g_M = g_{ij}\, dx^i \otimes dx^j$, $g_B = g_{\alpha\beta}\, dy^\alpha \otimes dy^\beta$. Metric $g = g_M \oplus g_B$ is given by

$$g = g_{ij}(dx^i - f^i_\alpha\, dy^\alpha) \otimes (dx^j - f^j_\beta\, dy^\beta) + g_{\alpha\beta}\, dy^\alpha \otimes dy^\beta. \tag{2.2}$$

Now the Dirac operator D on fiber M_y is defined completely by g_{ij}, but the family of D depends on the structure of fibered space X and thus depends on f^i_α. The components of the Levi-Civita connection defined by the metric g_{ij} are denoted by Γ^k_{ij}. Let the components of the Levi-Civita connection ∇^L be defined as follows:

$$\begin{aligned}
&\nabla^L_{\partial^H_\alpha}\partial^H_\beta = \Gamma^k_{\alpha\beta}\partial_k + \Gamma^\gamma_{\alpha\beta}\partial^H_\gamma,\\
&\nabla^L_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k + \Gamma^\gamma_{ij}\partial^H_\gamma,\\
&\nabla^L_{\partial^H_\alpha}\partial_j = \Gamma^k_{\alpha j}\partial_k + \Gamma^\gamma_{\alpha j}\partial^H_\gamma,\\
&\nabla^L_{\partial_i}\partial^H_\alpha = \Gamma^k_{i\alpha}\partial_k + \Gamma^\gamma_{i\alpha}\partial^H_\gamma.
\end{aligned} \tag{2.3}$$

After some calculations, we obtain

$$\begin{aligned}
&\Gamma^k_{ij} = \Gamma^k_{ij}\ \Gamma^\gamma_{\alpha\beta} = \Gamma^\gamma_{\alpha\beta},\\
&\Gamma^\alpha_{ij} = \Gamma^\alpha_{ji} = -\tfrac{1}{2}g^{\alpha\beta}(\partial^H_\beta g_{ij} + g_{ik}f^k_{\beta,j} + g_{jk}f^k_{\beta,i}),\\
&\Gamma^i_{\alpha j} = \Gamma^i_{j\alpha} - f^i_{\alpha,j}\,\Gamma^i_{j\alpha} = -g_{\alpha\beta}g^{ik}\Gamma^\beta_{jk},\\
&\Gamma^i_{\alpha\beta} = -\Gamma^i_{\beta\alpha} = \tfrac{1}{2}(\partial^H_\alpha f^i_\beta - \partial^H_\beta f^i_\alpha),\\
&\Gamma^\beta_{\alpha i} = \Gamma^\beta_{i\alpha} = g_{ij}g^{\beta\gamma}\Gamma^j_{\gamma\alpha}.
\end{aligned} \tag{2.4}$$

These formulas enable us to calculate the curvature of the connection ∇. In proving the family index theorem, we need to know only its components restricted to TM, namely, $R^i_{\ j}(A,B)$, $AB \in TX$. These components are

$$\begin{aligned}
&R^{\ i}_{jkl} = R^{\ i}_{jkl},\\
&R^{\ i}_{j\alpha\beta} = \partial^H_\alpha\Gamma^i_{\beta j} + \Gamma^l_{\beta j}\Gamma^i_{\alpha l} - (\alpha\leftrightarrow\beta) - 2\Gamma^l_{\alpha\beta}\Gamma^i_{lj},\\
&R^{\ i}_{jk\alpha} = \Gamma^i_{\alpha j,k} - \partial^H_\alpha\Gamma^i_{jk} - \Gamma^i_{kl}\,\Gamma^l_{\alpha j} - \Gamma^i_{\alpha l}\Gamma^l_{jk} - \Gamma^i_{lj}f^l_{\alpha,k},
\end{aligned} \tag{2.5}$$

where indices α,β indicate that the components are taken in basis ∂^H_α. The following formula can be easily proved by use of Eqs. (2.4) and (2.5), although the calculation is rather tedious:

$$\begin{aligned}
R_{ij\alpha\beta} = {} & -g_{jk}\Gamma^l_{i[\alpha}\Gamma^k_{l\beta]} - g_{ik}\Gamma^k_{\alpha\beta,j} + g_{jk}\Gamma^k_{\alpha\beta,i}\\
& - \Gamma^k_{\alpha\beta}g_{ij,k} + 2\Gamma^l_{\alpha\beta}\Gamma^k_{li}g_{kj}.
\end{aligned} \tag{2.6}$$

It is not easy to see the fact that the components of curvature R_{ab} in the orthonormal frame are asymmetric in indices a and b. This fact will become obvious later. Since in Sec. III we will use the tangent space approximation to calculate the path integral, here we present some formulas that will be useful there. Around a given point x_0 on the fiber M_y, we can always choose a coordinate transformation $\bar{x}^i = \bar{x}^i(x,y)$ such that at x_0, $\bar{g}_{ij} = \delta_{ij}$. Now $\bar{f}^i_\alpha$ are given by (2.1). Since curvature $R^{\ i}_{jkl}$ is determined by g_{ij}, we can always choose normal coordinates such that around point x_0,

$$g_{ij}(x,y) = \delta_{ij} - \tfrac{1}{3}R_{ikjl}(x_0,y)x^kx^l + O(x^3), \tag{2.7}$$

where we assumed that at x_0, $x^i = 0$. This formula is still valid when y varies in a trivialized open set of X.

At point x_0, the curvatures in Eq. (2.5) are given by

$$\begin{aligned}
&R_{ij\alpha\beta}(x_0,y) = \Gamma^i_{\alpha\beta,i} - \Gamma^i_{l\alpha}\Gamma^j_{l\beta} - (i\leftrightarrow j),\\
&R_{ijk\alpha}(x_0,y) = \Gamma^i_{\alpha j,k} - \Gamma^i_{kj,l}f^l_\alpha
\end{aligned} \tag{2.8}$$

where all quantities denote those at point x_0. Around x_0 we have

$$\begin{aligned}
&\Gamma^i_{jk} = -\tfrac{1}{3}(R_{ikjl} + R_{ijkl})x^l + O(x^2),\\
&\Gamma^i_{\alpha j} = \tfrac{1}{2}(f^j_{\alpha,i} - f^i_{\alpha,j}) - \tfrac{1}{6}(R_{ikjl} + R_{iljk})x^kf^l_\alpha + O(x^2).
\end{aligned} \tag{2.9}$$

Using the above equations and (2.8), at last we derive the following useful formula:

$$R_{ijk\alpha} = \Gamma^j_{k\alpha,i} - \Gamma^i_{k\alpha,j}. \tag{2.10}$$

Still, those above quantities are at x_0, and from Eqs. (2.8) and (2.10) we see that the curvature is asymmetric in indices i and j.

In the case of the equivalent family index theorem, we introduce an automorphism group G acting fiber-wise on X. This group G is lift to act on $\Delta(M) \otimes E$ and $C(\Delta(M) \otimes E)$, and the latter is the set of sections of $\Delta(M) \otimes E$. Group G is assumed to commute with the Dirac operator D, namely, if $g \in G$, $gD = Dg$. Also, $g\gamma_5 = \gamma_5 g$, and G preserves the grading of $\Delta(M) \otimes E$.

If g has a fixed point set $M_y(g)$ on each fiber M_y, then $M_y(g)$ is supposed to be diffeomorphic to each other for different y's. The fixed point set of X is another fibered space over base space B with fiber space $M(g)$. Let x^a denote coordinates of $M_y(g)$ and the rest of the dimensions described by x^r. On $M_y(g)$, TM_y is decomposed as $\{\partial_a, \partial_r\}$, and metric g_{ij} is grouped into two parts g_{ab} and g_{rs}. Since g preserves the metric g_{ij} and $M(g)$ is its fixed point set, any geodesic in $M(g)$ determined by the induced metric g_{ab} is a geodesic in M also.[10] This implies that the connection of metric g_{ij} on $M(g)$ has no Γ^r_{ab} and Γ^b_{ar} components.[10] This in turn implies that the components $R^{\ a}_{bcd}$ of curvature on $M(g)$ determined by metric g_{ij} are identical to those completely determined by the metric g_{ab}. Similarly preservation of the metric (2.2) makes $f^a_{\alpha,r} = f^r_{\alpha,a} = 0$ on $M_y(g)$. Using these facts and formulas $\Gamma^a_{bc,r} = 0$ on $M_y(g)$ [these are equivalent to $g_{ab,r} = 0$ or $\Gamma^a_{br} = 0$], in the normal coordinates we derive the following formulas for quantities that will appear in Sec. IV:

$$\begin{aligned}
&R_{ab\alpha\beta} = \Gamma^b_{\alpha\beta,a} - \Gamma^a_{c\alpha}\Gamma^b_{c\beta} - (a\leftrightarrow b),\\
&R_{abc\alpha} = \Gamma^b_{c\alpha,a} - \Gamma^a_{c\alpha,b},\\
&R_{rs\alpha\beta} = \Gamma^s_{\alpha\beta,r} - \Gamma^r_{t\alpha}\Gamma^s_{t\beta} - (r\leftrightarrow s),\\
&R_{rst\alpha} = \Gamma^s_{t\alpha,r} - \Gamma^r_{t\alpha,s}.
\end{aligned} \tag{2.11}$$

Note that in the above equations, $\Gamma^a_{b\alpha}$ is independent of f^r_α. Also note the fact that $\Gamma^a_{r\alpha} = \Gamma^r_{a\alpha} = 0$ is used.

III. THE FAMILY INDEX THEOREM

In the path integral of the index theorem,[5,6] one introduces the formula

$$\begin{aligned}
\mathrm{Index}(Q) &= \mathrm{str}\exp(-\beta Q^2)\\
&= \int [dx\, d\psi]\exp\Big(-\int_0^\beta L_E\, d\tau\Big),
\end{aligned} \tag{3.1}$$

where Q^2 is the Hamiltonian of the supersymmetric system. Here in dealing with the family index, we have a similar formula, because the family index can also be expressed as a supertrace. This is to assume that Bismut's result is known. To independently derive the formula, we might need some supersymmetry. But unfortunately, from the formula that will be given below, one can hardly find out a supersymmetry. I will observe this problem in the future.

In Ref. 2, Bismut proved that the analog of Hamiltonian Q^2 in the family index is $\frac{1}{2\beta}I^\beta$, namely,

$$\operatorname{str}\exp(-\beta H) = \operatorname{str}\exp(-\tfrac{1}{2}I^\beta) = \overline{\mathrm{Ch}}(\mathrm{Ind}\, D).$$

Bismut proved the following identity[2]:

$$\begin{aligned} H &= \frac{1}{2\beta} I^\beta = -\frac{1}{2}(D_a + V_a)^2 + \frac{1}{8}R + V, \\ V_a &= e_a^i\Big(\frac{i}{\sqrt{2\beta}}\langle S(\partial_i)\partial_j f_a\rangle \psi^j\, dy^\alpha \\ &\quad + \frac{1}{4\beta}\langle S(\partial_i) f_\alpha f_\beta\rangle dy^\alpha\, dy^\beta\Big), \qquad (3.2)\\ V &= -\frac{1}{2}\psi^i\psi^j F_{ij} + \frac{1}{4\beta} dy^\alpha\, dy^\beta F_{\alpha\beta} \\ &\quad + \frac{i}{\sqrt{2\beta}}\psi^i\, dy^\alpha F_{i\alpha}, \end{aligned}$$

where tensor S is defined in Sec. I, F is the curvature of bundle E to which Dirac operator D is coupled. If we consider dy^α as some anticommunting parameters, then (3.2) can be viewed as a Hamiltonian of a system. Note that in V_a and V, there is no term containing $\dot{x}^i$, so the corresponding Lagrangian should be simple. However we have to note that when a gauge field is present, H is no longer a Hamiltonian containing only dynamic variables x^i and ψ^i, but also an additional set of fields $\bar{\eta}_a$, η^a, since in (3.2), F is matrix valued. According to Ref. 6, we can simply treat these fields as matrices; the path integral will not involve any integral over these variables. We will use prescription

$$\overline{\mathrm{Ch}}(\mathrm{Ind}\, D) = \operatorname{tr}\int [dx\, d\psi]\exp\Big(-\int_0^\beta d\tau\, L_E\Big). \qquad (3.3)$$

We now prove the following theorem.

Theorem: Using a certain strategy of operator ordering, the corresponding Lagrangian of H is

$$L = \tfrac{1}{2}\dot{x}^a\dot{x}^a + (i/2)\psi_a(\dot{\psi}_a + \omega_c^{ab}\psi^b\dot{x}^c) + iV_a\dot{x}^a - V, \qquad (3.4)$$

where $\dot{x}^a = e_i^a\dot{x}^i$ and $\psi^a = e_i^a\psi^i$.

Proof: We define $H = p_a\dot{x}^a + \frac{i}{2}\psi_a\dot{\psi}_a - L$. But now $\dot{x}^a = -i(p_a + V_a) = -i(D_a + V_a)$. Substitute it into the definition of H and assume that in L there is a term $\frac{1}{4}R_{ijkl}\psi^i\psi^j\psi^k\psi^l$ (it is zero when the anticommutative property of ψ is incorporated). Then the Hamiltonian $H = -\frac{1}{2}(D_a + V_a)^2 + \frac{1}{8}R + V$ obtained.[11]

At present, it seems to us that there is no longer supersymmetry existing in the Lagrangian (3.11). Nevertheless, we can still use the formula

$$\operatorname{tr}\int [dx\, d\psi]\exp\Big(-\int_0^\beta d\tau\, L_E\Big)$$

to calculate the family index. First, let us consider the simpler case when E is not present. Under the change of parameters $dy^\alpha \to (1/\sqrt{\pi i})\, dy^\alpha$ in L_E, the corresponding result is Ch(Ind D) (see Ref. 2). It can be seen that, in V_a,

$$\langle S(\partial_i)\partial_j f_a\rangle = \Gamma_{ij}^\gamma g_{\gamma\alpha} = -g_{ik}\Gamma_{j\alpha}^k,$$
$$\langle S(\partial_i) f_\alpha f_\beta\rangle = g_{ij}\Gamma_{\beta\alpha}^j.$$

Here connections are defined as in Sec. II. Therefore we have the following Euclidean Lagrangian:

$$\begin{aligned} L_E &= -L(-i\tau) \\ &= \tfrac{1}{2}g_{ij}\dot{x}^i\dot{x}^j + \tfrac{1}{2}g_{ij}\psi^i(\dot{\psi}^j + \Gamma_{kl}^j\dot{x}^k\psi^l) + V_i\dot{x}^i, \qquad (3.5)\end{aligned}$$
$$V_i = -(\sqrt{i/2\pi\beta})g_{ik}\Gamma_{j\alpha}^k\psi^j\, dy^\alpha + (1/4\pi\beta i)g_{ij}\Gamma_{\beta\alpha}^j\, dy^\alpha\, dy^\beta.$$

When $\beta\to 0$, we use the tangent space approximation, i.e., we expand fields around a constant configuration. Suppose the reasoning used by Goodman in Ref. 6 is known. Then the expansions and the path integral measure are

$$\begin{aligned} x^i &= x_0^i + \sqrt{\beta}\sum x_n^i\exp\Big(\frac{i2\pi n\tau}{\beta}\Big), \\ \psi^i &= \Big(\frac{i}{2\pi\beta}\Big)^{1/2}\Big[\psi_0^i + \sqrt{\beta}\sum\psi_n^i\exp\Big(\frac{i2\pi n\tau}{\beta}\Big)\Big], \qquad (3.6)\\ [dx\, d\psi] &= \prod_{i,n\geqslant 1}[dx_0^i\, d\psi_0^i][2\pi n\, dx_n^i\, dx_{-n}^i d\psi_n^i d\psi_{-n}^i]. \end{aligned}$$

Under the expansions, the free part of the action is

$$\sum_{n\geqslant 1} 2\pi^2 n^2 x_{-n}^i x_n^i A - n\psi_{-n}^i\psi_n^i + O(\beta), \qquad (3.7)$$

and the interacting part of the Lagrangian is

$$\begin{aligned} &\frac{i}{4\pi\beta}[\Gamma_{kj,l}^i\psi_0^i\psi_0^j + \Gamma_{\alpha\beta,l}^k\, dy^\alpha\, dy^\beta - 2\Gamma_{i\alpha,l}^k\psi_0^i\, dy^\alpha]x^l\dot{x}^k \\ &\quad + \Big(\frac{i}{2\pi\beta}\Big)^{1/2}\Gamma_{i\alpha}^k\, dy^\alpha\psi^i\dot{x}^k + O\Big(\frac{1}{\sqrt{\beta}}\Big). \qquad (3.8)\end{aligned}$$

In the bracket all quantities are taken at point x_0. After integration, the contribution of $x^l\dot{x}^k$ is asymmetric in indices l,k, and the part proportional to $1/\beta$ tends to zero when $\beta\to 0$. So the interacting part of the action is

$$\begin{aligned} &\frac{i}{4\pi}\Big[\frac{1}{2}R_{lkij}\psi_0^i\psi_0^j + \frac{1}{2}(\Gamma_{\alpha\beta,l}^k - \Gamma_{\alpha\beta,k}^l) \\ &\quad\times dy^\alpha\, dy^\beta - (\Gamma_{i\alpha,l}^k - \Gamma_{i\alpha,k}^l)\psi_0^i\, dy^\alpha\Big] \\ &\quad\times\Big(\sum_n 2\pi n i x_n^k x_{-n}^l\Big) - \Gamma_{i\alpha}^k\, dy^\alpha\sum_n n x_n^k\psi_{-n}^i, \qquad (3.9)\end{aligned}$$

where the form (2.7) of g_{ij} is used. Note that here modes ψ_n^i are not decoupled in interaction, so in the path integral we first integrate these modes. Using measure $\Pi_{i,n>0}[d\sqrt{n}\,\psi_n^i\psi_{-n}^i]$ and Gaussian integrand

$$\exp\Big(\sum_{n>0} n\psi_{-n}^i\psi_n^i + \Gamma_{i\alpha}^k\, dy^\alpha\sum_n n x_n^k\psi_{-n}^i\Big),$$

we integrate modes ψ_n^i, yielding a term in effective action,

$$-\frac{i}{4\pi}\Gamma_{i[\alpha}^l\Gamma_{i\beta]}^k\, dy^\alpha\, dy^\beta\Big(\sum_n 2\pi n i x_n^k x_{-n}^l\Big). \qquad (3.10)$$

We then have the effective action at the limit of $\beta\to 0$,

$$\frac{i}{4\pi}\left[\frac{1}{2} R_{lkij}\psi_0^i\psi_0^j + \frac{1}{2} R_{lk\alpha\beta}\, dy^\alpha\, dy^\beta - R_{lki\alpha}\psi_0^i\, dy^\alpha\right]\left(\sum_n 2\pi n i x_n^k x_{-n}^l\right), \tag{3.11}$$

where we have used formulas, presented in Sec. II, of curvatures at x_0 in normal coordinates. Note that the term in the bracket of the above equation is not the curvature of the connection ∇ restricted to TM. But we can do the transformation $\psi_0^i \to -\psi_0^i$ in the path integral measure, as then the sign of term $R_{lki\alpha}$ is reversed. Further, remember that for the Grassmanian integral, we have the formula $\int da\, f(a) = \int da\, f(a+b)$, where b is an a-number. We shift ψ_0^i by an amount $-f_\alpha^i\, dy^\alpha$, and then the term in the bracket in Eq. (3.18) becomes

$$R_{lk} = \tfrac{1}{2}(R_{lkij}X^iX^j + R_{lk\alpha\beta}\, dy^\alpha\, dy^\beta + 2R_{lki\alpha}X^i dy^\alpha)$$

and $X^i = \psi_0^i - f_\alpha^i\, dy^\alpha$. Because the measure of zero modes is $[dx_0^i\, d\psi_0^i]$, the integral may replaced by a form integrated over fiber M, provided we replace ψ_0^i formally by dx^i. Finally, we perform an integration of modes x_n^i obtaining the formula of family index theorem,

$$\mathrm{Ch}(\mathrm{Ind}\, D) = \int \hat{A}\left(\frac{R}{4\pi}\right),$$

$$\hat{A}\left(\frac{R}{4\pi}\right) = \left[\det\left(\frac{iR}{4\pi}\left(\sinh\frac{iR}{4\pi}\right)^{-1}\right)\right]^{1/2}. \tag{3.12}$$

Next we come to the case with a gauge field present. We use Eq. (3.10). Now we have an additional term in the Euclidean Lagrangian:

$$\bar{\eta}A_i\eta\dot{x}^i - \tfrac{1}{2}\bar{\eta}F_{ij}\psi^i\psi^j + (1/4\pi\beta i)\bar{\eta}F_{\alpha\beta}\, dy^\alpha\, dy^\beta + (\sqrt{i/2\pi\beta})\bar{\eta}F_{i\alpha}\psi^i\, dy^\alpha.$$

Here the transformation $dy^\alpha \to (\sqrt{1/\pi i})dy^\alpha$ is performed and the term $\bar{\eta}A_i\eta\dot{x}^i$ is due to covariant derivative D_i. As shown by Goodman,[6] normal coordinates can be chosen such that around x_0, $A_i = -\tfrac{1}{2}F_{ij}(X_0)x^j + O(x^2)$. Therefore $\bar{\eta}A_i\eta\dot{x}^i = \tfrac{1}{2}\bar{\eta}F_{ij}\eta x^i\dot{x}^j + O(\beta)$. This term does not contribute when $\beta\to 0$. After substituting expansions (3.13), the additional term in action reads

$$(1/2\pi i)\bar{\eta}\left[\tfrac{1}{2}F_{ij}\psi_0^i\psi_0^j + \tfrac{1}{2}F_{\alpha\beta}\, dy^\alpha\, dy^\beta - F_{i\alpha}\psi_0^i\, dy^\alpha\right]\eta. \tag{3.13}$$

Perform transformations $\psi_0^i \to -\psi_0^i$ and $\psi_0^i \to \psi_0^i - f_\alpha^i\, dy^\alpha$, and the term in the bracket of the above equation is transformed to be

$$F = \tfrac{1}{2}F_{ij}X^iX^j + \tfrac{1}{2}F_{\alpha\beta}\, dy^\alpha\, dy^\beta + F_{i\alpha}X^i\, dy^\alpha.$$

The trace of

$$\exp(-(1/2\pi i)\bar{\eta}F\eta)$$

is

$$\mathrm{tr}\exp(-(1/2\pi i)F) = \mathrm{Ch}(F).$$

So, we reach the family index formula

$$\mathrm{Ch}(\mathrm{Ind}\, D) = \int \hat{A}\left(\frac{R}{4\pi}\right)\mathrm{Ch}(F). \tag{3.14}$$

IV. THE EQUIVARIANT FAMILY INDEX THEOREM

The equivariant family index theorem can be proved similarly. Before doing so, we mention here that this theorem concerning the Ramond–Dirac operator in loop space has been applied in string theory.[9,12] First, Schellekens and Warner used the family index theorem to calculate mixed anomaly in strings, and then this theorem was proved by Pilch *et al.*[12] While the index theorem of the Ramond–Dirac operator was conjectured by Witten by generalizing the ordinary index theorem in finite dimensions, it was proved by Alvarez *et al.* and Li via a path integral for the supersymmetric nonlinear σ-model.[13] More interestingly, I pointed out[9] that Pilch and Warner's vacuum bundle over some infinite-dimensional manifold[14] is related to the equivariant family index theorem, and one concludes the results in Ref. 12. To proceed further, one needs to calculate corresponding curvatures; indeed our present work is inspired by this purpose. The integral formula presented in this section, to the best of my knowledge, has not appeared in literature before.

Let us calculate the equivariant family index. As in Sec. II, G is an automorphic group acting on fibered space X. Any element $g\in G$ commutes with the Dirac operator and γ_5, so one can consistently define the equivariant family index by $\mathrm{str}\, g\exp(-\tfrac{1}{2}I^\beta)$. It can be easily verified by the same arguments of Ref. 2 that this definition coincides with the usual definition. By this definition, the equivariant version is then cast into a representation of the path integral; the effect of inserting factor g amounts to imposing twisting boundary conditions on dynamical variables. Only those quantities on a fixed point set contribute, since under limit $\beta\to 0$, fluctuations around constant configurations dominate the contribution. As in Sec. II, we choose coordinates x^a for fixed fiber $M(g)$ and coordinates x^r for normal bundle to $M(g)$. On the normal bundle, g can be block diagonalized to rotate pairs x^r and $x^{\bar{r}}$. If we write $X^r = x^r + ix^{\bar{r}}$, then

$$gX^r = e^{i\theta_r}X^r. \tag{4.1}$$

Similarly, since ψ^r is tangential to the normal bundle, we define $\Psi^r = \psi^r + i\psi^{\bar{r}}$, and then $g\Psi^r = \exp(i\theta_r)\Psi^r$. We make the following expansions:

$$X^r = \sqrt{\beta}\sum (a_n^r + ib_n^r)e^{i(2\pi n+\theta_r)\tau/\beta}, \tag{4.2}$$

$$\Psi^r = \sum (c_n^r + id_n^r)e^{i(2\pi n+\theta_r)\tau/\beta}.$$

The part containing X^r and Ψ^r in free action is

$$\sum_n \frac{1}{2}(2\pi n+\theta_r)^2[(a_n^r)^2 + (b_n^r)^2] - (2\pi n+\theta_r)c_n^r d_n^r, \tag{4.3}$$

and the measure is

$$[dx^r\, d\psi^r] = -\frac{1}{2\pi}da_0^r\, db_0^r\, dc_0^r\, dd_0^r \times\prod_{n\neq 0}(-n\, da_n^r\, db_n^r\, dc_n^r\, dd_n^r). \tag{4.4}$$

To perform the path integral, we first integrate modes ψ_n^a and c_n^r, d_n^r out. The term

$$(\sqrt{i/2\pi\beta})\Gamma^b_{a\alpha}\, dy^\alpha\, \psi^a \dot{x}^b \tag{4.5}$$

gives rise to

$$\frac{1}{2}\sum_n \Gamma^c_{a\alpha}\Gamma^b_{a\beta}\, dy^\alpha\, dy^\beta\, x^b_n x^c_{-n}. \tag{4.6}$$

But if θ_r is not of form $2\pi n$, the terms

$$(\sqrt{i/2\pi\beta})dy^\alpha\, \psi^a(\Gamma^r_{a\alpha}\dot{x}^r + \Gamma^{\bar{r}}_{a\alpha}\dot{x}^{\bar{r}}), \tag{4.7}$$

$$(\sqrt{i/2\pi\beta})dy^\alpha(\Gamma^a_{r\alpha}\psi^r + \Gamma^a_{\bar{r}\alpha}\psi^{\bar{r}})\dot{x}^a$$

do not contribute. In fact these terms are zero, according to Sec. II. Another term which will contribute is

$$(\sqrt{i/2\pi\beta})dy^\alpha(\Gamma^r_{s\alpha}\psi^s\dot{x}^r + \cdots), \tag{4.8}$$

and it gives the contribution

$$-(i/4\pi)\left[(\Gamma^r_{s\alpha}\Gamma^{\bar{r}}_{s\beta} + \Gamma^r_{s\alpha}\Gamma^{\bar{r}}_{\bar{s}\beta})dy^\alpha\, dy^\beta x^r x^{\bar{r}} + \cdots\right]. \tag{4.9}$$

Supposing curvature restricted on the normal bundle can be block diagonalized, then the interacting part of the Lagrangian, not including the gauge field, is

$$(i/4\pi\beta)R_{ab}x^a\dot{x}^b + (i/4\pi\beta)(R_{r\bar{r}}x^r\dot{x}^{\bar{r}} + R_{\bar{r}r}x^{\bar{r}}\dot{x}^r), \tag{4.10}$$

where

$$\begin{aligned} R_{ab} &= \tfrac{1}{2}R_{abcd}\psi^c_0\psi^d_0 + \tfrac{1}{2}R_{ab\alpha\beta}\, dy^\alpha\, dy^\beta \\ &\quad - R_{abc\alpha}\psi^c_0\, dy^\alpha, \\ R_{r\bar{r}} &= \tfrac{1}{2}R_{r\bar{r}ab}\psi^a_0\psi^b_0 + \tfrac{1}{2}R_{r\bar{r}\alpha\beta}\, dy^\alpha\, dy^\beta \\ &\quad - R_{r\bar{r}a\alpha}\psi^a_0\, dy^\alpha. \end{aligned} \tag{4.11}$$

The curvature components are defined in Sec. II. We transform ψ^a_0 as in the last section, and then integrate modes x^a_n. With notation $\Omega_r = R_{r\bar{r}}$, after integrating modes x^a_n, x^r_n, we obtain from the action given by Eq. (4.10),

$$\begin{aligned} &\hat{A}\left(\frac{R}{4\pi}\right)\prod_r \frac{1}{2i}\left(\sin\frac{\bar{\theta}_r}{2}\right)^{-1}, \\ &\bar{\theta}_r = \theta_r + \frac{i}{2\pi}\Omega_r, \\ &R = \{R_{ab}\}. \end{aligned} \tag{4.12}$$

To include the gauge field, we assume that g acts on the ath dimension of the fiber space of bundle E by a factor $\exp(-i\theta_a)$, namely,

$$\eta^a \to e^{i\theta_a}\eta^a, \quad \bar{\eta}_a \to e^{-i\theta_a}\bar{\eta}_a. \tag{4.13}$$

Let F be defined as in the last section, and as a matrix it is diagonalized as $F^a_b = F_a\delta^a_b$. On subspace $\{u^a\bar{\eta}_a|0\rangle\}$, g acts like a matrix $g = \exp(-i\theta_a)\delta^a_b$. We have to calculate $\operatorname{tr}\exp(-F/2\pi i)$. Using the above ansätze we obtain

$$\begin{aligned} \operatorname{tr} g \exp\left(-\frac{F}{2\pi i}\right) &= \sum_a \exp\left(-i\theta_a + \frac{iF_a}{2\pi}\right) \\ &= \sum_a \exp(-i\bar{\theta}_a). \end{aligned} \tag{4.14}$$

Combining Eqs. (4.12) and (4.14), we finally reach the equivariant family index theorem,

$$\begin{aligned} &\operatorname{str} g \exp\left(-\frac{I^\beta}{2}\right) \\ &= \int_{M(g)} \hat{A}\left(\frac{R}{4\pi}\right)\prod_r \frac{1}{2i\sin\bar{\theta}_r/2}\sum_a \exp(-i\bar{\theta}_a). \end{aligned} \tag{4.15}$$

V. CONCLUSION

We have shown in this paper how one can prove the family index theorem by the use of a path integral once the Quillen superconnection is recognized. Our proof is not a naive one, for we have assumed a theorem of Bismut. Nevertheless, our proof indicates that perhaps one can start from a supersymmetric quantum system to derive the ingredients we derived from the Hamiltonian given by Bismut. This demands insight to the relationship between the family index theorem and the geometry of determinant bundles.[7,8] However, formulas given in Secs. II and IV will prove helpful in evaluating quantities presented in Ref. 9, in order to calculate anomalies in strings.

ACKNOWLEDGMENTS

I would like to thank Prof. R. Ruffini for his kind hospitality at the Physics Department, Rome University and ICRA. I am also grateful to Prof. L. Z. Fang for his sustained encouragement.

[1] M. F. Atiyah, and I. M. Singer, "The index of elliptic operators. IV," Ann. Math. **93**, 119 (1971).

[2] J. M. Bismut, "The Atiyah–Singer index theorem for families of Dirac operators: Two heat equation proofs," Invent. Math. **83**, 91 (1986).

[3] D. Quillen, "Superconnections and the Chern characters," Topology **24**, 89 (1985); "Determinants of Cauchy–Riemann operators over a Riemann surface," Func. Anal. Appl. **19**, 31 (1985).

[4] E. Witten, "Constraints on supersymmetry breaking," Nucl. Phys. B **202**, 253 (1982).

[5] L. Alvarez-Gaumé, "Supersymmetry and the Atiyah–Singer index theorem," Commun. Math. Phys. **90**, 161 (1983); D. Friedan and P. Windey, "Supersymmetric derivation of Atiyah–Singer index and the chiral anomaly," Nucl. Phys. B **235**, 395 (1984).

[6] M. W. Goodman and E. Witten, "Global symmetries in four and higher dimensions," Nucl. Phys. B **271**, 21 (1986); M. W. Goodman, "Proof of character-valued index theorems," Commun. Math. Phys. **107**, 391 (1986).

[7] J. M. Bismut, and D. S. Freed, "The analysis of elliptic families. I," Commun. Math. Phys. **106**, 159 (1986); II, **107**, 103 (1986).

[8] D. S. Freed, "Determinants, torsion, and strings," Commun. Math. Phys. **107**, 483 (1987); S. Della Pietra, V. Della Pietra, and L. Alvarez-Gaumé, "The chiral determinant and the eta invariant," *ibid.* **109**, 691 (1987); S. Della Pietra and V. Della Pietra, "Parallel transport in the determinant line bundle: The zero index case," *ibid.* **110**, 573 (1987); "The nonzero index case," **111**, 11 (1987).

[9] L. Alvarez-Gaumé, J.-B. Bost, G. Moore, P. Nelson, and C. Vafa, Commun. Math. Phys. **112**, 503 (1987); M. Li, "Equivariant determinant bundles and supersymmetric closed strings," to be published in Phys. Lett. B.

[10] B.-Y. Chen, *Total Mean Curvature and Submanifolds of Finite Type* (World Scientific, Singapore, 1984).

[11] In fact if we do not include the term proportional to the curvature in the Lagrangian, then in the Hamiltonian there is no term $\frac{1}{8}R$. This does not matter because at the limit $\beta\to 0$ this term does not contribute.

[12] A. N. Schellekens, and N. P. Warner, Phys. Lett. B **171**, 317 (1986); "Anomalies, characters and strings," Nucl. Phys. B **287**, 317 (1987); K. Pilch, A. N. Schellenkens, and N. P. Warner, "Path integral calculation of string anomalies," *ibid.* **287**, 362 (1987).

[13] E. Witten, "Elliptic genera and quantum field theory," Commun. Math. Phys. **109**, 525 (1987); "The index of the Dirac operator in loop space," Princeton University preprint PUPT-1050, 1987; O. Alvarez, T. P. Killingback, M. Mangano and P. Windey, "String theory and loop space index theorems," Commun. Math. Phys. **111**, 1 (1987); K. K. Li, "Character-valued index theorems in supersymmetric string theories," MIT preprint, CTP-1460, 1987.

[14] B. Bowick, and S. G. Rajeev, "String theory as the Kahler geometry of loop space," Phys. Rev. Lett. **58**, 535 (1987); Nucl. Phys. B **293**, 348 (1987); K. Pilch and N. P. Warner, "Holomorphic structure of superstring vacua," Class. Quantum Gravit. **4**, 1183 (1987); M. Li, "Holomorphic structure of closed strings: Left moving sector," Phys. Lett. B **208**, 89 (1988).

Diagonalization and Hamiltonian structures of hyperbolic systems

Fahrünisa Neyzi
Department of Physics, Boğaziçi University, Bebek, Istanbul, Turkey

(Received 8 February 1989; accepted for publication 12 April 1989)

A simple algorithm for casting two-dimensional hyperbolic systems into the Hamiltonian form is given. First-order conserved quantities are found and expressed in terms of eigenvalues. It is shown that for some systems the dependence of the eigenvalues on the field variables can be used for classification.

I. INTRODUCTION

Extensive results on Hamiltonian structures, symmetries, and conservation laws have been obtained by Sheftel'[1] and Olver and Nutku[2], which elucidate the connection between the infinite sequence of conserved densities and the Hamiltonian structure for systems of hydrodynamic type with two dependent variables. In order to utilize these results, it is necessary first to formulate the relevant evolution equations as a Hamiltonian system.

In this paper, we present a simple algorithm for finding a transformation of the dependent variables that maps the system into the Hamiltonian form. If such a form already exists, then we may use this method for casting the system into alternative Hamiltonian forms; thereby determining the equivalence class of the system.

To this end, we first diagonalize the system. In the diagonal form, the symmetries of the system are most clearly manifest[3] and as we further note in Sec. II, some first-order conserved quantities, which are related to shock conditions,[4] can be readily singled out. In Sec. III we show that all possible one-to-one variable transformations, which cast a system into the Hamiltonian form, are solutions of one differential equation. As examples, we use the Born–Infeld[5] equation and the dispersionless Kadomtsev–Petviashvili equation, as reduced by Kodama.[6] In Sec. IV we point out that the hyperbolic systems with a very rich structure, like the Poisson equation, are all characterized by eigenvalues that are linear functions of the Riemann invariants of the system.

The variables that put the system into a Hamiltonian form are both solutions of a single equation. In cases where a Hamiltonian exists, all solutions of this equation are conserved quantities. However, a new Hamiltonian structure exists if and only if two solutions with nonvanishing Jacobian can be matched together by satisfying an additional equation.

II. DIAGONALIZATION, AND FIRST-ORDER CONSERVED QUANTITIES RELATED TO SHOCK CONDITIONS

We consider a system of two first-order partial differential equations in the form

$$\begin{aligned} &p_t + Ap_x + Bq_x = 0, \\ &q_t + Cp_x + Dq_x = 0, \end{aligned} \tag{2.1}$$

where A, B, C, and D are functions of p and q. Then following Lax's construction,[4] we find the eigenvalues λ and μ of the matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

We consider only hyperbolic systems, therefore λ and μ are real and distinct. Riemann invariants r and s for the system are the solutions of the following:

$$r_p = [C/(\lambda - A)]r_q, \quad s_p = [C/(\mu - A)]s_q. \tag{2.2}$$

Having found a set of Riemann invariants we can write the system of Eq. (2.1) in the diagonal form:

$$\begin{aligned} &r_t + \lambda r_x = 0, \\ &s_t + \mu s_x = 0. \end{aligned} \tag{2.3}$$

After differentiation of these equations, with respect to x and some manipulation, we obtain

$$\begin{aligned} &\left(\frac{d}{dt} + \lambda \frac{d}{dx}\right) z + \lambda_r e^{-h} z^2 = 0, \\ &\left(\frac{d}{dt} + \mu \frac{d}{dx}\right) \tilde{z} + \mu_s e^{-k} \tilde{z}^2 = 0, \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} &h_s = \lambda_s/(\lambda - \mu), \quad k_r = \mu_r/(\mu - \lambda), \\ &z = e^h r_x, \quad \tilde{z} = e^k s_x. \end{aligned} \tag{2.5}$$

Now Eqs. (2.4) can be compared to the simple initial value problem[4]

$$\frac{dz}{dt} = a(t) z^2, \quad z(0) = m, \tag{2.6}$$

in the interval $(0,T)$. If $z(t)$ is a solution of this problem, for any A satisfying $0 < A < a(t)$, the duration for the solution, is less than $(mA)^{-1}$. Analogously, for Eq. (2.4), a solution cannot be continued beyond one of the following:

$$(-\lambda_r(0) \max r_x(0))^{-1}, \quad (-\mu_s(0) \max s_x(0))^{-1}. \tag{2.7}$$

Now if we define

$$\begin{aligned} &M = e^{-h}/z, \\ &N = e^{-k}/\tilde{z}, \end{aligned} \tag{2.8}$$

after some manipulation Eq. (2.4) reduces to

$$(M \pm N)_t + (\mu M \pm \lambda N)_x = 0, \tag{2.9}$$

if and only if

 0022-2488/89/081695-04$02.50

$$\lambda_r e^{-2h} \pm \mu_s e^{-2k} = 0. \tag{2.10}$$

Therefore, for any hyperbolic system that satisfies Eq. (2.10), there exists a conserved density

$$e^{-2h}/r_x \pm e^{-2k}/s_x. \tag{2.11}$$

We have derived this first-order density directly from Eq. (2.4), which, as we have seen, also determines the duration of stability of the system. Therefore the conserved densities of this kind[7] and shock conditions are intimately related.

III. FROM DIAGONAL TO HAMILTONIAN SYSTEMS

A two-component hyperbolic system is in the Hamiltonian form if it can be formulated as

$$\begin{pmatrix} u \\ v \end{pmatrix}_t + \begin{pmatrix} 0 & D_x \\ D_x & 0 \end{pmatrix} \begin{pmatrix} \partial_u \\ \partial_v \end{pmatrix} \mathscr{H}(u,v) = 0, \tag{3.1}$$

where ∂_u and ∂_v indicate partial derivatives and D_x is a total derivative operator with respect to x. Here, $H(u,v)$ is the Hamiltonian density, such that through

$$H = \int \mathscr{H}(u,v)dx \tag{3.2}$$

it defines the Hamiltonian energy. If a general hyperbolic system, defined in Eq. (2.1), can be reduced to this Hamiltonian form, then it is possible to examine other properties of the system, like the Lax operator representation[8] or the existence of infinite series of commuting Hamiltonians.[1] Furthermore, Sheftel'[1] has shown that starting from the Hamiltonian form, a recurrence operator can be constructed and its powers generate an entire series of Hamiltonians.

Here we suggest a method for passing from a general form, Eq. (2.1), to the Hamiltonian form, Eq. (3.1), via the diagonal form, Eq. (2.3). The reason for following this route can be found in Sheftel's work,[3] where it is shown that the group analysis is maximally simplified if the system of type (2.1) is expressed in the diagonal form.

Therefore, having passed to the diagonal form as suggested in Sec. II, we note that the two independent solutions of the partial differential equation

$$-u_{rs} = [\lambda_s/(\lambda - \mu)]u_r + [\mu_r/(\mu - \lambda)]u_s, \tag{3.3}$$

will give us the desired transformation. Then, expressing the eigenvalues λ and μ as a function of the solutions of (3.3), u and, say v, we can construct the derivatives of the Hamiltonian density $\mathscr{H}(u,v)$:

$$\begin{aligned} \mathscr{H}_{uv} &= (\lambda + \mu)/2, \\ \mathscr{H}_{vv} &= (\lambda - \mu)/2V, \\ \mathscr{H}_{uu} &= [(\lambda - \mu)/2]V, \end{aligned} \tag{3.4}$$

where

$$\frac{-V_r}{2V}\left(\frac{v_s}{v_r}\right) = \frac{\lambda_s}{\lambda - \mu}, \quad \frac{-V_s}{2V}\left(\frac{v_r}{v_s}\right) = \frac{\mu_r}{\mu - \lambda}, \tag{3.5a}$$

$$v_r = Vu_r, \quad v_s = -Vu_s, \tag{3.5b}$$

$$v_r u_s + u_r v_s = 0. \tag{3.5c}$$

It can easily be seen that Eq. (3.3) is just the integrability condition of these equations.

Different sets of solutions of Eq. (3.3) will allow us to construct different possible Hamiltonian forms for a hyperbolic system. As an example, let us first consider the Born–Infeld equation. It has been shown[5] that this system is equivalent to

$$\begin{aligned} p_t - (p^{-2} + q^{-2})p_x + 2pq^{-3}q_x &= 0, \\ q_t + 2qp^{-3}p_x - (p^{-2} + q^{-2})q_x &= 0. \end{aligned} \tag{3.6}$$

This is already in the Hamiltonian form, with the Hamiltonian density given by

$$\mathscr{H} = -(pq^{-1} + qp^{-1}). \tag{3.7}$$

When we diagonalize the system, we find that the Riemann invariants are

$$\begin{aligned} r &= f(p^{-1} - q^{-1}), \\ s &= g(p^{-1} + q^{-1}), \end{aligned} \tag{3.8}$$

in terms of which Eq. (3.6) reduces to

$$r_t - G^2(s)r_x = 0,$$
$$s_t - F^2(r)s_x = 0,$$

where

$$\begin{aligned} F &= f^{-1}, \\ G &= g^{-1}. \end{aligned} \tag{3.9}$$

Then substituting these values for the eigenvalues, in terms of the new variables in Eq. (3.3), we obtain

$$u_{FG} = [2/(F^2 - G^2)](Gu_F - Fu_G). \tag{3.10}$$

Two obvious solutions are

$$u = (F+G)^{-1}, \quad v = (F-G)^{-1}, \tag{3.11}$$

which lead us back to the original form (3.6). Another solution set for Eq. (3.10) is easily found. Since uv is always a solution, we try to find its partner through Eq. (3.5c) and this yields

$$\tilde{u} = (F+G)^{-1}(F-G)^{-1}, \quad \tilde{v} = F^2 + G^2. \tag{3.12}$$

In terms of these variables we get

$$\begin{aligned} \lambda &= -s^2 = -(\tilde{u} - \tilde{v}^{-1}), \\ \mu &= -r^2 = -(\tilde{u} + \tilde{v}^{-1}), \\ V &= -\tilde{v}^2. \end{aligned} \tag{3.13}$$

Therefore using Eq. (3.4), we obtain the following system:

$$\begin{aligned} \tilde{u} - \tilde{u}\tilde{u}_x - \tilde{v}^{-3}\tilde{v}_x &= 0, \\ \tilde{v}_t - \tilde{v}\tilde{u}_x - \tilde{u}\tilde{v}_x &= 0. \end{aligned} \tag{3.14}$$

These are the equations of a Chaplygin gas.[5] This transformation has been found by Verosky. Here we have used it as an example to show that transformations to new Hamiltonian forms can be obtained directly from Eq. (3.10).

As our second example, we use the system that has been derived by Kodama as a reduced version of the dispersionless Kodomtsev–Petviashvili equation.[6] One other reduced version is the classical shallow water equation, which has a hierarchy of Hamiltonian forms and conserved quantities. Therefore we check to see if this alternate form is also a Hamiltonian. The Kodama equation is

$$\begin{aligned} p_t &= pp_x + q_x, \\ q_t &= (q-p)p_x. \end{aligned} \tag{3.15}$$

Going back to Eq. (2.2), we find its Riemann invariants:

$$r = p - \chi^{1/2} - 2\ln(\chi^{1/2} - 2),$$
$$s = p + \chi^{1/2} - 2\ln(\chi^{1/2} + 2),$$

where

$$\chi = p^2 + 4(q - p). \tag{3.16}$$

In terms of the Riemann invariants, the eigenvalues turn out to be

$$\lambda = -\tfrac{1}{2}r - \ln(F-2),$$
$$\mu = -\tfrac{1}{2}s - \ln(F+2),$$

with

$$F \equiv f^{-1}, \quad y = s - r, \qquad f(y) = 2y + 2\ln[(y-2)/(y+2)]. \tag{3.17}$$

Substituting in Eq. (3.3) to obtain the transformation that will map the system into the Hamiltonian form, we get

$$u_{rs} = [F'/F(F-2)]u_r + [F'/F(F+2)]u_s, \tag{3.18}$$

where the prime indicates a derivative with respect to $(s-r)$. One solution can be found and is given by

$$u = \int \left(\frac{F^2 - 4}{F^2}\right)^{1/2} dy. \tag{3.19}$$

However, its companion v does not exist because Eq. (3.5c) indicates that v must be a function of $(r+s)$. Therefore we conclude that we cannot cast the Kodama system into the Hamiltonian form using the algorithm presented in this section.

IV. CLASSIFYING MULTI-HAMILTONIAN SYSTEMS

In this section we build on the results of Nutku[9] and Olver and Nutku[2] to classify, through the material in the previous two sections, some hyperbolic systems that are endowed with a very rich structure.

Multi-Hamiltonian structure is a generalization of the elementary Hamiltonian structure, in that we can express Eq. (3.1) in a vector form,

$$\bar{u}_t + \mathbf{D}_0 E[H] = 0, \tag{4.1}$$

where

$$\mathbf{D}_0 = \begin{pmatrix} 0 & D_x \\ D_x & 0 \end{pmatrix} \tag{4.2}$$

is the Hamiltonian operator and $E = E_{\bar{u}}$ denotes the Euler operator. Now a system admits multi-Hamiltonian structure if it is possible to express it as

$$\bar{u}_t + \mathbf{D}_1 E[H'] = 0, \tag{4.3}$$

where $\mathbf{D}_1$ is a skew-symmetric operator and the Hamiltonians H and H' are compatible with respect to Poisson brackets determined by $\mathbf{D}_0$ and $\mathbf{D}_1$:

$$\{H,H'\}_i = \int E[H]\mathbf{D}_i E[H']dx, \quad i = 0,1.$$
$$= 0. \tag{4.4}$$

The large class of one-dimensional hyperbolic systems that exhibit this structure are discussed in Ref. 2. Here we show that these special systems, namely equations of gasdynamics, the Born–Infeld equation, Poisson equation, and another system that is mentioned as "the curious choice" in Ref. 2, are all seen to be related when formulated in their diagonal form.

Consider a very simple case of Eq. (2.3), where λ and μ are linear in r and s,

$$r_t + (ar + bs + c)r_x = 0, \qquad s_t + (br + as - c)s_x = 0, \tag{4.5}$$

where a, b, and c are constants. Following the procedure of Sec. III, we try to put this system into Hamiltonian form. First let us consider

$$c = 0, \quad a \neq b \neq 0. \tag{4.6}$$

Then Eq. (3.3) reduces to

$$-u_{rs} = [b/(a-b)(r-s)](u_r - u_s). \tag{4.7}$$

There exists a general solution for this second-order partial differential equation.[10] One simple solution set for (u,v) is

$$u = k(r+s), \qquad v = l(r-s)^\alpha, \quad \alpha = (a+b)/(a-b), \tag{4.8}$$

where k and l are arbitrary constants. We find V using Eq. (3.5) and this yields the following second derivatives for the Hamiltonian:

$$H_{uv} = [(a+b)/2k]u,$$
$$H_{uu} = [(a+b)/2k]v, \tag{4.9}$$
$$H_{vv} = [(a-b)^2/2(a+b)]kl^{2(a-b)(a+b)^{-1}} \times v^{(a-3b)(a+b)^{-1}}.$$

Choosing the values of the arbitrary constants k and l to be

$$k = \tfrac{1}{2}, \quad l = [4/(\gamma-1)]^{2(1-\gamma)^{-1}}, \qquad a = \tfrac{1}{4}(1+\gamma), \quad b = \tfrac{1}{4}(3-\gamma), \tag{4.10}$$

the Hamiltonian form for Eq. (4.5), under the restrictions defined by Eq. (4.6), turns out to be

$$u_t + uu_x + v^{\gamma-2}v_x = 0, \qquad v_t + vu_x + uv_x = 0. \tag{4.11}$$

This is the gasdynamics system.

Now let us turn our attention to the constraints defined by Eq. (4.6) and see what is revealed when these constraints are relaxed. If $a = 0$, then $\gamma = 1$. This case yields the Born–Infeld equation. If $b = 0$, then $\gamma = 3$. Then the system is decoupled if expressed in terms of r and s. For $a = -b$, Eq. (4.7) yields the following for the new variables (u,v):

$$u = k(r+s), \qquad v = l\ln(s-r). \tag{4.12}$$

Then, when k and l are chosen to be 2 and a is taken to be 1 for convenience, we have

$$\begin{pmatrix} u \\ v \end{pmatrix}_t + \begin{pmatrix} 0 & e^v \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}_x = 0. \tag{4.13}$$

As noted in Ref. 2 this system, though not part of the gasdynamics family, has an especially rich Hamiltonian structure.

Finally, let us consider the case where a equals b. If c is still zero then our system is not hyperbolic, since the eigenvalues λ and μ are equal. Still we can solve the modified form of Eq. (3.3),

$$\lambda_s u_r + \mu_r u_s = 0,\quad \lambda_s v_r + \mu_r v_s = 0, \tag{4.14}$$

with

$$\lambda_s = \mu_r = a. \tag{4.15}$$

Then we can use the simplest solution to obtain the system

$$u = r + s,\quad v = r - s,\quad u_t + u u_x = 0,\quad v_t + u v_x = 0. \tag{4.16}$$

Now let us consider the case, where c does not equal zero but a equals b. Then Eq. (3.3) becomes

$$u_{rs} = (a/2c)(u_s - u_r), \tag{4.17}$$

the two simple solutions of which we identify as u and v;

$$u = a(r+s),\quad v = e^{a(r-s)/c}. \tag{4.18}$$

In terms of these variables the system is now very familiar in the Hamiltonian form

$$\begin{pmatrix} u \\ v \end{pmatrix}_t + \begin{pmatrix} u & v^{-1} \\ v & u \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}_x = 0. \tag{4.19}$$

This is the Poisson equation.[11]

Finally let us go back to Eq. (4.11). Its Riemann invariants are

$$r = f([(\gamma - 1)/2]u + v^{(\gamma-1)/2}),\quad s = g([(\gamma - 1)/2]u - v^{(\gamma-1)/2}). \tag{4.20}$$

Here f and g are arbitrary functions. Defining their inverse functions as F and G, respectively, we have, using Eq. (4.7),

$$[2(\gamma-1)/(3-\gamma)](F-G)u_{FG} = u_G - u_F. \tag{4.21}$$

The solution that takes us back to Eq. (4.11), which, of course, is already in the Hamiltonian form, turns out to be the simplest solution,

$$\frac{(F+G)}{(\gamma-1)} = \alpha,\quad [(F-G)/2]^{2/(\gamma-1)} = \beta. \tag{4.22}$$

Here α and β are used instead of u and v to avoid confusion. In terms of these variables, Eq. (4.21) becomes

$$u_{\alpha\alpha} = u_{\beta\beta}\beta^{3-\gamma}. \tag{4.23}$$

All conserved densities of zeroth order are readily seen to be a solution when we rename

$$H_{uu} = H_{vv}v^{3-\gamma}. \tag{4.24}$$

Therefore we conclude that all the solutions of Eq. (4.21) are conserved quantities when expressed in terms of the "primary" variable set, which gives the Hamiltonian form. There may be an infinite number of them. However, the system may still be cast into the diagonal form only in one set, because each solution u must be matched with another solution, say v that, in addition, has to satisfy Eq. (3.5c) in order to define an alternative Hamiltonian structure.

ACKNOWLEDGMENTS

I would like to thank M. Arik and Y. Nutku for their interest and help.

[1] M. B. Sheftel', Funct. Anal. Appl. **20**, 227 (1986).
[2] P. J. Olver and Y. Nutku, J. Math. Phys. **29**, 1610 (1988).
[3] M. B. Sheftel', Theor. Math. Phys. **56**, 878 (1983).
[4] P. D. Lax, J. Math. Phys. **5**, 611 (1964).
[5] M. Arik, F. Neyzi, Y. Nutku, P. J. Olver, and J. M. Verosky, submitted to J. Math. Phys.
[6] Y. Kodama, Prog. Theor. Phys. Suppl. **94**, 184 (1988).
[7] J. M. Verosky, J. Math. Phys. **27**, 3061 (1986).
[8] B. A. Dubrovin and S. A. Novikov, Sov. Math. Dokl. **27**, 665 (1983).
[9] Y. Nutku, J. Math. Phys. **28**, 2579 (1987).
[10] E. T. Copson, *Partial Differential Equations* (Cambridge U. P., New York, 1975).
[11] Y. Nutku, J. Math. Phys. **26**, 1237 (1988).

On the Cauchy problem for Yang–Mills equations with external current

Z. Świerczyński
Institute of Physics, Jagellonian University, Reymonta 4, 30-059 Kraków, Poland

(Received 2 May 1988; accepted for publication 22 February 1989)

The Cauchy problem for Yang–Mills equations with the external current in $(3+1)$-dimensional space-time is considered. The formulation in which only the spatial part of the current four-vector is given is studied. The global existence theorem is proved.

I. INTRODUCTION

In this paper we shall consider the Cauchy problem for classical Yang–Mills equations with the external current in $(3+1)$-dimensional space-time. We would like to begin with a brief comparison with the initial data problem in classical electrodynamics concentrating on the existence and uniqueness of the solution. It can be proved that if the initial data and the external current satisfy appropriate regularity conditions (which are not very restrictive) then Maxwell's equations have a solution if and only if the external current satisfies the continuity equation. Thus we obtain a simple criterion that enables us to decide whether the theory is consistent. In the non-Abelian theory the situation is more complicated. The covariant "continuity equation" involves gauge potentials. Therefore we could not regard it as a condition imposed on the current four-vector alone. To avoid this difficulty we will assume that only the spatial part of the current is given while the charge density is determined from the non-Abelian Gauss' law. It is easy to see that in electrodynamics the continuity equation determines the charge density uniquely, provided that its initial value and space components of the current are known. Thus in this case our formulation reduces to the usual one. Also the regularity conditions, which have to be imposed on the fields and the external current, are stronger than in the Abelian case. Since Yang–Mills equations are nonlinear the functional space we use should be chosen in such a way that the multiplication of functions can be defined.

Now we would like to discuss the problem of uniqueness of the solution. The Maxwell theory can be formulated entirely in terms of field strengths. The advantage of this description consists of involving gauge invariant quantities. However, this description is not interesting for us, since it does not possess any simple generalization to the non-Abelian theory. Therefore we shall use the potentials. Since the potentials are not gauge invariant quantities the solution is not uniquely determined by the initial data (one can always perform a gauge transformation that does not change the initial data). We may eliminate the gauge equivalent solutions by imposing a constraint (gauge condition). It is sufficient if it is chosen in such a way that each gauge transformation preserving the constraint and the additional data is time independent, for example, $A_0 = 0$ or $\partial_\mu A^\mu = 0$. In the non-Abelian gauge theory the situation is somewhat different. Since the covariant derivatives of the field strength tensor are not gauge invariant quantities, the presence of the external current causes some of the gauge transformations not to be symmetries of the system. Thus by performing a gauge transformation of the solution of the Yang–Mills equations we generally obtain a solution with different external current. On the other hand, the equations admit the existence of solutions having the same initial data but not related by a gauge transformation. To ensure uniqueness of the solution we shall use the constraint $A_0 = 0$. However, unlike in the Maxwell theory it cannot generally be considered as a gauge fixing condition since it is impossible to satisfy it by a suitable choice of gauge. Using a gauge transformation that changes the external current [see Eqs. (3) in the next section] we can associate with each of the removed solutions the solution with $A_0 = 0$ but corresponding to a different external current.

It follows from the above discussion that the model we shall consider is described by the spatial part of the Yang–Mills equations, i.e., the non-Abelian Ampère's law. Replacing the time derivative of the potential by the electric field (which is correct since $A_0 = 0$) we can write it as a system of equations of the first order in time. The Yang–Mills field at any time is described by specifying the values of the potential and the electric field. For convenience we will assume that the initial data are given at $t = 0$. In this paper we prove that this Cauchy problem has a unique, global (i.e., defined for all $t \geqslant 0$) solution. If the external current vanishes, the charge density, defined here as the covariant divergence of the electric field, becomes time independent (because $A_0 = 0$) and may be treated as a given function. Thus we obtain the Yang–Mills equations with a static source, which were considered in many papers (see Ref. 1 and references therein). In our approach, however, the Yang–Mills equations are supplemented by the constraint $A_0 = 0$.

The proof presented in this paper is a modification of the proof given in Ref. 2. The first part of the proof is based on Segal's general existence theory for semilinear evolution equations.[3] Using this theory we show that there always exists a solution defined on some finite time interval. Next we estimate the norm of the solution and thus establish its global existence. Discussing the regularity properties of the solution we have used the methods of Refs. 4 and 5. For the reader's convenience we shall utilize notation similar to that of Ref. 2.

II. SOLUTIONS FOR SMALL TIME INTERVALS

In our approach (i.e., $A_0 = 0$, j_0 determined from Gauss' law) the Yang–Mills field with the external current is described by the following system of equations:

 0022-2488/89/081699-08$02.50

$$\frac{\partial}{\partial t} A_k = E_k, \tag{1a}$$

$$\frac{\partial}{\partial t} E_k = D_l F_{lk} + j_k. \tag{1b}$$

Here A_k, $k = 1,2,3$, are gauge potentials, $E_k = F_{0k}$ is an electric field, j_k is the external current, $D_l = \partial_l + [A_l,\cdot]$ is a covariant derivative, and F_{lk} is the spatial part of the field strength tensor:

$$F_{lk} = \partial_l A_k - \partial_k A_l + [A_l, A_k]. \tag{1c}$$

We will assume that the gauge group G is the matrix group $\mathrm{Gl}_R(n)$ or $\mathrm{Gl}_c(n)$. The fields A_k, E_k, and F_{lk} and external current j_k are Lie-algebra-valued functions on Minkowski space-time. Equation (1a) relates E_k to the gauge potentials A_k, (1b) is the spatial part of the Yang–Mills equations (non-Abelian Ampère's law). The non-Abelian Gauss' law has the form

$$D_k E_k = j_0, \tag{1d}$$

where j_0 is the color charge density. As we have explained in the Introduction, the quantities j_k and j_0 are treated in quite different manners. The external current j_k is an *a priori* given function while the charge density j_0 is defined by Eq. (1d). Taking the covariant three-dimensional divergence of (1b) and using (1c) we obtain the covariant "continuity equation"

$$\frac{\partial}{\partial t} j_0 = D_k j_k. \tag{1e}$$

Equations (1a)–(1e) may be written in Lorentz covariant form:

$$\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] = F_{\mu\nu}, \tag{2a}$$

$$D^\mu F_{\mu\nu} = -j_\nu, \tag{2b}$$

$$D^\mu j_\mu = 0, \tag{2c}$$

where $D_0 \equiv \partial/\partial t$ and $A_0 = 0$. In Minkowski space we use the metric tensor of signature $(-+++)$. The gauge symmetry of (2) [and (1)] is restricted by the presence of the external current. However, if we choose a smooth, G-valued function $\mathscr{U}$ and define the transformed fields

$$\begin{aligned} \hat{A}_\mu &= \mathscr{U} A_\mu \mathscr{U}^{-1} + \mathscr{U}\, \partial_\mu \mathscr{U}^{-1}, \\ \hat{F}_{\mu\nu} &= \mathscr{U} F_{\mu\nu} \mathscr{U}^{-1}, \\ \hat{j}_0 &= \mathscr{U} j_0 \mathscr{U}^{-1}, \end{aligned} \tag{3}$$

then they will satisfy the Yang–Mills equations with the transformed external current

$$\partial_\mu \hat{A}_\nu - \partial_\nu \hat{A}_\mu + [\hat{A}_\mu, \hat{A}_\nu] = \hat{F}_{\mu\nu}, \tag{4a}$$

$$\hat{D}^\mu \hat{F}_{\mu\nu} = -\hat{j}_\nu, \tag{4b}$$

$$\hat{D}^\mu \hat{j}_\mu = 0, \tag{4c}$$

where $\hat{j}_k = \mathscr{U} j_k \mathscr{U}^{-1}$, $\hat{D}_\mu = \partial_\mu + [\hat{A}_\mu,\cdot]$ [this is true even if the gauge potential in (2) does not satisfy the constraint $A_0 = 0$]. The transformation (3) may be regarded as a symmetry of (2) only if it fulfills the condition $\hat{j}_k \equiv \mathscr{U} j_k \mathscr{U}^{-1} = j_k$. We will use transformation (3) in our considerations concerning the global existence of the solution.

Follow in Ref. 2 we split the electric field present on the right-hand side of Eq. (1a) into the longitudinal and transverse parts

$$E_k^L = \partial_k \{ -(1/4\pi r) * \partial_l E_l \}, \tag{5a}$$

$$E_k^T = (\delta_{kl}\, \partial_m \partial_m - \partial_k \partial_l) \{ -(1/4\pi r) * E_l \}. \tag{5b}$$

Here the star denotes the convolution

$$\left(\frac{1}{4\pi r} * \rho\right)(x) = \int \frac{1}{4\pi|x-y|} \rho(y) d^3y. \tag{6}$$

We modify (1b) by substituting into (5a) for $\partial_l E_l$ the expression obtained from (1d). The system of equations (1a), (1b), (1d), and (1e) becomes

$$\frac{\partial}{\partial t} A_k = E_k^T + \partial_k \left\{ -\frac{1}{4\pi r} * (j_0 + [E_l, A_l]) \right\}, \tag{7a}$$

$$\frac{\partial}{\partial t} E_k = D_l F_{lk} + j_k, \tag{7b}$$

$$\frac{\partial}{\partial t} j_0 = D_l j_l, \tag{7c}$$

$$D_l E_l - j_0 = 0. \tag{7d}$$

The reason for replacing (1) with (7) is given in Ref. 2.

Now we shall prove the local existence and uniqueness of the solution of the Cauchy problem. Applying Segal's general theory of semilinear evolution equations[3] we shall show that for any sufficiently regular external current j_k and initial values $(A(0,\cdot) E(0,\cdot) j_0(0,\cdot))$ Eqs. (6a)–(6c) have a unique solution defined on some time interval $[0,T)$. Clearly, the choice of $t=0$ as the instant for which the initial conditions are specified does not limit the generality of our considerations. Since the constraint (7d) is conserved in time each solution of Eqs. (7a)–(7c) with the Cauchy data satisfying (7d) is also a solution of Eqs. (1). In the subsequent considerations time plays a different role than spatial variables; the fields are treated as a functions of time with values lying in some Hilbert space.

We denote by H_r the Hilbert space of Lie-algebra-valued functions on R^3 quadratically integrable together with their first r derivatives,

$$H_r = \left\{ f\colon\ \|f\|_{H_r} = \left(\sum_{p=0}^{r} \sum_{k=1,2,3} \int_{R^3} \mathrm{Tr}(\partial_{k_1} \cdots \partial_{k_p} f)^2 d^3x \right)^{1/2} < \infty \right\},$$

where the operator Tr differs by a minus sign from the usual one. Since elements of the Lie algebra are anti-Hermitian this convention implies that $\mathrm{Tr}\, X^2 \geqslant 0$ for any X belonging to the Lie algebra of G. It is easy to see that $f \in H_r$ is equivalent to the statement that each matrix component of f belongs to the usual Sobolev space H_r. Next we form the Hilbert space

$$\mathscr{H}_r = \{(A,E,j_0):\ A_k \in H_{r+1},\ E_k \in H_r,\ j_0 \in H_r;\ k=1,2,3\}$$

$$\simeq H^3_{r+1} \oplus H^3_r \oplus H_r,\ r=1,2,\ldots .$$

The symbols A_k, E_k, and j_0 are used here to denote elements of space H_{r+1} and H_r; H^3_r denotes the direct sum $H_r \oplus H_r \oplus H_r$.

We shall consider an abstract problem in the space $\mathscr{H}_r$, corresponding to Eqs. (7a)–(7c):

$$\frac{du}{dt} = \mathscr{A}_r u + J_r(t,u), \tag{8a}$$

$$u(0) = u_0. \tag{8b}$$

Here u is a function of time with values in the space $\mathscr{H}_r$:

$$u(t) = \begin{pmatrix} A(t,\cdot) \\ E(t,\cdot) \\ j_0(t,\cdot) \end{pmatrix}, \tag{9a}$$

where $A(t,\cdot)\in H^3_{r+1}$, $E(t,\cdot)\in H^3_r$, $j_0(t,\cdot)\in H_r$. Operator $\mathscr{A}_r$ is a linear operator in $\mathscr{H}_r$:

$$\mathscr{A}_r u = \begin{pmatrix} E^T_k \\ \partial_l \partial_l A_k - \partial_k \partial_l A_l \end{pmatrix} \tag{9b}$$

with the domain

$$D_{\mathscr{A}_r} = \{(A,E,j_0):\ A^T_k \in H_{r+2},\ \ A^L_k \in H_{r+1},\ \ E^T_k \in H_{r+1},\ \ E^L_k \in H_r,\ \ j_0 \in H_r;\ \ k=1,2,3\}.$$

Here J_r: $R_+ \times \mathscr{H}_r \to \mathscr{H}_r$ is a function

$$J_r(t,u) = \begin{pmatrix} \partial_k\{-(1/4\pi r)*(j_0 + [E_l,A_l])\} \\ \partial_l[A_k,A_l] - [A_l,F_{lk}] + j_k \\ \partial_l j_l + [A_l,j_l] \end{pmatrix}. \tag{9c}$$

The external current is assumed to be a continuous function of time with values in the space H_{r+1}. We shall omit the subscript r when this will not lead to confusion.

It may be proved[2] that $\mathscr{A}_r$ is an infinitesimal generator of a bounded, strongly continuous one-parameter group in $\mathscr{H}_r$ and that J_r is a continuous, Lipshitzian-in-u function (in fact all terms present in J_r are C^∞ functions, except for the external current, which is continuous by assumption). Since the proof is almost the same as in Ref. 2, we shall not repeat it here. The situation would be quite different if we had used (1) instead of (7). It turns out that the operator corresponding to the linear part of Eqs. (1a) and (1b) is not a generator of a bounded, strongly continuous one-parameter group in the space $H^3_{r+1} \oplus H^3_r$. Thus we would have to use some other space, for example the space $H^3_r \oplus H^3_r$. The continuity of J_r implies then that $r \geqslant 2$. We have chosen the space $\mathscr{H}_r$ because it is easier to prove the global existence of the solution in this space.

It follows from the Segal general theory[3] that for any initial data $u_0 \in \mathscr{H}_r$ the integral equation corresponding to the Cauchy problem (8),

$$u(t) = e^{t\mathscr{A}} u_0 + \int_0^t e^{(t-s)\mathscr{A}} J(s,u(s))ds, \tag{10}$$

possesses a unique solution defined on some interval $[0,T)$, where $T>0$ usually depends on u_0. Besides, for the maximal interval of existence of the solution we have $T=\infty$ or $\lim_{t\to T}\|u(t)\| = \infty$. If, in addition, $u_0 \in D_{\mathscr{A}}$, j_k is of the C^1 class, then $u(t)$ is a C^1 function taking values in $D_{\mathscr{A}}, j$ and satisfying Eq. (8).[3] It also can be proved that the solution satisfying initially the constraint (7d) satisfies it throughout its whole interval of existence $([0,T))^2$. The connection between solutions of the abstract problem (8) and Eqs. (1) may be established by using Lemma (7) of Ref. 2 (see below).

Now we would like to investigate the smoothness of the solution. Our considerations are based on the methods used in Refs. 3 and 4. Let us assume that there exists a bounded, strongly continuous one-parameter semigroup with the infinitesimal generator $\mathscr{B}$ such that the function $u_h(t) = e^{h\mathscr{B}} u(t)$ satisfies the equation

$$u_h(t) = e^{t\mathscr{A}} u_h(0) + \int_0^t e^{(t-s)\mathscr{A}} K(h,s,u_h(s))ds, \tag{11}$$

where $u(t)$ is a solution of Eq. (10) with $u_0 \in D_{\mathscr{B}}$ and K is a C^1 function satisfying the condition

$$K(0,t,u) = J(t,u). \tag{12}$$

We denote by $w(t)$ the solution of the equation

$$w(t) = e^{t\mathscr{B}} \mathscr{B} u_0 + \int_0^t e^{(t-s)\mathscr{A}} \left\{ \frac{\partial J}{\partial u}(s,u(s)) \cdot w(s) + \frac{\partial K}{\partial h}(0,s,u(s)) \right\} ds. \tag{13}$$

Using (10), (11), and (13) we obtain

$$
\begin{aligned}
&\left\|\frac{1}{h}(u_h(t)-u(t))-w(t)\right\| \\
&\quad\leqslant\|e^{t\mathscr{A}}\|\left\|\frac{1}{h}(u_h(0)-u_0)-\mathscr{B}u_0\right\| \\
&\qquad+\int_0^t\|e^{(t-s)\mathscr{A}}\|\left\{\left\|\frac{1}{h}\left[K(h,s,u_h(s))-J(s,u_h(s))\right]-\frac{\partial K}{\partial h}(h,s,u(s))\right\|\right. \\
&\qquad+\left\|\frac{\partial K}{\partial h}(h,s,u(s))-\frac{\partial K}{\partial h}(0,s,u(s))\right\|+\left\|\frac{1}{h}\left[J(s,u_h(s))-J(s,u(s))\right]-\frac{\partial J}{\partial u}(s,u(s))\cdot(u_h(s)-u(s))\right\| \\
&\qquad\left.+\left\|\frac{\partial J}{\partial u}(s,u(s))\cdot\left(\frac{1}{h}(u_h(s)-u(s))-w(s)\right)\right\|\right\}ds \\
&\quad\leqslant\|e^{t\mathscr{A}}\|\left\|\frac{1}{h}(u_h(0)-u_0)-\mathscr{B}u_0\right\|+\int_0^t f(h,s,u(s))ds+\int_0^t\|e^{(t-s)\mathscr{A}}\|\left\|\frac{\partial J}{\partial u}(s,u(s))\right\| \\
&\qquad\times\left\|\frac{1}{h}(u_h(s)-u(s))-w(s)\right\|ds,
\end{aligned}
\tag{14}
$$

where f is some continuous function and $f(0,s,u(s))=0$. It follows from Gronwall's lemma[6] that

$$\lim_{h\to 0}\|(1/h)(u_h(t)-u(t))-w(t)\|=0. \tag{15}$$

Thus, we have proved that if $u_0\in D_{\mathscr{B}}$ then $u(t)\in D_{\mathscr{B}}$ for $t\in[0,T)$ and $\mathscr{B}u(t)=w(t)$.

Let us now assume that the components of the external current treated as functions with values in the space H_{r+j-l}, for $l=0,\dots,j$, are of class C^l, and that $u(t)$, $0\leqslant t<T$, is a solution of Eq. (10) in the space $\mathscr{H}_r$ with $u_0\in\mathscr{H}_{r+j}$ for some $j\geqslant 1$. Taking $\mathscr{B}_k=\partial_k$, $D_{\mathscr{B}_k}=\{u\in\mathscr{H}_r:\mathscr{B}_k u\in\mathscr{H}_R\}$, we obtain that

$$u(t)\in\mathscr{H}_{r+1}=\bigcap_{k=1}^{3}D_{\mathscr{B}_k},\quad 0\leqslant t<T.$$

Thus $u(t)=e_{r+1}u_1(t)$, where u_1 is a solution of Eq. (10) in the space $\mathscr{H}_{r+1}$, and e_{r+1} is the natural embedding of $\mathscr{H}_{r+1}$ into $\mathscr{H}_r$. Repeating this argument $j-1$ times, we obtain that $u(t)=e_{r+1}\cdots e_{r+j}u_j(t)$, $u_j(t)\in\mathscr{H}_{r+j}$, $0\leqslant t<T$. Since u_j, e_{r+j}, and $\mathscr{A}_{r+j-1}e_{r+j}$ are continuous, the left-hand side of the equation

$$\frac{d}{dt}e_{r+j}u_j=\mathscr{A}_{r+j-1}e_{r+j}u_j+J_{r+j-1}(t,e_{r+j}u_j) \tag{16}$$

also has to be continuous. Repeating the argument based on Eq. (16) (in which j is replaced in succession by $j-1,\dots,1$) we obtain that u_k is of class C^{j-k} (with respect to t) for $k=0,\dots,j$.

To complete our considerations concerning the existence of solutions of Eqs. (1) for small time intervals we have to establish a connection between the classical equations (1) and the abstract problem (8). We use Lemma (6) of Ref. 2 which states that if f is a C^k function mapping a real interval I into H_{2+j-k}, for $k=0,\dots,j$, then it is a C^j function on $I\times R^3$. Besides

$$\frac{df}{dt}=\frac{\partial f}{\partial t},\dots,\frac{d^jf}{dt^j}=\frac{\partial^jf}{\partial t^j}, \tag{17}$$

i.e., the derivatives of f of treated as a curve in H_2 coincide with its partial derivatives. Combining this lemma with our previous considerations we obtain that each solution of Eqs. (1) with initial data lying in H_{2+j} and the external current being a C^l function of time with values in H_{2+j-l}, for $l=0,\dots,j$, has C^{j+1} potentials A_n and C^j electric fields E_n (treated as a functions of variables x^0, x^1, x^2, x^3).

III. EXISTENCE OF A GLOBAL SOLUTION

Now it remains for us to prove that the Cauchy problem possesses a global solution. The regularity conditions, which have to be imposed on the fields, are somewhat more restrictive than in the previous section. It will be assumed that u_0 and $u(t)$ belong to the space $\mathscr{H}_4$. The external current is supposed to be a C^l function of time with values of H_{4-l}, for $l=0,1,2$. We again follow Ref. 2. First we show that the L^∞ norm of the function $F_{\mu\nu}(t,\cdot)$ and $j_0(t,\cdot)$ is bounded on each finite interval $[0,T)$. Using this it can be proved that the norm of the solution in the space $\mathscr{H}_1$ is also bounded. It follows that the solution is global in time, i.e., $T=\infty$.

We denote by $T_{\mu\nu}$ the energy-momentum tensor for the Yang–Mills field

$$T^{\mu\nu}=\operatorname{Tr}\{F^{\mu\alpha}F^{\nu}{}_{\alpha}-\tfrac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\}, \tag{18}$$

where $\eta^{\mu\nu}=\operatorname{diag}(-1,1,1,1)$ is metric tensor. Using (1b) we obtain

$$\partial_\mu T^{\mu 0}=-\operatorname{Tr}E_mj_m. \tag{19}$$

If the external current does not vanish the energy of the Yang–Mills field

$$
\begin{aligned}
\overline{E}(t)&=\int_{x^0=t}T_{00}(x)d^3x \\
&=\int_{x^0=t}\operatorname{Tr}\left\{\frac{1}{2}E_mE_m+\frac{1}{4}F_{mn}F_{mn}\right\}d^3x
\end{aligned}
\tag{20}
$$

may be time dependent. However, using (19) and the Schwarz inequality it can be easily proved that

$$\left|\frac{d}{dt}\overline{E}(t)\right|\leqslant(2\overline{E}(t))^{1/2}\|\mathbf{j}(t)\|_{L^2}. \tag{21}$$

From this inequality it follows that

$$\overline{E}\leqslant\left(\overline{E}(0)^{1/2}+\int_0^t\|\mathbf{j}(s)\|_{L^2}ds\right)^2, \tag{22}$$

where

$$\|\mathbf{j}(s)\|_{L^2}=\left(\int_{x^0=s}\operatorname{Tr}j_k(x)j_k(x)dx\right)^{1/2}.$$

We also use the notation

$$|\mathbf{j}(x)| = (\mathrm{Tr}\, j_k(x) j_k(x))^{1/2}$$

and similar notation for the fields A_μ, $F_{\mu\nu}$, and j_0.

Let us pick up an arbitrary point p in Minkowski space with $x_p^0 \geqslant 0$ (x_p^μ for $\mu = 0,1,2,3$ are coordinates of the point p). We will show that

$$|F(p)| \equiv \Big(\sum_{\mu,\nu} \mathrm{Tr}\, F_{\mu\nu}(p)^2 \Big)^{1/2}$$

and $|j_0(p)| \equiv (\mathrm{Tr}\, j_0(p)^2)^{1/2}$ may be estimated by some functions of the initial data and x_p^0. First we find an estimation on the integral

$$\begin{aligned} I(p) &= \int_{K_p} T_{0\mu}\, dS^\mu \\ &= \int_{K_p} \frac{1}{2} \mathrm{Tr} \Big\{ \tfrac{1}{4} (F_{\mu\nu} \hat{l}^\mu \hat{m}^\nu)^2 \\ &\quad + \sum_A (F_{\mu\nu} \hat{l}^\mu \hat{e}_A^\nu)^2 + (F_{\mu\nu} \hat{e}_1^\mu \hat{e}_2^\nu)^2 \Big\}, \end{aligned} \tag{23}$$

where K_p is the part of the past light cone with vertex at point p, contained between the initial data surface $t = 0$ and point p:

$$K_p = \{y:\ y^0 - x_p^0 = -r,\quad x_p^0 \geqslant y^0 \geqslant 0\},$$

$$r = \Big(\sum_{k=1}^{3} (y^k - x_p^k)^2 \Big)^{1/2},$$

$$\hat{l} = -\frac{\partial}{\partial t} + \frac{\partial}{\partial r}, \quad \hat{m} = \frac{\partial}{\partial t} + \frac{\partial}{\partial r},$$

and $\hat{e}_1$ and $\hat{e}_2$ are spacelike four-vectors satisfying the conditions

$$\hat{e}_A^\mu \hat{l}_\mu = \hat{e}_A^\mu \hat{m}_\mu = 0, \quad \hat{e}_A^\mu \hat{e}_\mu^B = \delta_A^B.$$

Applying Gauss' theorem to Eq. (19) we obtain

$$I(p) - \int_{B_p} T_{00}(y) d^3y = -\int_{\widetilde{K}_p} \mathrm{Tr}\, j_k(y) E_k(y) d^4y. \tag{24}$$

Here B_p is a three-dimensional ball of radius $r_0 = x_0^p$ contained in the initial surface $t = 0$,

$$B_p = \{y:\ r < r_0,\quad y^0 = 0\},$$

$$\widetilde{K}_p = \{y:\ r \leqslant x_p^0 - y^0, 0 \leqslant y^0 \leqslant x_p^0\},$$

$$\partial \widetilde{K}_p = K_p \cup B_p.$$

Using the inequality $\int T_{00} d^3y \geqslant 0$ and the Schwarz inequality we find that

$$\begin{aligned} I(p) &\leqslant \overline{E}(0) - \int_0^{x_p^0} ds \int_{r<s=x^0} \mathrm{Tr}\, j_k(y) E_k(y) d^3y \\ &\leqslant \overline{E}(0) + 2 \int_0^{x_p^0} \|\mathbf{j}(s)\|_{L^2} \overline{E}(s)^{1/2}\, ds = \widetilde{E}(s), \end{aligned} \tag{25}$$

where we have denoted by $\widetilde{E}(t)$ the right-hand side of the last inequality in (25).

Now we shall use the transformation (3). The function $\mathscr{U}$ may be chosen in such a way that the following conditions hold[2,7]:

$$\hat{A}_\mu(y)(y^\mu - x_p^\mu) = 0, \tag{26a}$$

$$\mathscr{U}(x_p) = 1. \tag{26b}$$

It is convenient to shift the origin of the coordinate system to the point p. Conditions (26) then become $\hat{A}_\mu(x) x^\mu = 0$, $\mathscr{U}(0) = 1$. In the new coordinates the initial data are given at time $t_0 = -r_0 = -x_p^0$. The new potentials $\hat{A}_\mu$ can be expressed in terms of the field strength tensor $\hat{F}_{\mu\nu}$ (see Refs. 2 and 7):

$$\hat{A}_\mu(x) = \int_0^1 d\lambda\, \lambda x^\mu \hat{F}_{\mu\nu}(\lambda x). \tag{27}$$

Hence

$$\partial_\mu \hat{A}^\mu(x) = \int_0^1 d\lambda\, \lambda^2 x^\alpha \{ [\hat{F}_{\alpha\mu}(\lambda x), \hat{A}^\mu(\lambda x)] + \hat{j}_\alpha(\lambda x) \}. \tag{28}$$

From (4) and the Bianchi identity it follows that

$$\hat{D}_\gamma \hat{D}^\gamma \hat{F}_{\alpha\beta} = 2[\hat{F}^\gamma{}_\alpha, \hat{F}_{\gamma\beta}] - \hat{D}_\alpha \hat{j}_\beta + \hat{D}_\beta \hat{j}_\alpha. \tag{29}$$

Separating the linear part and using the retarded Green's function for the wave equation we obtain from (29) the following integral equation for $F_{\alpha\beta}$:

$$\begin{aligned} \hat{F}(0)_{\alpha\beta} = \hat{F}_{\alpha\beta}^{\mathrm{in}}(0) - \frac{1}{4\pi} \int_{K_p} r\, dr\, d\Omega \{ &-2\partial_\gamma [\hat{A}^\gamma, \hat{F}_{\alpha\beta}] + [\partial_\gamma \hat{A}^\gamma, \hat{F}_{\alpha\beta}] + [\hat{A}^\gamma, [\hat{A}_\gamma, \hat{F}^{\alpha\beta}]] \\ &+ 2[\hat{F}^\gamma{}_\alpha, \hat{F}_{\gamma\beta}] - \hat{D}_\alpha \hat{j}_\beta + \hat{D}_\beta \hat{j}_\alpha \}|_{t=-r}, \end{aligned} \tag{30a}$$

where

$$\hat{F}_{\alpha\beta}^{\mathrm{in}}(0) = \frac{1}{4\pi} \int_{S^2} d\Omega \Big\{ r_0 \frac{\partial \hat{F}_{\alpha\beta}}{\partial t} + r_0 \frac{\partial \hat{F}_{\alpha\beta}}{\partial r} + \hat{F}_{\alpha\beta} \Big\} \Big|_{t=t_0,\ r=r_0} = \frac{1}{4\pi} \int_{S^2} d\Omega \{ r_0 \hat{m}^\mu \partial_\mu \hat{F}_{\alpha\beta} + \hat{F}_{\alpha\beta} \}|_{t=t_0,\ r=r_0} \tag{30b}$$

is the solution of the wave equation $\partial^\gamma \partial_\gamma \hat{F}_{\alpha\beta}^{\mathrm{in}} = 0$ for the same initial data as for $\hat{F}_{\mu\nu}$, i.e., $\hat{F}_{\mu\nu}^{\mathrm{in}}|_{t=t_0} = \hat{F}_{\mu\nu}|_{t=t_0}$, $\partial_0 \hat{F}_{\mu\nu}^{\mathrm{in}}|_{t=t_0} = \partial_0 \hat{F}_{\mu\nu}|_{t=t_0}$, $S^2 = \partial B_p$, and $d\Omega$ is an element of area on the unit sphere.

After substituting $\partial_\mu A^\mu$ from formula (28), Eq. (30a) becomes

$$\begin{aligned} \hat{F}_{\mu\nu}(0) = \hat{F}_{\alpha\beta}^{\mathrm{in}}(0) - \frac{1}{4\pi} \int_{K_p} r\, dr\, d\Omega \Big\{ &-2\, \partial_\gamma [\hat{A}^\gamma, \hat{F}_{\alpha\beta}] + \Big[\int_0^1 d\lambda\, \lambda^2 x^\gamma [\hat{F}_{\gamma\mu}(\lambda x), \hat{A}^\mu(\lambda x)], \hat{F}_{\alpha\beta}(x) \Big] \\ &- [\hat{A}^\gamma, [\hat{A}_\gamma, \hat{F}_{\alpha\beta}]] - 2[\hat{F}^\gamma{}_\alpha, \hat{F}_{\gamma\beta}] + \int_0^1 d\lambda\, \lambda^2 x^\gamma [\hat{j}_\gamma(\lambda x), \hat{F}_{\alpha\beta}] - \hat{D}_\alpha \hat{j}_\beta + \hat{D}_\beta \hat{j}_\alpha \Big\} \Big|_{t=-r}. \end{aligned} \tag{31}$$

Repeating the arguments given in Ref. 2 one can obtain an estimate of the terms not involving j_μ; namely, the sum of $F_{\mu\nu}(0)$ and the integral of three last terms is bounded by the following expression:

$$C_1\widetilde{E}(r_0)\int_0^{r_0} dr\|F(-r)\|_{L^\infty} + C_2\widetilde{E}(r_0)^{1/2}\Big(\int_0^{r_0} dr\|F(-r)\|^2_{L^\infty}\Big)^{1/2} + K_1(t_0) + K_2(t_0), \tag{32}$$

where the C_i are constants, the K_i depend on the initial data, and

$$\|F(t)\|_{L^\infty} = \sup_{x^0=t}|F(x)|.$$

The first integral involving $\hat{j}_\mu$ may be estimated in the following way:

$$\Big|\int_{K_p} r\,dr\,d\Omega\int_0^1 d\lambda\,\lambda^2[x^\gamma\hat{j}_\gamma(\lambda x),\hat{F}_{\alpha\beta}(x)]\Big|$$

$$\leqslant 2\int_0^1 d\lambda\,\lambda^{1/2}\Big(\int_{K_p} r^2\,dr\,d\Omega\,\lambda^3\Big|\frac{x_\gamma}{r}j_\gamma(\lambda x)\Big|^2\Big)^{1/2}\Big[4\pi\int_0^1\|F(-r)\|^2_{L^\infty}r^2\,dr\Big]^{1/2}$$

$$\leqslant\Big\{C_3\|j_0(t)\|^2_{L^2} + C_4\Big(\|A(t_0)\|^2_{L^2} + r_0^2\int_0^{r_0}\overline{E}(t)dt\Big)\int_0^{r_0}\|\mathbf{j}(-r)\|^2_{L^\infty}dr\Big\}\Big(r_0^2\int_0^{r_0}dr\|F(-r)\|^2_{L^\infty}\Big)^{1/2}, \tag{33}$$

where we have used the relations $|x^\gamma\hat{j}_\gamma| = |x^\gamma j_\gamma|$ and $|\hat{F}| = |F|$ and, moreover, Eqs. (1a) and (1e) to express j_0 in terms of the electric field, external current, and initial data.

We evaluate the next integral using (1e):

$$\int_{K_p} r\,dr\,d\Omega\,\hat{D}_l\hat{j}_0\Big|_{t=-r}$$

$$=\int_{K_p} r^2\,dr\,d\Omega\Big\{\partial_l\Big(\frac{1}{r}\Big)\hat{j}_0|_{t=-r} + \Big(-\frac{x^l}{r^2}[\hat{A}_0,\hat{j}_0] + \frac{x^l}{r^2}\hat{D}_k\hat{j}_k + \frac{x^l}{r^3}\hat{j}_0 + \frac{1}{r}[\hat{A}_l,\hat{j}_0]\Big)\Big|_{t=-r}\Big\}. \tag{34}$$

Performing the integration over x_1 in the first term in (34) we get

$$\int_{K_p} r^2\,dr\,d\Omega\,\partial_1\Big(\frac{1}{r}\hat{j}_0\Big|_{t=-r}\Big) = \frac{1}{r_0}\int_{x_2^2+x_3^2<r_0^2} d^2x\{\hat{j}_0|_{t=-r_0,\,x_1=(r_0^2-x_2^2+x_3^2)^{1/2}} - \hat{j}_0|_{t=-r_0,\,x_1=-(r_0^2-x_2^2+x_3^2)^{1/2}}\}. \tag{35}$$

Similarly performing the integration over x_2 and x_3 we obtain the integrals involving only the initial values of $\hat{j}_0$ which are bounded because $|\hat{j}_0| = |j_0|$. To estimate the second integral on the rhs of (34) we use the formula (27)

$$\Big|\int_{K_p} r^2\,dr\,d\Omega\,\frac{x^l}{r^2}[\hat{A}_0,\hat{j}_0]\Big|_{t=-r}\Big|$$

$$\leqslant 2\int_0^1 d\lambda\,\lambda^{-1/2}\Big(\int_{K_p} r^2\,dr\,d\Omega\,\lambda^3|l^\mu F_{\mu 0}(\lambda x)|^2\Big)^{1/2}\Big(4\pi\int_0^{r_0} dr\,r^2\|j_0(-r)\|^2_{L^\infty}\Big)^{1/2}$$

$$\leqslant C_5E(r_0)^{1/2}\Big(r_0^2\int_0^{r_0} dr\,\|j_0(-r)\|^2_{L^\infty}\Big)^{1/2}. \tag{36}$$

Similarly,

$$\Big|\int_{K_p} r^2\,dr\,d\Omega\,\frac{x^l}{r^2}[\hat{A}_l,\hat{j}_0]\Big|_{t=-r}\Big| \leqslant C_6\widetilde{E}(r_0)^{1/2}\Big(r_0^2\int_0^{r_0} dr\|j_0(-r)\|^2_{L^\infty}\Big)^{1/2}. \tag{37}$$

For the term involving $\hat{D}_k\,\hat{j}_k$ we obtain

$$\Big|\int_{K_p} r^2\,dr\,d\Omega\,\frac{x^l}{r^2}\hat{D}_k\hat{j}_k\Big|_{t=-r}\Big|$$

$$\leqslant\int_{K_p} r\,dr\,d\Omega|D_kj_k|_{t=-r}|$$

$$\leqslant C_7\int_{K_p} r\,dr\,d\Omega|\partial_kj_k|_{t=-r}| + \Big(\int_{K_p} r\,dr\,d\Omega|\mathbf{j}|\,|_{t=-r}\Big)\Big(C_8\|A(t_0)\|_{L^\infty} + C_9\int_0^{r_0} dr\|F(-r)\|_{L^\infty}\Big). \tag{38}$$

The remaining integral in (34) and the last integral in (31) are easy to estimate:

$$\left|\int_{K_p} r^2\, dr\, d\Omega\, \frac{x^l}{r^3}\hat{j}_0\right|_{t=-r}\right| \leqslant 4\pi \int_0^{r_0} dr \| j_0(-r)\|_{L^\infty}, \tag{39}$$

$$\left|\int_{K_p} r\, dr\, d\Omega\, \hat{D}_0 \hat{j}_k\right|_{t=-r}\right| \leqslant \int_{K_p} r\, dr\, d\Omega \left|\partial_k j_k\right|_{t=-r}. \tag{40}$$

The estimates (32)–(40) yield

$$|F(0)| \leqslant f_1(r_0) + f_2(r_0)\int_0^{r_0} dr\|F(-r)\|_{L^\infty} + f_3(r_0)\left(\int_0^{r_0} dr\|F(-r)\|^2_{L^\infty}\right)^{1/2}$$
$$+ f_4(r_0)\int_0^{r_0} dr\| j_0(-r)\|_{L^\infty} + f_5(r_0)\left(\int_0^{r_0} dr\| j_0(-r)\|^2_{L^\infty}\right)^{1/2}, \tag{41}$$

where the functions $f_k(r_0)$ are bounded on each finite interval $[0,T)$ and depend on the initial data and the external current. From (41) it follows that

$$\|F(t)\|_{L^\infty} \leqslant f_1'(t) + f_2'(r_0)\int_0^t dr\| F(-r)\|_{L^\infty} + f_3'(r_0)\int_0^t dr\| j_0(-r)\|_{L^\infty}. \tag{42}$$

The inequality (42) is written in the old coordinate system (in which the initial data are given at $t=0$).

We also need an estimate on $\| j_0(t)\|_{L^\infty}$. Using (1a) and (1b) it is not difficult to show that

$$\| j_0(t)\|_{L^\infty} \leqslant \| j_0(0)\|_{L^\infty} + \int_0^t ds\|\partial_k j_k(s)\|_{L^\infty} + C\left(\int_0^t ds\|\mathbf{j}(s)\|_{L^\infty}\right)\left(\|A(0)\|_{L^\infty} + t\int_0^t ds\|F(s)\|_{L^\infty}\right). \tag{43}$$

Next we define

$$N(t) = \|F(t)\|_{L^\infty} + \| j_0(t)\|_{L^\infty}. \tag{44}$$

Since $u(t)\in\mathscr{H}_4$ it follows from Sobolev inequalities that $N(t)$ is continuous.[2] Combining (41) and (42) we obtain

$$N(t) \leqslant f(t) + g(t)\int_0^t ds N(s) \tag{45}$$

with appropriate $f(t)$ and $g(t)$. Applying Gronwall's lemma we see that $N(t)$ is bounded on each finite interval $[0,T)$. Equation (1a) implies that $\|A\|_{L^\infty}$ is also bounded.

In order to complete the proof we have to estimate the $\mathscr{H}_1$ norm of the solution. Since

$$\|u\|_{\mathscr{H}_1} \leqslant C(\mathscr{R}_0 + \mathscr{E}_0 + \mathscr{R}_0 + \mathscr{E}_0), \tag{46}$$

where

$$\mathscr{R}_0 = \frac{1}{2}\int_{R^3} \operatorname{Tr} j_0 j_0, \tag{47a}$$

$$\mathscr{E}_0 = \frac{1}{2}\int_{R^3} \operatorname{Tr}\{E_l E_l + (\partial_k A_l)(\partial_k A_l) + m^2 A_l A_l\}, \quad m>0, \tag{47b}$$

$$\mathscr{R}_1 = \frac{1}{2}\int_{R^3} \operatorname{Tr}(\partial_l j_0)(\partial_l j_0), \tag{47c}$$

$$\mathscr{E}_1 = \frac{1}{2}\int_{R^3} \operatorname{Tr}\{(\partial_k E_l)(\partial_k E_l) + (\partial_k\partial_l A_n)(\partial_k\partial_l A_n)\}, \tag{47d}$$

it is enough to show that $\mathscr{R}_k$ and $\mathscr{E}_k$ are bounded. Using Eqs. (1) and proceeding in a manner similar to Ref. 2 we have

$$\left|\frac{d}{dt}\mathscr{R}_0\right| \leqslant (\|\partial_k j_k\|_{L^\infty} + \|\mathbf{j}\|_{L^2}\|A\|_{L^\infty})\mathscr{R}_0^{1/2}, \tag{48a}$$

$$\left|\frac{d}{dt}\mathscr{E}_0\right| \leqslant (C_0 + C_1\|F\|_{L^\infty} + \mathscr{R}_0)\mathscr{E}_0, \tag{48b}$$

$$\left|\frac{d}{dt}\mathscr{R}_1\right| \leqslant (\|\partial_m j_m\|_{H_1} + 4\|\mathbf{j}\|_{L^\infty}\mathscr{E}_0 + \|\mathbf{j}\|_{H_1}\|A\|_{L^\infty})\mathscr{R}_1^{1/2}, \tag{48c}$$

$$\left|\frac{d}{dt}\mathscr{E}_1\right| \leqslant \mathscr{R}_1 + (C_2 + C_3\|A\|_{L^\infty})\|\mathbf{j}\|^2_{H_1} + (C_4 + C_5\|A\|^2_{L^\infty} + C_6\|F\|^2_{L^\infty})\mathscr{E}_0 + (C_7 + C_8\|A\|^2_{L^\infty})\mathscr{E}_1. \tag{48d}$$

Hence we obtain the desired estimation. Thus the proof is complete.

IV. ENDING REMARKS

In this paper we have considered the Cauchy problem for Eqs. (1), i.e., for the Yang–Mills field with the external current. We have proved that if the initial data satisfy the condition

$$A_k(0,\cdot)\in H_5,\quad E_k(0,\cdot)\in H_4,\quad D_kE_k\in H_4,$$

and the external current is a C^l function with values in H^{4-l}, for $l=0,1,2$, then there exists a global solution (since the two first conditions imply $[A_k,E_k]\in H_4$, the third one is equivalent to $\partial_kE_k\in H_4$). The solution is uniquely determined by the initial data. Besides, $A_k(t,\cdot)\in H_5$, $E_k(t,\cdot)\in H_4$, $\partial_kE_k(t,\cdot)\in H_4$, for $t\geqslant 0$. The potential A and the longitudinal part of the electric field E^L are C^3 functions of the variables $x^0,...,x^3$; the transverse part is a C^2 function.

Our first remark concerns solutions of Eqs. (1) in local spaces. The condition $f\in H_r$ implies restrictions of the behavior of f at infinity. Since the speed of propagation of field disturbance described by Eqs. (1) is bounded by the velocity of light these restrictions are not necessary. We define H_r^{loc} as a space of functions from R^3 to the Lie algebra of G satisfying the condition $gf\in H_r$ for an arbitrary C^∞ function g with compact support (here $f\in H_r^{\text{loc}}$). Using the method of Ref. 5 it can be proved that everything we have said above about global solutions remains valid if we replace each H_r by H_r^{loc}.

Let us recall that in our formulation of the Cauchy problem for the Yang–Mills equations with the external current, j_0 is not arbitrary and $A_0=0$. However, it follows from our considerations that if Eqs. (2) are satisfied then the spatial part of the external current j_μ, for $t\geqslant 0$, the initial values of the potentials $A_\mu|_{t=0}$, and their time derivatives $\partial_0A_\mu|_{t=0}$ uniquely determine all quantities invariant under the transformation (3) e.g., $\operatorname{Tr} j_0 j_\mu$. Thus in any case we cannot treat all four components of the external current and the initial data as arbitrary functions. On the other hand, the Yang–Mills equations do not determine ∂_0A_0 so if we do not impose an extra constraint (e.g., $A_0=0$) the solution is not unique.

We would like to end our paper with some remarks about Lagrangian formulations of the Yang–Mills theory with external current. It is easy to see that Eq. (1b) follows from the Lagrangian density

$$\mathscr{L}=\operatorname{Tr}\{\tfrac12(\partial_0A_k)(\partial_0A_k)-\tfrac14F_{lk}F_{lk}+A_kj_k\}. \tag{49}$$

We may also use

$$\mathscr{L}'=\operatorname{Tr}\{-\tfrac14F_{\mu\nu}F^{\mu\nu}+A_\mu j^\mu\}. \tag{50}$$

Taking the variation of the action corresponding to $\mathscr{L}'$ with respect to A_μ and j_0 we obtain the following Euler–Lagrange equations:

$$D^\mu F_{\mu\nu}=-j_\nu, \tag{51a}$$

$$A_0=0. \tag{51b}$$

One of the possible generalizations of the Lagrangian $\mathscr{L}'$ has the form

$$\mathscr{L}''=\operatorname{Tr}\{-\tfrac14F_{\mu\nu}F^{\mu\nu}+A_\mu j^\mu[w]+h[w]\}, \tag{52}$$

where j^μ and h are functionals of a new field w. To obtain $\mathscr{L}'$ from $\mathscr{L}''$ we have to take $j^\mu=(w,\mathbf{j})$, $h=0$. If we choose $j^\mu=\tilde{j}^\mu+\partial^\mu w$, $h=0$, we get the constraint $\partial_\mu A^\mu=0$. The Lagrangian $\mathscr{L}''$ establishes some connection between the form of the dependence of the current on w and the constraint imposed on the potentials. However, this connection seems to have no significance in the considerations concerning the Cauchy problem. For example, we can take $j^\mu=\tilde{j}^\mu+\partial^\mu w$ and $A_0=0$ instead of $\partial_\mu A^\mu=0$.

Sometimes we obtain a well-posed Cauchy problem treating the covariant "current continuity equation" (2c) as a constraint imposed on the potentials. An example is SU(2) gauge theory with the external charge (the current has the form $j^\mu=\delta_0^\mu\rho$, where ρ is a Lie-algebra-valued function satisfying the condition $\partial_0\operatorname{Tr}\rho^2=0$).

ACKNOWLEDGMENTS

The authors thank H. Arodź, K. Heller, and P. Serda for a critical reading of the manuscript and H. Arodź for suggesting this problem.

This work was supported in part by the Polish Ministry of Science and Higher Education, Project CPBP01.03-1.7.

[1]H. Arodź, Phys. Rev. D **35**, 4024 (1987).
[2]D. M. Eardley and V. Moncrief, Commun. Math. Phys. **83**, 171, 193 (1982).
[3]I. Segal, Ann. Math. **78**, 393 (1963).
[4]I. Segal, J. Funct. Anal. **33**, 175 (1979).
[5]J. Ginibre and G. Velo, Commun. Math. Phys. **82**, 1 (1981).
[6]P. Hartman, *Ordinary Differential Equations* (Wiley, New York, 1964).
[7]C. Cronström, Phys. Lett. B **90**, 267 (1980).

On the iterative solutions of some nonlinear eigenvalue problems

Kazimierz Wanelik[a)]
Institute of Physics, Jagellonian University, Reymonta 4, 30-059 Cracow, Poland

(Received 1 March 1988; accepted for publication 15 February 1989)

A simple refinement procedure of iteration methods for nonlinear systems is presented. It is based upon the introduction of a numerical measure of the "separation" between iterates, which themselves depend on parameters to be determined by requirements on the magnitude of this "separation." As a simple but informative application of this idea the refined Picard iteration of the null Dirichlet problem for $\Delta u + \lambda f(x,u) = 0$ is considered. To illustrate the use of this technique the equation of Bratu and the so-called Gel'fand problem is observed. A discussion follows.

I. INTRODUCTION

There exists an interesting refinement procedure for the iterative solution of functional equations suggested by Rayski,[1] within the context of Lippmann–Schwinger-type integral equations. This method, called the refined Born approximation, has been discussed in several papers.[2] The practical efficiency of the method has been confirmed in recent calculations of some atomic and molecular properties.[3] By virtue of its generality, Rayski's method may be applied to a large variety of physical problems. Moreover, this kind of refinement technique diversifies the iteration method itself.

The present paper is intended as a first step in treating nonlinear systems by Rayski-type refinement of Picard's iteration. Our main interest is in questions such as nonuniqueness and critical dependence on parameters connected with the intrinsic behavior of solutions of nonlinear systems.

Specifically, we investigate the positive solutions of the following nonlinear eigenvalue problem:

$$Lu = \lambda f(x,u), \quad x\in\Omega, \tag{1.1a}$$

$$u = 0, \quad x\in\partial\Omega, \tag{1.1b}$$

where L is the negative Laplacian,

$$L \equiv -\Delta; \tag{1.1c}$$

Ω, a bounded (open) domain in R^m; λ, a real parameter; and f, a given nonlinear function of its second argument. The importance of this problem in mathematical analysis, in geometry, and in various applications, has been recognized for many years.[4–9] Let us recall that this type of problem arises in many questions of mathematical physics: existence of solitary waves, nonlinear field equations, problems of false vacuum, nonlinear heat equations, nonlinear diffusion equations, etc.

A most striking feature of problem (1.1) is that the positive solutions of (1.1) need not be unique. The purpose of this paper is to relate results for this problem obtained by the standard Picard method to those established by the refined one. Problem (1.1), as well as some generalizations of it, were treated by Keller and Cohen[9] and by Laetsch,[10] who originally developed the monotone iterative methods placed in the context of Banach spaces by Amann.[11] Our approach presented in the paper is somewhat inspired by their treatment of such problems.

In Sec. II of this paper we state some previously established results on the positive solutions of (1.1) under rather mild monotonicity conditions on f; the Schauder fixed point theorem, Picard's iteration, and the perturbation method are applied to the problem. Rayski's refinement procedure of an iterative method is described in Sec. III. In Sec. IV the procedure is applied to Picard's method for the solution of problem (1.1). We look at the equation of Bratu ($m = 1$) and the so-called Gel'fand problem ($m = 2$). As will be seen, the refined Picard iterations yield much more information on the solutions than the standard Picard scheme. In this case the main advantage of the refined iteration is that it yields quantitative improvements and, furthermore, the qualitative results are much deeper than those obtained in the ordinary approach (at the same stage in the iterations). In order to illustrate these qualitative improvements, we consider the limit of vanishing λ; in this limit the standard scheme corresponds to the perturbation method (λ being the perturbation parameter). It is very encouraging to note that the refined technique, although very simple and not more cumbersome than the plain version of the iteration, predicts results with surprisingly small errors and yields remarkable results on the multiplicity of solutions and the bifurcation picture, even in the simplest possible version (and in the lowest-order approximation). The significance of our results is finally discussed, and a few remarks made in support of the method, in Sec. V.

II. STATEMENT OF THE PROBLEM AND SOME PREVIOUSLY ESTABLISHED RESULTS

The example of a nonlinear system, which we study in this paper, is concerned with the nonlinear eigenvalue problem of the form (1.1) defined on $\bar{\Omega}$ under certain conditions on f. We seek positive solutions $u(\lambda;x) > 0$ on Ω, and investigate the set of values of λ for which such solutions exist. The boundary $\partial\Omega$ is assumed "sufficiently smooth." (This last is merely an announcement that our concerns are only with those technical difficulties arising from considerations other than regularity of $\partial\Omega$.[12]) The nonlinearity $f(x,v)$ is a smooth function of its arguments on the $(m+1)$-dimensional half-cylinder $x\in\Omega$, $v\geqslant 0$. The problem for (1.1) consists in determining a function $u(\lambda;x)\in[C^2(\Omega)\cap C(\bar{\Omega})]$

[a)] Present address: Department of Theoretical Physics, 1 Keble Road, Oxford, OX1 3NP, England.

 0022-2488/89/081707-06$02.50

that satisfies (1.1) in the pointwise sense. In addition to the smoothness requirements mentioned above, the conditions to be imposed on f will be one or more of the following:

$$f(x,v) \geqslant \beta > 0, \quad \text{on } \Omega \text{ for all } v \geqslant 0, \tag{2.1a}$$

$$f(x,v) \text{ is nondecreasing in } v \text{ for fixed } x, \tag{2.1b}$$

$$f(x,v) \geqslant g(x)v > 0, \quad \text{for all } v > 0. \tag{2.1c}$$

In order to present the results of the present paper, we shall need some previously established facts. Rather than rederive these results we shall simply state them and refer the reader to Refs. 9, 13, and 14 where they are proved.

Property 1: Only positive λ can be in the spectrum Λ, the set of all values of λ for which (1.1) has a positive solution.[9,13]

Property 2: There is a finite (critical) number $\lambda_c > 0$, such that for $\lambda < \lambda_c$ the system (1.1) has at least one positive solution, while for $\lambda > \lambda_c$ the system has no positive solution.[9,13,14]

Property 3: If $S = S(\lambda)$, the set of all positive solutions of (1.1) (with a certain λ) is not empty, then there exists the minimal solution $\underline{u}(\lambda;x)$, in the sense that $\underline{u}(\lambda;x) \leqslant u(\lambda;x)$ for any other positive solution.[9,13] Moreover, $\underline{u}(\lambda;x)$ is an increasing function of λ on Λ for each x.[9,13]

Property 4: If λ is a point in the spectrum Λ, then the sequence $\{u_n(\lambda;x)\}$, defined by the standard Picard iteration,

$$u_0 \equiv 0, \tag{2.2a}$$

$$Lu_n = \lambda f(x,u_{n-1}(x)), \quad x \in \Omega, \tag{2.2b}$$

$$u_n = 0, \quad x \in \partial\Omega, \quad n \geqslant 1, \tag{2.2c}$$

is monotone increasing and converges uniformly to the minimal solution, that is,[9,13]

$$\lim_{n\to\infty} u_n(\lambda;x) = \underline{u}(\lambda;x). \tag{2.3}$$

It should be pointed out that the minimal solution is unique and therefore that it has the same symmetries as the problem (1.1); the spectrum may be closed or open depending on the particular shape of $f(x,v)$.[5–10,13,15] The length of the interval $(0, \lambda_c)$ is variable and depends on f and the geometry of Ω.[5–10,13–15]

Now let us look for the positive solutions of (1.1) and (2.1) that satisfy, on Ω, the following condition:

$$\lim_{\lambda\to 0} u(\lambda;x) = 0. \tag{2.4a}$$

In this case we may apply the perturbation method in the neighborhood of $\lambda = 0$. We proceed by assuming that the unknown function $u(\lambda;x)$ can be expanded on Ω as $\lambda \to 0$, according to

$$u(\lambda;x) = \sum_{m=1}^{\infty} u_m(x)\lambda^m, \tag{2.4b}$$

where the quantities u_m are functions to be determined from (1.1), and the (known) nonlinearity $f(x,v)$ can be expanded on $\Omega \times [0,\rho)$, according to

$$f(x,v) = \sum_{m=0}^{\infty} \varphi_m(x)v^m. \tag{2.5}$$

Since it is more convenient to deal with integrals than derivatives, we shall rewrite the problem (1.1) as the following integral equation:

$$u(x) = \lambda \int_\Omega dy\, g(x,y) f(y,u(y)), \quad x \in \Omega, \tag{1.1'}$$

where $g(x,y)$ is the (positive) Green function for (L,Ω), vanishing on $\partial\Omega$. Now if we substitute (2.4b) and (2.5) in Eq. (1.1'), and equate the terms of the same order in λ on both sides, we get the following set of expressions for u_m, which ensures that $u(\lambda;x)$ given by (2.4b) satisfies (formally) Eq. (1.1'):

$$u_1(x) = \int_\Omega dy\, g(x,y)\varphi_0(y),$$

$$u_2(x) = \int_\Omega dy\, g(x,y)\varphi_1(y)u_1(y),$$

$$\vdots$$

$$u_{m+1}(x) = \int_\Omega dy\, g(x,y)(\varphi_1 u_m + \varphi_2 u_{m,2} + \cdots + \varphi_p u_{m,p} + \cdots + \varphi_m u_1^m)(y), \tag{2.6a}$$

where

$$u_{m,p} = \sum_P u_{i_1} u_{i_2} \cdots u_{i_p}, \tag{2.6b}$$

where the sum is taken over $P \equiv (i_1, i_2, \ldots, i_p)$, all p permutations of numbers $\{1,2,3,\ldots,m\}$ with repetitions to p, and that $i_1 + i_2 + i_3 + \cdots + i_p = m$ [in particular, we have $u_{m,1} = u_m(y)$, $u_{m,m} = u_1^m(y)$].

Apparently, this perturbative result corresponds to Picard's iterates in the limit of vanishing λ [compare (2.2)]. Moreover, in view of property 4 it is clear that the solution (2.4)–(2.6) corresponds to the minimal positive solution of (1.1). This means that for λ sufficiently small the series (2.4)–(2.6) converges uniformly on $\Omega \times [0,\rho)$, that is,

$$\sum_{m=1}^{\infty} u_m(x)\lambda^m = \underline{u}(x,\lambda). \tag{2.7}$$

III. OUTLINE OF THE METHOD

Now let us describe Rayski's refinement of an iterative method. In the interest of simplicity and clarity it is best to discuss the one-step stationary case of iterative processes, examples of which are Picard's method and Newton's method. The sequence of iterates is then defined recursively by the following relation:

$$u_{n+1} = Tu_n, \quad n = 0,1,2,\ldots, \tag{3.1}$$

where the iteration is started by choosing a (suitable) first approximation to u. Rayski's method is a very general one that can be used whenever one imposes a (proper) numerical measure $Q[v,w]$ of the "separation" between iterates v and w. For simplicity we restrict our considerations to a pair of consecutive iterates u_n, u_{n+1}.[2] The principle involved is the following. If at any stage in the above successive definitions we have $u_{n+1} = u_n$, then $u_n = u_{n+1} \equiv u$ is a fixed point of the operator T, that is,

$$u = Tu, \tag{3.2}$$

and the procedure is terminated. In general this will not happen, but the functional $Q[u_n,u_{n+1}]$ will become more nearly constant ($=Q[u,u]$) as n is taken larger. We are now starting with an initial estimate $u_0(\alpha)$ labeled with a certain number of continuous parameters α to be determined. Considered as a functional of the pair $u_n(\alpha),u_{n+1}(\alpha)$, the quantity Q reduces to a simple function of the parameters α,

$$q(\alpha)\equiv Q\,[u_n(\alpha),u_{n+1}(\alpha)]\,. \tag{3.3}$$

Each set of values $\hat{\alpha}$ for which

$$q(\hat{\alpha})=Q\,[u,u]\,, \tag{3.4}$$

defines an approximate fixed point of T,

$$\hat{u}\equiv u_{n+1}(\hat{\alpha})\,. \tag{3.5}$$

More precisely, if the zeroth-order approximation to u is a function of N unknown real parameters, then these N parameters can be found from N requirements of the type (3.4); this means, for example, from the following system of equations[2,16]:

$$Q\,[u_{n+i},u_{n+i+1}]=Q\,[u,u]\,, \tag{3.6}$$

for $i=0,1,2,\ldots,N-1$.

The success of the method depends essentially on the choice of the initial iterate and the measure of the separation. These must be simple enough to lend themselves easily to the calculation, but must be judicious for the solutions obtained to be close to the exact ones. Apparently, the method described above can be easily developed toward a "variation–iteration" procedure.[2] The initial approximation to u and the functional Q should be chosen in such a way that one is able to find "good" $\hat{\alpha}$'s [in the sense of (3.5)] after performing few iterations.

There is one remark that has to be added: In case the operator T in (3.1) is nonlinear we get

$$Q\,[u,u]=Q\,[u_n,Tu_n](\alpha)\,, \tag{3.7}$$

which is a nonlinear algebraic equation for the determination of the α's. This leads in a natural way to the problem of multiplicity of the solutions of Eq. (3.2) and related topics. Actually, Rayski's procedure is related to the solution of nonlinear functional equations by appealing (to some extent) to certain nonlinear algebraic equations derived from these functional ones. Moreover, one may employ Rayski's procedure as a search algorithm to obtain initial approximations, and then subsequently approach the solutions by applying a fast iterative method.

IV. APPLICATIONS

In order to illustrate the idea outlined in Sec. III, we consider the refined Picard's method applied to (1.1). In the case of the problem for (1.1), the method of collocation, Galerkin's method, and the least-squares method provide examples of the measure Q.[2,16] To further simplify considerations we assume that the initial guess is (positive on Ω and) linearly dependent on its parameters, that is,

$$u_0(\lambda;x)=\sum_{m=1}^{N}\alpha_m\chi_m(x)\,, \tag{4.1}$$

where $\{\chi_m\}$ is a set of linearly independent functions we have at our disposal. When this quantity is substituted in (2.2b) and (2.2c), we obtain for all $x\in\Omega$,

$$u_n=u_n(x;\alpha),\quad n\geqslant 1\,. \tag{4.2}$$

Whatever the choice of $\alpha_1,\alpha_2,\ldots,\alpha_N$, we cannot expect $u_n(x;\alpha)$ and $u_{n+1}(x;\alpha)$ to be equal on Ω,

$$u_n(x;\alpha)=u_{n+1}(x;\alpha) \tag{4.3}$$

[unless, accidentally, the exact solution of (1.1) is of this form]. One is therefore led to a number of different ways of satisfying (4.3) approximately. The idea is to choose $\alpha_1,\ldots,\alpha_N$ so that the two sides of (4.3) "are equal as nearly as possible." There are various criteria (that is various Q's) for "nearly equal." We might require, for instance, that the two sides be equal at N suitable chosen points $x_1,x_2,\ldots,x_N\in\Omega$, or, following (3.6), demand that

$$u_{n+i}(x_0;\alpha)=u_{n+i+1}(x_0;\alpha),\quad i=0,1,\ldots,N-1\,, \tag{4.4}$$

where x_0 is a suitable chosen point in Ω.

The following reasoning may be useful in choosing the point x_0. Let us define the functional,

$$\lambda_c\,[v(x;\alpha)]\equiv\sup_{\alpha\in R}\left(\min_{x\in\Omega}\frac{v(x;\alpha)}{\int_\Omega dy\,g(x,y)f(y,v(y;\alpha))}\right); \tag{4.5a}$$

we expect that the value

$$\lambda_c\equiv\lambda_c\,[u_n(x;\alpha)] \tag{4.5b}$$

should approximate the critical value λ_c mentioned in property 2. On the other hand, we expect that (4.5) singles out a point, which can be substituted in (4.4) as x_0. For instance, in case we are dealing with "autonomous" equations of the form

$$-\Delta u=f(u) \tag{4.6a}$$

in a ball,

$$\Omega=B_R\equiv\{x\in R^m:\|x-a\|<R\}\,, \tag{4.6b}$$

choosing the trial function $u_0(x;\alpha)\equiv\alpha$, we have

$$\begin{aligned}\lambda_c=\lambda_c\,[u_0\equiv\alpha]&=\sup_R\frac{\alpha}{f(\alpha)}\min_{B_R}\left[\int_{B_R}dy\,g(x,y)\right]^{-1}\\&=\sup_R\frac{\alpha}{f(\alpha)}\left[\int_{B_R}dy\,g(a,y)\right]^{-1}.\end{aligned} \tag{4.7}$$

In this case in the lowest order of Picard's method, we get from (4.4) the following nonlinear algebraic equation for the determination of $\hat{\alpha}=\hat{\alpha}(\lambda)$:

$$u_0(a;\alpha)=\lambda\int_{B_R}dy\,g(a,y)f(y,u_0(y;\alpha))\,. \tag{4.8}$$

A. The equation of Bratu

To demonstrate the utility of the method, we consider the simple example of the equation of Bratu; that is, the following nonlinear integral equation[17,18]:

$$u(x)=\lambda\int_a^b dy\,g(x,y)e^{u(y)},\quad x\in(a,b)\,, \tag{4.9a}$$

where the kernel is a Green's function defined as follows:

$$g(x,y)=\frac{(b-x)(y-a)}{b-a},\quad y\leqslant x\,,$$

$$= \frac{(b-y)(x-a)}{b-a}, \quad y \geqslant x. \tag{4.9b}$$

One can show that the function $u(x)$ which satisfies this equation is any solution of the differential equation:

$$\frac{d^2u}{dx^2} + \lambda e^u = 0, \tag{4.10a}$$

which also satisfies the two-point boundary condition

$$u(a) = 0 = u(b). \tag{4.10b}$$

Now we shall treat the problem by the refined Picard iteration. We expect to reproduce in the limit of vanishing λ the perturbative result, that is,

$$\begin{aligned} \mathbf{u}(\lambda;x) &= \lambda(x/2)(b-x) \\ &\quad + \lambda^2(x/24)(b-x)(b^2+bx-x^2) \\ &\quad + \lambda^3(x/1440)(b-x)(9b^4+9b^3x-b^2x^2 \\ &\quad - 16bx^3+8x^4) + \cdots, \end{aligned} \tag{4.11}$$

for $x \in (0,b)$, where for convenience it is assumed that $a=0$.

It can be proved that the series (2.4b) converges uniformly to the unique solution of (1.1) for

$$\lambda < \rho/4FG\mu(\Omega), \quad x \in \Omega, \quad |v| < \rho, \tag{4.12a}$$

where

$$F \equiv \sup_{\Omega \times [0,\rho)} f(x,v), \tag{4.12b}$$

$$G \equiv \sup_{\Omega \times \Omega} g(x,y). \tag{4.12c}$$

The solution is unique in the sense that for small λ there is no other solution of (1.1) that lies in a sufficiently small neighborhood of the zero function throughout Ω. Since e^v is an entire function of v, we get from (4.12) that the radius of convergence of the series (4.11) is

$$\lambda < \max_{\rho > 0} (\rho/b^2 e^\rho) = 1/b^2 e. \tag{4.13}$$

Now, let us employ the simplest possible version of Rayski's method, that is,

$$u_0(x;\alpha) \equiv \alpha \quad \text{(a constant parameter)}; \tag{4.14}$$

and let us perform only one iteration (2.2). Then we get from (4.7) the following approximate result for the critical value of λ:

$$\lambda_c = \lambda_c[u_0 \equiv \alpha] = \sup_R(\alpha e^{-\alpha})\left[\int_0^b dy\, g\left(\frac{b}{2},y\right)\right]^{-1} = \frac{8}{b^2 e}. \tag{4.15}$$

Fitting the parameter α from Eq. (4.8) in the point $x_0 = b/2$, we arrive at the following equation:

$$\alpha e^{-\alpha} = \lambda(b^2/8). \tag{4.16}$$

In view of the above considerations we expect that there exists a solution of Eq. (4.16) satisfying the following condition:

$$\lim_{\lambda \to 0} \alpha(\lambda) = 0. \tag{4.17}$$

Apparently, this choice of the parameter α should reproduce the results obtained by the standard Picard iteration. Indeed, one can verify that, if the rhs of Eq. (4.16) is a number less than $1/e$, then the smallest root of this equation is given by the convergent expansion

$$\hat{\alpha}_0(\lambda) = \sum_{m=1}^{\infty} \frac{m^{m-1}}{m!}\left(\lambda \frac{b^2}{8}\right)^m. \tag{4.18}$$

On the other hand, it can be easily seen that there exists a second root of Eq. (4.16), $\hat{\alpha}_\infty = \hat{\alpha}_\infty(\lambda)$, which can be determined from the formula

$$\delta = [\hat{\alpha} - (\lambda b^2/8)e^{\hat{\alpha}}]/(\hat{\alpha}-1), \tag{4.19a}$$

where

$$\hat{\alpha}_\infty = \hat{\alpha} + \delta. \tag{4.19b}$$

This formula can be used as an iterative device in which $\hat{\alpha}$ is replaced by successively determined values. Thus we see that by successive applications of formulas (4.18) and (4.19) one can obtain $\hat{\alpha}_{0,\infty}$ to any specified accuracy. The important thing to note is that the second solution of (4.16) is nonperturbative,

$$\hat{\alpha}_\infty(\lambda) = O(-\log \lambda b^2), \quad \text{as } \lambda \to 0, \tag{4.20}$$

and that as $\lambda \to \lambda_c$, the two roots of Eq. (4.16) tend toward a limiting root $\hat{\alpha}_c$ between them. When $\lambda = \lambda_c$, the limit root is the unique solution of Eq. (4.16). For $\lambda > \lambda_c$, Eq. (4.16) has no real solution.

Thus one may conjecture that the bifurcation picture that follows from the above analysis holds for solutions of problem (4.9), and the point $[\lambda_c, u(x,\lambda_c)]$ is then a turning point. Hence, in the case of Bratu's equation, (4.15) yields the estimate for the critical value of λ mentioned in property 2; (4.18) and (4.19) provide the approximations for solutions, that is,

$$\underline{\hat{u}}(\lambda;x) = \hat{\alpha}_0(\lambda)(4/b^2)x(b-x), \tag{4.21a}$$

$$\bar{\hat{u}}(\lambda;x) = \hat{\alpha}_\infty(\lambda)(4/b^2)x(b-x). \tag{4.21b}$$

Moreover, we expect that (4.18) and (4.19) exhibit L^∞ bounds for both solutions of (4.9) as $\lambda \to 0$,

$$\|\underline{u}\|_\infty = O(\hat{\alpha}_0), \tag{4.22a}$$

$$\|\bar{u}\|_\infty = O(\hat{\alpha}_\infty). \tag{4.22b}$$

Comparing the above conjectures with the results of a theorem of Bratu,[17] we are able to confirm them.[5,19] The value of $\lambda_c b^2$ obtained in (4.15) is within 16% of its correct value 3.513... . On the other hand, (4.14) is a very bad approximation to $u(x)$, which vanishes for $x=0$ and $x=b$. One may take the polynomial initial guess $\tilde{u}_0(x;\alpha) = \alpha x(b-x)$, and a somewhat tedious calculation yields 3.448... , a value of $\lambda_c b^2$, which lies within 1.9% of its correct value (in the lowest-order approximation).

B. The Gel'fand problem

This section deals with a particular case of problem (1.1), the so-called Gel'fand problem,[5,20]

$$-\Delta u = \lambda e^u, \quad \text{in } \Omega \subset R^2, \tag{4.23a}$$

$$u = 0, \quad \text{on } \partial\Omega. \tag{4.23b}$$

If $\Omega = \Omega_R \equiv \{(x,y) \in R^2 \colon r \equiv |z| < R\}$ (circle), where the complex variable $z = x + iy$, the radially symmetric perturbative result reads

$$\begin{aligned}\underline{u}(\lambda;r) = \lambda\,[(R-r)/4](R+r) &+ \lambda^2[(R-r)/64](3R^3+3R^2r-Rr^2-r^3)\\ &+\lambda^3[(R-r)/384](5R^5+5R^4r-\tfrac{5}{2}R^3r^2\\ &-\tfrac{5}{2}R^2r^3+\tfrac{1}{2}Rr^4+\tfrac{1}{2}r^5)+\cdots\,,\end{aligned}\tag{4.24}$$

and the same analysis as in Sec. IV A can be carried out. We get the following approximate critical value of λ:

$$\lambda_c = 4/R^2 e\,, \tag{4.25}$$

and the approximations for solutions

$$\hat{\underline{u}}(\lambda;x) = \hat{\alpha}_0(\lambda)(1/R^2)(R^2-r^2)\,, \tag{4.26a}$$

$$\hat{\bar{u}}(x) = \hat{\alpha}_\infty(\lambda)(1/R^2)(R^2-r^2)\,, \tag{4.26b}$$

where $\hat{\alpha}_{0,\infty}$ are the two roots of the following equation:

$$\alpha e^{-\alpha} = \lambda(R^2/4)\,. \tag{4.27}$$

In this case explicit justification is possible. Introducing the complex variables z and $\bar{z}$, we can rewrite Eq. (4.3a) in the following form:

$$-\frac{\partial^2}{\partial z\,\partial \bar{z}}\,u = \frac{\lambda}{4}\,e^u\,. \tag{4.28a}$$

It was shown by Liouville[21] that the general solution of Eq. (4.28a) is

$$u(z,\bar{z}) = \log\{|f'(z)|^2/[1+(\lambda/8)|f(z)|^2]^2\}\,, \tag{4.28b}$$

where $f(z)$ is an analytic function, which can be determined with the help of the boundary condition. In the case of the null Dirichlet problem on Ω_R, the radially symmetric solutions can be calculated explicitly,[5,21]

$$u_i(\lambda;r) = \log b_i/[1+(\lambda b_i/8)r^2]^2\,, \tag{4.29a}$$

where

$$b_i = \frac{32}{\lambda^2R^4}\left[1-\frac{\lambda R^2}{4}+(-1)^i\left(1-\frac{\lambda R^2}{2}\right)^{1/2}\right],\quad i=1,2\,. \tag{4.29b}$$

Obviously, symmetric solutions exist only if

$$\lambda \leqslant \lambda_c = 2/R^2\,. \tag{4.30}$$

It follows from a uniqueness argument that the minimal solution has the same symmetries as the problem itself. This means that the minimal solution $\underline{u}(\lambda;x)$ is radially symmetric and corresponds to $u_1(\lambda;r)$. It can be shown that the circle has only radially symmetric solutions.[5,21]

It is easy now to check explicitly that in the limit of vanishing λ we have

$$\|\underline{u}\|_\infty = O(\hat{\alpha}_0),\quad \|\bar{u}\|_\infty = O(\hat{\alpha}_\infty)\,. \tag{4.31}$$

However, for $m>2$ or for arbitrary Ω, the explicit justification analogous to that just described is impossible and so qualitative discussion is required.[20,22]

V. CONCLUDING REMARKS

The results of this paper can be extended to a more general class of the nonlinear Dirichlet problem. Since certain generalizations are immediate, we have restricted ourselves to problem (1.1) in order to keep the paper reasonably short. Moreover, if we define the Nemytskii operator $F(v)(x) \equiv f(x,v(x))$, and let $GF(v) \equiv \int_\Omega dy\, g(x,y) \times f(y,v(y))$, it can be shown[23] that problem (1.1) is equivalent to the Hammerstein equation $u = \lambda GF(u)$ in the Banach space $C(\bar{\Omega})$ with norm $\|u\|_\infty = \max_{x\in\Omega}|u(x)|$. It is clear now that possible applications can be extended to multidimensional integral equations of the form $u = \lambda\tilde{G}\tilde{F}u$. We wish to mention at this point that the refined iteration is especially worth recommending for iterative Lipschitzian operators[24] T with the Lipschitz constant $\tau < \infty$,

$$\|T(Tv) - Tv\| \leqslant \tau\|Tv - v\|\,,$$

and for iterative methods of higher orders.[25]

Let us end with the remark that there are some recent techniques that often appear able to overcome the attractive or repulsive properties of solutions if the starting points are chosen sufficiently near a solution. Hence it would appear that in general it would be most efficient to use a search algorithm, such as the refined iteration, to obtain initial approximations and subsequently approach the solutions by applying a fast iterative method.[26]

ACKNOWLEDGMENTS

The author wishes to thank Professor Jerzy Rayski for suggesting this problem and for helpful discussions.

This work was supported in part under Project CPBP 01.03, No. 1.7/87.

[1] J. Rayski, Acta Phys. Polon. B **85**, 631 (1974).

[2] J. Rayski, J. Comput. Appl. Math. **3**, 31 (1977); S. Kopeć, Acta Phys. Polon. A **65**, 471 (1984); **68**, 775 (1985); Phys. Scr. **34**, 353 (1986).

[3] F. Mrugała and G. Staszewska, Acta. Phys. Polon. A **54**, 99 (1978); G. Staszewska, *ibid.* **56**, 317 (1979); Chem. Phys. **58**, 93 (1981); **59**, 289 (1981); G. Staszewska and T. Orlikowski, Acta Phys. Polon. A **63**, 531 (1983).

[4] T. Kusano and S. Oharu, Indiana Univ. Math. J. **34**, 85 (1985).

[5] C. Bandle, *Isoperimetric Inequalities and Applications* (Pitman, London, 1980).

[6] M. G. Crandall and P. H. Rabinowitz, Arch. Rational Mech. Anal. **58**, 207 (1985).

[7] C. Bandle, Arch. Rational Mech. Anal. **58**, 219 (1975).

[8] D. D. Joseph and T. S. Lundgren, Arch. Rational Mech. Anal. **49**, 241 (1973).

[9] H. B. Keller and D. S. Cohen, J. Math. Mech. **16**, 1631 (1967).

[10] T. Laetsch, J. Math. Mech. **18**, 1905 (1969).

[11] H. Amann, *Nonlinear Operators in Ordered Banach Spaces and Some Applications to Nonlinear Boundary Value Problems* (Springer, New York, 1976).

[12] See, e.g., J. L. Lions and E. Magenes, *Nonhomogeneous Boundary Value Problems and Applications* (Springer, Berlin, 1972), Vol. 1 imposes C^∞ regularity on $\partial\Omega$, but notes that this is excessive and we invoke the same spirit without going into details (see also Ref. 9).

[13] R. S. Cantrell and C. Cosner, Indiana Univ. Math. J. **34**, 517 (1985).

[14] M. S. Berger, *Nonlinearity and Functional Analysis* (Academic, New York, 1977); M. A. Krasnoselskii, *Positive Solutions of Operator Equations* (Noordhoff, Groningen, 1964).

[15] J. P. Keener and H. B. Keller, J. Differential Equations **16**, 103 (1974).

[16] L. Collatz, *Numerical Treatment of Differential Equations* (Springer, Berlin, 1960).

[17] H. T. Davis, *Introduction to Nonlinear Differential and Integral Equations* (Dover, New York, 1962).

[18] P. B. Bailey, L. F. Shampine, and P. E. Waltman, *Nonlinear Two-Point Boundary Value Problems* (Academic, New York, 1968).

[19] T. Laetsch, Indiana Univ. Math. J. **20**, 1 (1970).

[20] I. M. Gel'fand, Sov. Usp. Mat. Nauk **14**, 87 (1959) [Am. Math. Soc. Transl. **29**, 295 (1963)].

[21] J. Liouville, J. Math. (Paris) **18**, 71 (1853); C. Bandle, Arch. Rational Mech. Anal. **58**, 219 (1975).

[22]F. Mignot, F. Murat, and J. P. Puel, "Variation d'un point de retournement par rapport au domain," in *Lecture Notes in Mathematics*, Vol. 782 (Springer, New York, 1980); D. D. Joseph and T. S. Lundgren, Arch. Rational Mech. Anal. **49**, 241 (1973).
[23]S. Agmon, A. Douglis, and L. Nirenberg, Commun. Pure Appl. Math. **12**, 623 (1959); D. H. Sattinger, "Topics in stability and bifurcation theory," in *Lecture Notes in Mathematics*, Vol. 333 (Springer, New York, 1973).
[24]J. M. Ortega and W. C. Rheinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables* (Academic, New York, 1970).
[25]B. Noble, "The numerical solution of nonlinear integral equations and related topics," in *Nonlinear Integral Equations*, edited by P. M. Anselone (Univ. Wisconsin Press, Madison, WI, 1964).
[26]R. E. Bellman and R. E. Kalaba, *Quasilinearization and Nonlinear Boundary Value Problems* (Elsevier, New York, 1965).

General form of the equation of motion for a point charge

Abraham Lozada[a)]

Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela

(Received 24 May 1988; accepted for publication 4 January 1989)

In this paper, using a distribution theory, the question of the compatible phenomenology associated with the electromagnetism of a classical point charge is addressed. The compatibility is restricted to the conservation laws of energy momentum and angular momentum. A general form of the equation of motion for a point charge is derived. Renormalization is not required. Previous results are discussed. It is shown that a distribution approach yields a deep insight into the problem.

I. INTRODUCTION

There are in the literature (see, e.g., Refs. 1–4), several equations to describe the motion of a point charge. Except for a few articles (see, e.g., Refs. 5–8), the overwhelming majority of works (see, e.g., Refs. 9–15) has been developed leading to the Lorentz–Dirac equation[1] as the equation of motion for a classical point charge. This circumstance could be misleading in the sense that the classical electrodynamics is compatible only with one equation of motion for a point charge.

A long time ago, an important fact related to the equation of motion of a point charge was addressed by Poincaré[16] and Von Laue.[17] They remarked that a purely electromagnetic charged body was impossible (classically) because the electric charge distribution by itself is unstable. They postulated that a phenomenological tensor $K^{\mu\nu}$, should be added to the electromagnetic energy-momentum tensor $\theta^{\mu\nu}_{\text{elm}}$, to give the total tensor

$$\theta^{\mu\nu} = K^{\mu\nu} + \theta^{\mu\nu}_{\text{elm}} . \qquad (1.1)$$

The phenomenological tensor $K^{\mu\nu}$ should be such that

$$\partial_\nu \theta^{\mu\nu} = 0 . \qquad (1.2)$$

In principle, if $\theta^{\mu\nu}_{\text{elm}}$ is given, the compatible $K^{\mu\nu}$ tensor is not unique.[18] For the case of a classical point charge the Poincaré–Von Laue's proposal raises the following question: *Which is the compatible phenomenology associated with the electromagnetism of a classical point charge?* In previous works, this question has not been totally considered, and only partial answers (through particular equations) have been obtained. Furthermore, many of these partial answers have the problem that they suppose a phenomenology that remains hidden behind unconventional mathematical procedures (such as cutoff procedures and other prescriptions). The main purpose of this work is to give a general answer to the question that has arisen because of Poincaré–Von Laue's proposal. For this task, a general enough phenomenology is explicitly exposed in order to obtain a theory that conserves energy momentum and angular momentum. Also, the additional restrictions that are necessary to extract some particular equations of motion are explicitly shown. In this sense, equations of various authors (see, e.g., Refs. 1–4) are particular cases of the general theory discussed in this work.

To extract an equation of motion, two methods have mainly been used. (i) The first method (see, e.g., Refs. 1, 4–9, 11–13) is based on classical field theory. This point of view has its origin in Dirac's formulation.[1] In this approach, the electromagnetic field is considered as a mathematical function of the space-time points even though the source in the Maxwell's equations is not a mathematical function. (ii) A mathematical distribution theory is used to discuss the closed theory of interaction between the point charge and the electromagnetic field.[14,15] In this approach, the field (as the source) is not a function of the space-time points, instead, it is a well-defined distribution.

In this paper, using a distribution theory for the energy-momentum tensor, it is proven that many equations of motion for a point charge are compatible with the classical electromagnetic theory. In fact, using a constructive form of the Hahn–Banach theorem, a general form of the energy-momentum tensor distribution is found that describes the interaction between a point charge and the electromagnetic field. From this distribution theory, a general form of the equation of motion of a point charge is derived. The general form yields well-known equations and new equations, moreover, it includes, as a particular case, a general form found previously.[4] To extract the equation of motion, energy momentum and angular momentum conservations are used. Since a rigorous distribution theory is employed, everything is finite, and no renormalization[1,13] is required. Also, it is shown that distribution theory offers a clearer insight than other approaches, since it shows that the possibility of extracting different equations is directly related to the different compatible phenomenologies (with electromagnetism) for a point charge.

With respect to the notation, the metric tensor g will have a signature of $+2$, and the speed of the light is taken as 1. When it will be convenient, indices on vectors and tensors will be omitted and scalar products will be indicated by a dot. Parentheses $(\cdot,\cdot)$ or brackets $[\cdot,\cdot]$ will denote symmetrization or antisymmetrization, respectively, of the enclosed indices (without a factor of one-half). The charge world line (CWL) is $z(\tau)$, where τ is the proper time, and $v(\tau) \equiv v$ ($v^2 = -1$) and $a(\tau) \equiv a (v\cdot a = 0)$ are the four-velocity and acceleration, respectively. Retarded coordinates will be used here (see, e.g., Refs. 13 and 19). Then for any space-time point x, we define $R \equiv x - z(\tau_r)$, $R^2 = 0$ $(R^0 > 0)$, $\rho \equiv -v\cdot R$, $u \equiv R/\rho - v$, τ_r being the value of the proper time on the intersection between the light cone, with the apex at x opening into the past, and the CWL.

[a)] Postal address: A. P. 47399, Caracas 1041-A, Venezuela.

 0022-2488/89/081713-08$02.50

Throughout this article, the closed region $\overline{\Omega}(\epsilon)$ shown in Fig. 1 will be used. In this figure, $B(E)$ and $B(\epsilon)$ represent two bounded segments of two Bhabha tubes,[10,13] of radius E and ϵ, respectively, where $0 \leqslant \epsilon < E < \infty$ and $C(\tau_i)$ represents a bounded segment of the future light cone with the apex at $z(\tau_i)$, with $i = 1,2$. By $\Omega(\epsilon)$ we shall mean the union of all the open sets contained in $\overline{\Omega}(\epsilon)$, for $\epsilon = 0$ we write $\Omega(0) \equiv \Omega$.

We adopt, as the set of test functions, the set of all functions that are infinitely differentiable and have compact support[20,21] in $\Omega(\epsilon)$. Topologized in the usual way,[20] this set will be denoted by $\mathscr{D}[\Omega(\epsilon)]$. A defined and continuous linear functional on $\mathscr{D}[\Omega(\epsilon)]$ is called a distribution[20,21] in $\Omega(\epsilon)$. It will be seen that this election of $\Omega(\epsilon)$ will facilitate the future calculations without restricting the generality of the theory.

The electromagnetic field $F^{\mu\nu}(x)$ satisfies Maxwell's equations:

$$\partial_\nu F^{\mu\nu}(x) = 4\pi j^\nu(x)\,, \tag{1.3}$$

$$\partial^\alpha F^{\beta\gamma} + \partial^\beta F^{\gamma\alpha} + \partial^\gamma F^{\alpha\beta} = 0\,, \tag{1.4}$$

where the current for a point particle with a charge e is defined by

$$j(x) = \int_{\tau_1}^{\tau_2} d\tau\, ev(\tau)\delta[x - z(\tau)]\,,$$

$$(j,\phi) \equiv \int_{\tau_1}^{\tau_2} d\tau\, ev(\tau)\phi[z(\tau)], \quad \text{for } \phi \in \mathscr{D}(\Omega)\,. \tag{1.5}$$

The electromagnetic field can be decomposed as

$$F = F_{\text{ext}} + F_{\text{ret}}\,, \tag{1.6}$$

where the nonsingular part F_{ext} satisfies Maxwell's equations for vacuum and F_{ret} is the retarded distribution solution of Maxwell's equations. Taylor[14] discussed this solution (the retarded Lienard–Wiechert fields, see, also, Ref. 19). It is known[14,19] that F_{ret} defines a regular distribution in the whole space-time, in particular, in Ω. This means that the value of the distribution at the testing function $\phi \in \mathscr{D}(\Omega)$ is

$$(F_{\text{ret}},\phi) \equiv \int_\Omega F_{\text{ret}}(x)\phi(x)d^4x\,. \tag{1.7}$$

The components of the total electromagnetic energy-momentum tensor for a point charge and an external electromagnetic field are

$$\theta^{\mu\nu}_{\text{elm}} = (1/4\pi)(F^{\mu\alpha}F^{\nu}{}_{\alpha} - \tfrac{1}{4}g^{\mu\nu}F^{\alpha\beta}F_{\alpha\beta})\,, \tag{1.8}$$

where $F^{\mu\nu}$ is given by (1.6). Equation (1.8) defines a regular distribution in $\Omega(\epsilon)$, whenever $\epsilon > 0$. Corresponding to the superposition shown in (1.6) we obtain (in obvious notation) the following splitting:

$$\theta_{\text{elm}} = \theta_{\text{ext}} + \theta_{\text{mix}} + \theta_{\text{ret}}\,. \tag{1.9}$$

Since F_{ext} is not singular, θ_{ext} and θ_{mix} define regular distributions[10] in Ω. The problem to be solved is to define θ_{ret} as a distribution in Ω. In retarded coordinates, in $\Omega(\epsilon)$, $\epsilon > 0$,

$$\theta_{\text{ret}} = \frac{e^2}{4\pi}\left(\frac{g}{2} + vv - uu\right)\frac{1}{\rho^4} + \frac{e^2}{4\pi}\left(a - a\cdot uu, \frac{R}{\rho}\right)\frac{1}{\rho^3} + \frac{e^2}{4\pi}[a^2 - (a\cdot u)^2]\frac{RR}{\rho^4}\,. \tag{1.10}$$

In this paper θ_{ret} will be split with respect to its integrability in Ω, then

$$\theta_{\text{ret}} = \theta_{ni} + \theta_i\,, \tag{1.11}$$

where

$$\theta_i = (e^2/4\pi)[a^2 - (a\cdot u)^2](RR/\rho^4)\,. \tag{1.12}$$

The symmetric tensor θ_i, although singular (as θ_{mix}), defines a regular distribution[10] in Ω. The real difficulty is to define θ_{ni} as a distribution in Ω.

The present work is organized as follows. In Sec. II, the most-general distribution form of the total energy-momentum tensor for a point charge is considered. In Sec. III, basic restrictions are imposed on the distribution theory obtained in Sec. II. From these restrictions, a general form for the equation of motion of the charge is extracted. In Sec. IV, additional conditions are imposed in order to derive particular equations of motion. Section V contains some concluding remarks and comments. Finally, Rowe's[10,19] method of calculating divergences is discussed in the Appendix.

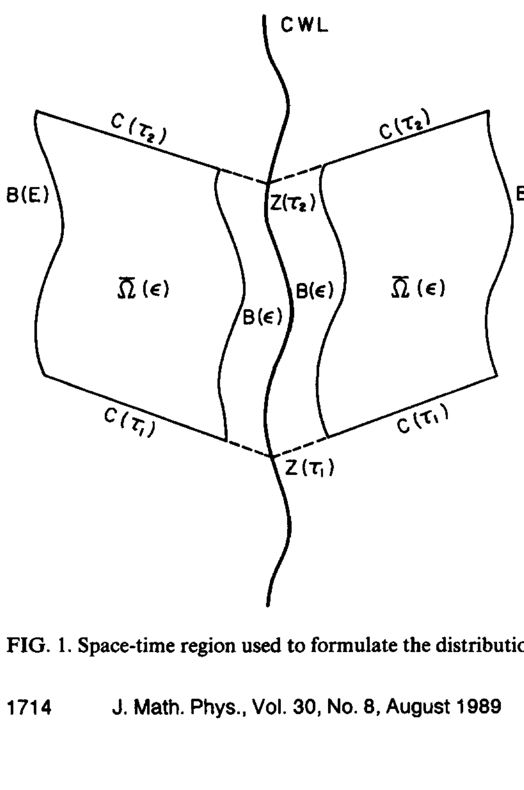


FIG. 1. Space-time region used to formulate the distribution theory.

II. DISTRIBUTION DEFINITION OF THE ENERGY-MOMENTUM TENSOR

In this section, the distribution form of the total energy-momentum tensor θ, for a point charge in an external electromagnetic field, is derived.

Equation (1.1) can be written as

$$\theta = K + \theta_{ni} + \theta_i + \theta_{\text{mix}} + \theta_{\text{ext}}\,. \tag{2.1}$$

To complete the distribution definition in Ω of θ, the distribution form of θ_{ni} and K should be determined. In order to do this, let us use the fact that θ_{ni} defines a regular distribution in $\Omega(\epsilon)$, $\epsilon > 0$. That is,

$$(\theta_{ni},\phi)=\int_{\Omega(\epsilon)}\theta_{ni}\phi\,d^4x,\quad \text{for } \phi\in\mathscr{D}[\Omega(\epsilon)],\quad \epsilon>0\,. \tag{2.2}$$

Our problem is to extend this functional onto the whole space, $\mathscr{D}(\Omega)$, in such a manner that the extended functional is linear and continuous on $\mathscr{D}(\Omega)$, and to determine the degree of arbitrariness of such an extension. According to the Hahn–Banach theorem, the linear functional θ_{ni} defined and continuous on $\mathscr{D}[\Omega(\epsilon)]$, $\epsilon>0$, has extensions to all of $\mathscr{D}(\Omega)$ which we shall call "renormalizations". (The term "renormalization" will be used only because the spirit of the distribution theory to be developed here is similar to the renormalization procedure of Bogoliubov, Parasiuk, and Hepp[22] in quantum field theory.) Since (2.2) defines a distribution of finite order[21] in $\Omega(\epsilon)$, it will be required that any "renormalization" preserves this property.

An extension of (2.2) for $\phi\in\mathscr{D}(\Omega)$ is easily obtained. In fact, considering that[15] in $\Omega(\epsilon)$, $\epsilon>0$,

$$\theta^{\mu\nu}_{ni}=\frac{e^2}{4\pi}\partial^2\left(\frac{R^\mu R^\nu}{\rho^4}\right)+\frac{e^2}{4\pi}\partial^\alpha\left[\left(a_\alpha+\frac{a\cdot R}{\rho}v_\alpha\right)\frac{R^\mu R^\nu}{\rho^4}\right], \tag{2.3}$$

we can obtain by partial integration of (2.2) that

$$(\theta_{ni},\phi)=\frac{e^2}{16\pi}\int_\Omega\frac{RR}{\rho^4}\partial^2\phi\,d^4x-\frac{e^2}{4\pi}\times\int_\Omega\frac{RR}{\rho^4}\left[a+\frac{(a\cdot R)}{\rho}v\right]\cdot\partial\phi\,d^4x,\quad \text{for } \phi\in\mathscr{D}[\Omega(\epsilon)],\quad \epsilon>0, \tag{2.4}$$

where Ω can be written for the region of integration [instead of $\Omega(\epsilon)$] since $\phi\in\mathscr{D}[\Omega(\epsilon)]$, $\epsilon>0$. Moreover, the right-hand side of (2.4) is defined for functions $\phi\in\mathscr{D}(\Omega)$. Then, Eq. (2.2) could be considered as a restriction of the following functional:

$$(\theta_{\text{reg}},\phi)=\frac{e^2}{16\pi}\int_\Omega\frac{RR}{\rho^4}\partial^2\phi\,d^4x-\frac{e^2}{4\pi}\int_\Omega\frac{RR}{\rho^4}\left[a+\frac{(a\cdot R)}{\rho}v\right]\cdot\partial\phi\,d^4x,\quad \text{for } \phi\in\mathscr{D}(\Omega)\,. \tag{2.5}$$

This functional is of finite order and is linear and continuous on $\mathscr{D}(\Omega)$. Therefore, (2.5) provides one of the "renormalizations" that should be looked for.[15]

Obviously, this continuation is not unique. If θ_{ren} is another "renormalization," then necessarily,

$$(\theta_{\text{ren}}-\theta_{\text{reg}},\phi)=0,\quad \text{for } \phi\in\mathscr{D}[\Omega(\epsilon)],\quad \epsilon>0\,. \tag{2.6}$$

That is, the distribution $\theta_{\text{ren}}-\theta_{\text{reg}}$ vanishes in $\Omega(\epsilon)$, $\epsilon>0$, then we have that $\theta_{\text{ren}}-\theta_{\text{reg}}$ vanishes in $U_{0<\epsilon<E}\Omega(\epsilon)$.[20] By the definition of the support of a distribution,[20] this means that the support of $\theta_{\text{ren}}-\theta_{\text{reg}}$ is at the segment of the CWL included in Ω.

Summarizing, it has been shown that *the most general "renormalization"* of (2.2) is

$$\theta_{\text{ren}}=\theta_{\text{reg}}+\Lambda\,, \tag{2.7}$$

where Λ is a distribution of finite order in Ω having support at the CWL.

For physical reasons, distributions whose supports consist of isolated points will not be considered. Then, let us consider the following general form of Λ:

$$\Lambda=\sum_{|\alpha|<N'}\int_{\tau_1}^{\tau_2}d\tau\,\Delta_\alpha(\tau)D^\alpha\delta[x-z(\tau)]\,, \tag{2.8}$$

where

$$D^\alpha\equiv\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\cdots\partial x_4^{\alpha_4}}\,,$$

α_j are non-negative integers such that $\Sigma_{j=1}^4\alpha_j=|\alpha|$; $N'<\infty$ is the order of Λ in Ω and $\Delta_\alpha(\tau)$ are integrable functions on (τ_1,τ_2) [additional restrictions on $\Delta_\alpha(\tau)$ will come when conservation laws will be examined]. It is easy to show that (2.8) [of which (1.5) is a particular case] defines a distribution in Ω. In fact, if $\mathscr{B}$ is a bounded set of $\mathscr{D}(\Omega)$,[20] it can be shown that $|(\Lambda,\phi)|<C$, $\forall\phi\in\mathscr{B}$, from which, it follows that Λ defines a distribution in Ω.[20]

Finally, the phenomenological energy-momentum tensor distribution K must satisfy:

$$(K,\phi)=0,\quad \text{for } \phi\in\mathscr{D}[\Omega(\epsilon)],\quad \epsilon>0\,. \tag{2.9}$$

Then, the general form of K is also of the type given in (2.8) (a distribution of finite order in Ω). In this manner, Eq. (2.1) can be written in the following form:

$$\theta=\sum_{|\alpha|<N}\int_{\tau_1}^{\tau_2}d\tau\,\Gamma_\alpha(\tau)D^\alpha\delta[x-z(\tau)]+\theta_{\text{reg}}+\theta_i+\theta_{\text{mix}}+\theta_{\text{ext}}\,, \tag{2.10}$$

where, the arbitrariness in the "renormalization" of θ_{ni} has been left to the material energy-momentum tensor K (in order to guarantee this, we must have $N\geqslant N'$). Then, according with Eq. (2.1), without restrictions on the θ definition, K and θ_{ni} can be chosen as

$$K=\sum_{|\alpha|<N}\int_{\tau_1}^{\tau_2}d\tau\,\Gamma_\alpha(\tau)D^\alpha\delta[x-z(\tau)] \tag{2.11}$$

and

$$\theta_{ni}=\theta_{\text{reg}}\,. \tag{2.12}$$

In this case the restriction $N\geqslant N'$ [N' is defined through Eqs. (2.7) and (2.8)] is equivalent to the trivial restriction $N\geqslant 0$. Equation (2.10) provides us with the general distribution form of the total energy-momentum tensor θ that will be considered in this work. The concrete value of the index N will be discussed in the next section. The conditions that allow the determination of N represent an important part of the theory.

III. BASIC REQUIREMENTS AND GENERAL FORM OF THE EQUATION OF MOTION

In this section, simple requirements are imposed on the general distribution definition of θ derived in Sec. II. From these restrictions, the general form of the equation of motion of a point charge is extracted.

Let us define the tensor:

$$M^{\lambda\mu\nu}\equiv x^\lambda\theta^{\mu\nu}-x^\mu\theta^{\lambda\nu}\,. \tag{3.1}$$

Equation (3.1) gives us a well-defined distribution because the multipliers x are infinitely differentiable.

Let us demand that:

(a) The theory conserves energy momentum, that is,

$$(\partial_\nu \theta^{\mu\nu},\phi)=0, \quad \text{for } \phi\in\mathscr{D}(\Omega). \tag{3.2}$$

(b) The theory conserves angular momentum, that is,

$$(\partial_\nu M^{\lambda\mu\nu},\phi)=0, \quad \text{for } \phi\in\mathscr{D}(\Omega). \tag{3.3}$$

(c) The index N [see, Eq. (2.10)] should have the lowest possible value consistent with (a) and (b).

Mathematically, condition (c) means that the order of the distribution K given by (2.11) is the lowest consistent with (3.2) and (3.3). Physically, condition (c) means that θ is "as little as possible" phenomenological.

It can be shown, that $N=0$ is inconsistent with (3.2) and the definition of the metric tensor g. The next order to consider is $N=1$. In this case, Eq. (2.10) gives

$$K^{\mu\nu}=\int_{\tau_1}^{\tau_2} d\tau\{\Gamma^{\mu\nu}(\tau)+\Gamma^{\mu\nu\lambda}(\tau)\partial_\lambda\}\delta[x-z(\tau)]. \tag{3.4}$$

We suppose that $\Gamma^{\mu\nu\lambda}v_\lambda(\tau)=0$ (this is not a restriction because an integration by parts can achieve it).

The technique for calculating the divergences of θ is explained in the Appendix. Equation (3.2) implies

$$\begin{aligned}(\partial_\nu\theta^{\mu\nu},\phi)=\int_{\tau_1}^{\tau_2} d\tau\Big\{&-\Gamma^{\mu\nu}\partial_\nu\phi[z(\tau)]+\Gamma^{\mu\nu\lambda}\partial_\nu\partial_\lambda\phi[z(\tau)]\\ &+\frac{e^2}{4}v^\mu g^{\nu\lambda}\partial_\nu\partial_\lambda\phi[z(\tau)]-e^2v^\mu a\cdot\partial\phi[z(\tau)]\\ &-\frac{e^2}{3}a^\mu v\cdot\partial\phi[z(\tau)]-eF^{\mu\nu}_{\text{ext}}[z(\tau)]v_\nu\phi[z(\tau)]\\ &+\tfrac{2}{3}e^2a^2v^\mu\phi[z(\tau)]\Big\}=0, \quad \text{for } \phi\in\mathscr{D}(\Omega).\end{aligned} \tag{3.5}$$

In order to extract information from (3.5), let us develop $\Gamma^{\mu\nu}$ and $\Gamma^{\mu\nu\lambda}$ as a linear combination of a basis of the respective tensor space. This gives the forms

$$\Gamma^{\mu\nu}=mv^\mu v^\nu+m^\mu v^\nu+n^\nu v^\mu+m_\perp^{\mu\nu}, \tag{3.6}$$

$$\Gamma^{\mu\nu\lambda}=h^{\lambda}v^\mu v^\nu+h^{\mu\lambda}v^\nu+q^{\nu\lambda}v^\mu+h^{\mu\nu\lambda}, \tag{3.7}$$

where each index that does not label a v, labels a quantity orthogonal to v.

Replacing Eqs. (3.6) and (3.7) in (3.5), and integrating by parts, Eq. (3.5) may now be written as

$$\begin{aligned}\int_{\tau_1}^{\tau_2} d\tau\Big\{&\Big[\frac{d}{d\tau}(mv^\mu+m^\mu)+\frac{e^2}{3}\dot a^\mu+\frac{2}{3}e^2a^2v^\mu-\frac{e^2}{4}\dot a^\mu-eF^{\mu\nu}_{\text{ext}}[z(\tau)]v_\nu-\frac{d}{d\tau}(v^\mu\dot h^{\alpha}v_\alpha)-\frac{d}{d\tau}(\dot h^{\mu\alpha}v_\alpha)\Big]\phi[z(\tau)]\\ &-\big[v^\mu n\cdot\partial\phi[z(\tau)]+m_\perp^{\mu\nu}\partial_\nu\phi[z(\tau)]+\tfrac{3}{4}e^2v^\mu a\cdot\partial\phi[z(\tau)]+v^\mu\dot h_\perp^{\lambda}\partial_\lambda\phi[z(\tau)]+a^\mu h^{\lambda}\partial_\lambda\phi[z(\tau)]+\dot h_\perp^{\mu\lambda}\partial_\lambda\phi[z(\tau)]\big]\\ &+\Big[v^\mu q_s^{\nu\lambda}+h_s^{\mu\nu\lambda}+\frac{e^2}{4}v^\mu g_\perp^{\nu\lambda}\Big]\partial_\nu\partial_\lambda\phi[z(\tau)]\Big\}=0,\end{aligned} \tag{3.8}$$

where

$$g_\perp\equiv g+vv, \tag{3.9}$$

$$\frac{dh^\mu}{d\tau}\equiv\dot h^\mu\equiv-\dot h^{\alpha}v_\alpha v^\mu+\dot h_\perp^\mu, \tag{3.10}$$

$$\frac{dh^{\mu\lambda}}{d\tau}\equiv\dot h^{\mu\lambda}\equiv-\dot h^{\mu\alpha}v_\alpha v^\lambda+\dot h_\perp^{\mu\lambda}, \tag{3.11}$$

$$q_s^{\nu\lambda}=(q^{\nu\lambda}+q^{\lambda\nu})/2, \tag{3.12}$$

$$h_s^{\mu\nu\lambda}=(h^{\mu\nu\lambda}+h^{\mu\lambda\nu})/2. \tag{3.13}$$

Using the lemma of the Appendix in (3.8), following similar steps [than those used with (3.2)] with (3.3), and doing some calculations, we obtain:

$$q_s^{\nu\lambda}=-(e^2/4)g_\perp^{\nu\lambda}, \tag{3.14}$$

$$h_s^{\mu\nu\lambda}=0, \tag{3.15}$$

$$h^{\nu\lambda}=q^{\nu\lambda}, \tag{3.16}$$

$$h^{[\mu,\nu]\lambda}=0, \tag{3.17}$$

$$m^\mu=n^\mu=-\dot h_\perp^\mu-(e^2/2)a^\mu+\dot q_A^{\alpha\mu}v_\alpha, \tag{3.18}$$

$$m_\perp^{\mu\nu}=-\tfrac{1}{2}(h^\mu,a^\nu), \tag{3.19}$$

$$[h^{\nu},a^\mu]+2[\dot q_A^{\mu\alpha}v_\alpha,v^\nu]+2\dot q_A^{\mu\nu}=0, \tag{3.20}$$

$$\begin{aligned}&\frac{d(mv^\mu)}{d\tau}-\frac{2}{3}e^2(\dot a^\mu-a^2v^\mu)-\ddot h^\mu(\tau)\\ &\quad-eF^{\mu\nu}_{\text{ext}}v_\nu-2\frac{d}{d\tau}(v^\mu\dot h\cdot v+\dot q_A^{\mu\lambda}v_\lambda)=0,\end{aligned} \tag{3.21}$$

where

$$q_A^{\mu\nu}=(q^{\mu\nu}-q^{\nu\mu})/2. \tag{3.22}$$

From Eqs. (3.20) and (3.21), we see that requirements (3.2) and (3.3) are not enough to determine a unique equation of motion for the charge. Therefore, there are many K tensors satisfying Poincaré–Von Laue's proposal for the particular case of a point charge.

Equations (3.20) and (3.21) are the general form of the equation of motion for a point charge consistent with requirements (a)–(c) (supposing that $\theta_{ni}=\theta_{\text{reg}}$; which is always possible redefining K, as it was done in Sec. II). Given h^μ and $q_A^{\mu\nu}$ in accordance with (3.20), an equation of motion is obtained through Eq. (3.21) (the quantity m is specified by $v\cdot v=-1$).

An equivalent, but clearer, form of Eqs. (3.20) and (3.21) can be obtained. Let us define the following antisymmetric tensor:

$$S^{\mu\nu}=2q_A^{\mu\nu}-[h^\mu,v^\nu], \tag{3.23}$$

from Eq. (3.23) and using $h^\mu v_\mu=q_A^{\mu\nu}v_\nu=0$, it follows that

$$h^\mu=S^{\mu\nu}v_\nu \tag{3.24}$$

and

$$2q_A^{\mu\nu}=S^{\mu\nu}+[S^{\mu\alpha}v_\alpha,v^\nu]. \tag{3.25}$$

Then, in terms of $S^{\mu\nu}(\tau)$, Eqs. (3.20) and (3.21) are written as

$$\dot{S}^{\mu\nu} + [\dot{S}^{\mu\alpha}v_\alpha, v^\nu] = 0\,, \tag{3.26}$$

$$\frac{d(mv^\mu)}{d\tau} - \frac{2}{3}e^2(\dot{a}^\mu - a^2v^\mu) - eF^{\mu\nu}_{\text{ext}}(\tau)v_\nu - \frac{d}{d\tau}(\dot{S}^{\mu\alpha}v_\alpha) = 0, \tag{3.27}$$

where m has been redefined.

In order to understand the physical meaning of the quantities m and S [which appear in Eqs. (3.26) and (3.27)], let us examine the different terms in Eq. (3.27). The term $eF^{\mu\nu}_{\text{ext}}v_\nu$ is the Lorentz force. The term $\frac{2}{3}e^2a^2v^\mu$ is the four-momentum (relative to the proper time of the charge) radiation rate leaving the charge. Therefore, it is reasonable to interpret the remaining terms as representing the four-momentum rate linked to the charge. Then

$$\frac{dp^\mu_b}{d\tau} \equiv \frac{d(mv^\mu)}{d\tau} - \frac{2}{3}e^2\dot{a}^\mu - \frac{d}{d\tau}(\dot{S}^{\mu\alpha}v_\alpha) \tag{3.28}$$

is the bound four-momentum rate. From (3.28) it follows that (except for an additive constant)

$$p^\mu_b = mv^\mu(\tau) - \tfrac{2}{3}e^2a^\mu(\tau) - \dot{S}^{\mu\alpha}v_\alpha\,. \tag{3.29}$$

The bound four-momentum (3.29) has the remarkable property of being a state function (it depends only on the proper time τ), that is, it does not depend on the history of the charge. The quantity m is defined from (3.29), as

$$m = -p^\mu_b v_\mu\,. \tag{3.30}$$

This means that the total energy that remains linked to the charge, at any one instant τ in the particle rest frame, is m. Then m must be interpreted as the experimental mass.

Now, if $\lim_{e\to 0}$ is taken in (3.26) and (3.27), a set of equations for a free particle with spin (with no electromagnetic moment associated with the spin, see, e.g., Ref. 23 and the works cited therein) is obtained. For this reason the tensor S will be called the spin tensor. The possibility of an equation such as (3.27) [which is equivalent to (3.21)] was mentioned in Ref. 15. It should be noted that from the point of view of the distribution theory developed here, Eqs. (3.26) and (3.27) are inextricably linked.

The set of equations (3.20) and (3.21) or (3.26) and (3.27), by themselves, are insufficient to define the particle motion. Therefore, electromagnetism plus the conservation laws (3.2) and (3.3) are insufficient to yield an unique equation of motion; to complete the system of equations, supplementary conditions must be introduced.

IV. ADDITIONAL RESTRICTIONS AND EQUATIONS OF MOTION

In this section, some complementary conditions that should be given in order to obtain particular equations of motion are discussed in the light of the general equation of motion. For this task, some examples are considered.

Before the presentation of some particular cases of Eqs. (3.26) and (3.27), let us note that these equations and the condition $v\cdot v = -1$ leave only three unknown quantities. Particularly, the Eq. (3.26) can be written as

$$(\dot{S})_\perp = 0\,, \tag{4.1}$$

where $(\cdot)_\perp$ means that the tensor is perpendicular to the four-velocity in all the indexes. To complete the system of Eqs. (3.26) and (3.27), let us start by giving values to $(\dot{S})_\parallel$. This is made in the following examples:

(i) Let us impose that

$$\dot{S}\cdot v = 0\,. \tag{4.2}$$

Then, Eqs. (3.26) and (3.27) lead to the conservation of the spin tensor and mass, and to the Lorentz–Dirac equation of motion.[1] So, a *spinning particle* is consistent with the Lorentz–Dirac equation whenever the spin tensor is conserved.[8]

In particular, if $S^{\mu\nu} = 0$, Rowe's theorem[15] about the uniqueness of the Lorentz–Dirac equation is obtained, but *without* additional hypothesis of simplicity.[15]

Substituting $S^{\mu\nu} = 0$ in Eqs. (3.18), (3.19), (3.24), and (3.25), Eqs. (3.6) and (3.7) may be written, in regards to the equation of motion, as

$$\Gamma^{\mu\nu}(\tau) = mv^\mu v^\nu - (e^2/2)(a^\mu, v^\nu)\,, \tag{4.3}$$

$$\Gamma^{\mu\nu\lambda}(\tau) = -(e^2/4)(g^{\mu\lambda}_\perp v^\nu + g^{\nu\lambda}_\perp v^\mu)\,. \tag{4.4}$$

Equations (4.3) and (4.4) represent the minimal phenomenology that must be postulated for deriving the Lorentz–Dirac equation. All previous deduction of this equation (see, e.g., Refs. 9–15) should suppose this kind of point charge. In fact, they must do it,[15] but it generally happens that (4.3) and (4.4) remain hidden behind cutoff procedures and other prescriptions (see, e.g., Tabensky's conditions on the second fundamental form of the surface considered in Ref. 6, p. 272).

(ii) Let us impose that

$$\dot{S}^{\mu\nu}v_\nu = b^\mu + (b\cdot v)v^\mu\,, \tag{4.5}$$

$b^\mu \equiv b^\mu(\tau)$ being any differentiable four-vector (in momentum units). Then, Eqs. (3.26) and (3.27) lead to

$$\dot{S}^{\mu\nu} = [v^\mu, b^\nu]\,, \tag{4.6}$$

$$\frac{d(Mv^\mu)}{d\tau} - \frac{2}{3}e^2(\dot{a}^\mu - a^2v^\mu) - e\,F^{\mu\nu}_{\text{ext}}v_\nu - \dot{b}^\mu = 0\,, \tag{4.7}$$

where $M \equiv m - (b\cdot v)$.

Except for some particular cases, Eq. (4.7) yields us a non-numerable set of *new equations* (in principle, b^μ is arbitrary). Let us show some particular cases of Eq. (4.7) that have been considered before.

Choosing $b^\mu = ka^\mu$, where k is a constant, Eqs. (4.6) and (4.7) yield

$$\dot{S}^{\mu\nu} = k\,[v^\mu, a^\nu]\,, \tag{4.8}$$

$$\frac{d(mv^\mu)}{d\tau} - \frac{2}{3}e^2(\dot{a}^\mu - a^2v^\mu) - e\,F^{\mu\nu}_{\text{ext}}v_\nu - k\dot{a}^\mu = 0\,. \tag{4.9}$$

For $k = -\frac{2}{3}e^2$, Eq. (4.9) reduces to Bonnor equation.[2,8] In this case, the change in the spin part of the radiated field is supplied by the change in the intrinsic mechanical angular momentum (spin) of the particle.[8] Substituting this value of S in Eqs. (3.24), (3.25), (3.6), and (3.7), we obtain a phenomenology consistent with Bonnor's equation. In this way, the consistency of this equation with the conservation laws is rigorously proven (in contraposition with the point made in Ref. 4, Bonnor's equation is consistent with angular momentum conservation[8]).

For $k=0$, Eq. (4.5) leads to (4.2).

For $k\neq 0$, $-\frac{2}{3}e^2$, a family of equations discussed before[24] [see, e.g., Eq. (6) of Ref. 7 and the consequences of it, under the change $\epsilon\to -3k\epsilon/2e^2$] is obtained.

Choosing now,

$$\begin{aligned}\dot b^{\mu} &= [\tfrac{4}{3}ke^2+2\dot I\,](\dot a^{\mu}-a^2v^{\mu})+I\,[\ddot a^{\mu}+(\ddot a\cdot v)v^{\mu}]\\ &\quad -d\,[I(\ddot a\cdot v)+2\dot I(\dot a\cdot v)+\ddot I\,]v^{\mu},\end{aligned}$$

where, I, k, and d are any functions of τ, differentiable enough. Then, Eq. (4.7) reduces to

$$\begin{aligned}&\frac{d(Mv^{\mu})}{d\tau}-\left[\frac{2}{3}e^2(2k+1)+2\dot I\right](\dot a^{\mu}-a^2v^{\mu})-eF^{\mu\nu}_{\rm ext}v_{\nu}\\ &\quad -I\,[\ddot a^{\mu}+(\ddot a\cdot v)v^{\mu}]\\ &\quad +d\,[I(\ddot a\cdot v)+2\dot I(\dot a\cdot v)+\ddot I\,]v^{\mu}=0\,.\end{aligned}\tag{4.10}$$

For $d=1$, Eq. (4.10) reduces to an equation obtained by Honig–Szamosi.[4] This equation is general but does not include many equations (such as Bonnor's equation). There exist some particular cases of Honig–Szamosi's equation that are interesting to mention. For $I=0$ and $k=-\frac{1}{2}$, there is no radiation effect. For $I=0$ and $k<-\frac{1}{2}$, this equation is "similar" to the Lorentz–Dirac equation, but, by contrast with the runaway solutions without external forces, the radiation damps the movement of the charge independently of the initial conditions.

For $k=d=0$ and $\dot I=0$, Eq. (4.10) leads to an equation proposed by Ringermacher[3] with M constant. Let us point out that the mass definitions of Refs. 3 and 4 *do not* coincide between them and, in general, they *cannot* be interpreted in the light of Eq. (3.30).

Let us notice, that giving conditions to $(\dot S)_{\parallel}$ in terms of kinematical variables is not the more general situation, since in this case Eqs. (3.26) and (3.27) are uncoupled.

Restrictions of other kind can be given, as it is shown in the following example.

(iii) Let us impose that

$$S^{\mu\nu}v_{\nu}=Ia^{\mu}\,,\tag{4.11}$$

where, I is any function of τ, differentiable enough. This condition has been used in theories of particles with spin (see, e.g., Ref. 25 and the works cited therein). Equations (3.26), (3.27), and (4.11) lead to a system of coupled differential equations. This system of equations has not been considered in the literature. A subcase of this system is $\dot q^{\mu\nu}_{A}\equiv 0$ [see, Eqs. (3.20) and (3.24)]. For this subcase the system of equations leads to Honig–Szamosi's equation,[4] with $k=0$.

For the particular case $I\equiv 0$, condition (4.11) is the most commmonly used in theories of particles with spin (see, e.g., Refs. 23 and 25). In this case, the system of coupled differential equations represented by Eqs. (3.26), (3.27), and (4.11) (with $I\equiv 0$) leads (as it is easy to show) to a conserved magnitude of the spin ($S^{\mu\nu}S_{\mu\nu}\equiv S^2$) and to a conserved rest mass m. The system of equations is a particular case of a set of equations obtained by Bhabha and Corben[26] as the equations describing the interaction between Maxwell's field and a point charge dipole. According to Bhabha and Corben[26] this system of equations [Eqs. (3.26), (3.27), and (4.11) with $I\equiv 0$] could explain the scattering of light by electrons for certain frequencies. This system of equations has not been studied in detail in the literature. If the radiation reaction terms are neglected in (3.27) (that is, terms proportional to e^2), a system of equations that was studied in detail by Corben[27] is obtained.

Finally, let us discuss the charge stability problem. In a functional approach, for $F^{\mu\nu}_{\rm ext}=0$, the condition of perfect stability[28] requires the vanishing of all the components of the stress tensor in the particle rest frame. Since outside the CWL the theory is purely electromagnetic, perfect stability cannot be satisfied without changing the theory there (cf. Ref. 28). This change is beyond the scope of the present paper. However, because in $\Omega(\epsilon)$ the trace of θ is zero, in a distribution approach a necessary condition for perfect stability is given by

$$(g_{\mu\nu}\theta^{\mu\nu},\phi)=-\int_{\tau_1}^{\tau_2}m\phi[z(\tau)]d\tau,\quad \text{for } \phi\in\mathscr{D}(\Omega)\,.\tag{4.12}$$

From (4.12) and the lemma of the Appendix, it follows that

$$a_{\mu}S^{\mu\nu}v_{\nu}=0\,.\tag{4.13}$$

In this manner, for $F^{\mu\nu}_{\rm ext}=0$, Eq. (4.13) represents a stability condition for a point charge that is satisfied by the requirement $a^{\mu}(\tau)=0$. This physical requirement does not impose strong restrictions on the possible equations of motion since we still have freedom to impose Cauchy data or boundary conditions in a suitable fashion.

If condition (4.12) is required for $F^{\mu\nu}_{\rm ext}\neq 0$, Eq. (4.13) follows again straightforwardly. In this case, many equations shown in this section do not satisfy (4.13); namely, Bonnor's equation is incompatible with (4.13) for $F^{\mu\nu}_{\rm ext}\neq 0$. Nevertheless, the fact that an equation does not satisfy the condition (4.13) for $F^{\mu\nu}_{\rm ext}\neq 0$ does not mean that such an equation should be rejected. Instead, it means that such an equation does not describe a fundamental or stable particle but it may be applied to composite unstable particles up to a certain stage of the theory (see, e.g., Ref. 2). Examples of equations satisfying Eq. (4.13) for $F^{\mu\nu}_{\rm ext}\neq 0$ are the Lorentz–Dirac equation (for $S^{\mu\nu}=0$) and the general equation with the condition $S^{\mu\nu}v_{\nu}=0$. In this sense, these last equations are strong candidates to be classical relativistic limits of quantum relativistic theories.

V. DISCUSSION

We have seen that, imposing energy-momentum and angular momentum conservations, *the electromagnetism of a point charge is compatible with many phenomenologies*, and hence, *many equations of motion for a point charge are compatible with classical electrodynamics*. The reason for this is clear, since there are many different point charges representing different physical characteristics and in accordance with these characteristics there are different equations of motion. In principle, from the point of view of Poincaré invariance, these different point charges can be realized (classically) in nature. An important property of the classical electrodynamics is that the conservation laws can be satisfied with a finite number of phenomenological parameters.

It is important to observe that the lack of completeness of the electromagnetic theory and the Poincaré invariance to yield a description for a point charge, *is not* an exclusive property of this model. In fact, from Poincaré–Von Laue's analysis, it follows that any closed theory of extended sources and electromagnetic fields *is phenomenological.* This is quite clear, if it is remembered that all the equations that determine the problem uniquely must be given (e.g., equations of state) and that the structure of the sources lies outside the scope of Maxwell's equations. This situation cannot yet be avoided in simple models of matter interacting with an electromagnetic field (such as a point charge model).

Other examples of equations of motion included in the general form (3.26) and (3.27) are still possible. However, there is not fundamental reason to show particular equations of motion without pointing out specific charge models. The purpose of Sec. IV was to show the compatibility of the general equation with previous findings. The fact that there are different point models of macroscopic or microscopic particles means that a variety of equations of motion may exist *simultaneously.* For instance: Bhabha and Corben[26] considered Eqs. (3.26), (3.27), and (4.11) with $I\equiv 0$ as describing, classically, electrons. The Lorentz–Dirac equation becomes necessarily the equation of motion of a spinless particle. Bonnor[2] has emphasized that his equation of motion is supposed to refer to macroscopic charged particles and not to fundamental ones of constant mass such as the electron. All these proposals represent *different physical situations* which could be eventually realized.

If we compare the present treatment to obtain an equation of motion with other approaches [see, e.g., (i) Sec. I], it can be seen that the distribution approach yields a clearer insight into the problem because quantities appearing in a distribution formulation are feasible of physical interpretation, while methods such as special ways to reach the CWL,[4–9,11–13] renormalization procedure, and cutoff prescriptions are not. In fact, ambiguities in these usual procedures are replaced, in a distribution theory, by ambiguities in an antisymmetric spin tensor and certain subsidiary conditions.

The distribution form developed in this article (which follows Rowe's main ideas[15,19]) has the advantage over earlier distribution approaches,[14,15] in that strong assumptions about the phenomenology are not imposed from the beginning. Apart from the general view, in the case of a particular equation, the additional necessary requirements to derive it can be obtained. This form of obtaining particular equations is advantageous for itself, because it permits a better comprehension of each equation. An explicit example of this was given in Sec. IV through the generalization of the Lorentz–Dirac equation to spinning particles.

An unique equation of motion could be justified, from a theoretical point of view, through the classical limit of quantum electrodynamics which is up to date an open question (see, e.g., Ref. 29 and the works cited therein). However, since the electromagnetic field can be coupled with different spinor or tensor fields, it is reasonable to expect that different quantum theories lead to different classical relativistic limits. Of course, whichever is that classical limit, it should be included in the general equation given by (3.26) and (3.27).

With the definition of "renormalization" adopted in this paper, there are an infinity of phenomenological parameters that can be considered. However, one could argue that as long as $\phi\in\mathscr{D}(\Omega)$ and $\int_\Omega \theta_{ni}\phi\, d^4x<\infty$, θ_{ni} is a well-defined functional, so θ_{ni} should be extended from this subspace to $\mathscr{D}(\Omega)$. If this requirement is adopted, in regards to the conservation laws, it does not change our results because we still have freedom on the K definition. Nevertheless, this requirement permits us to determine the least upper bound $N^{(0)}$ of the index N' in (2.8). In fact, it is not difficult to see with heuristic arguments that $N^{(0)}=1$ should be the solution. If $N=N^{(0)}$ [see, Eq. (2.10)] is taken as the definition of the theory "as electromagnetic as possible,"[9] which is natural, it follows that the approach in this paper is "as electromagnetic as possible."

ACKNOWLEDGMENTS

The author is very much indebted to Dr. P. L. Torres, who suggested this problem, for stimulating discussions and comments and for a critical reading of the manuscript. The author thanks also the journal referee who called his attention to the charge stability question.

APPENDIX: DIVERGENCE OF THE ENERGY-MOMENTUM TENSOR

In this Appendix, Rowe's[10,19] technique for calculating the divergence of the distribution θ is explained. Some steps that are not explicitly justified in Refs. 10 and 19, are worked out here. For this purpose let us introduce the following definition.

For $\phi(x)\in\mathscr{D}(\Omega)$, we define

$$\phi(\epsilon,\tau)\equiv\frac{1}{4\pi}\int_\Sigma d\Omega\,\phi(x)\,, \tag{A1}$$

where Σ is the bidimensional surface determined by the intersection of the future light cone with apex at $z(\tau)$ and the Bhabha tube of radius ϵ; $d\Omega$ is the solid-angle element for the inertial frame with time axis $v^\mu(\tau)$, see, e.g., (A16) in Ref. 13.

The calculation of the divergence of K is direct from its definition. We only consider the calculation of $(\partial\cdot\theta_{\rm reg},\phi)$, the others divergences can be calculated in a similar way. Then, from (2.5) we have that

$$\begin{aligned}-&(\theta^{\mu\nu}_{\rm reg},\partial_\nu\phi)\\ &=-\frac{e^2}{16\pi}\int_\Omega\frac{R^\mu R^\nu}{\rho^4}\partial^2\partial_\nu\phi\, d^4x\\ &\quad+\frac{e^2}{4\pi}\int_\Omega\frac{R^\mu R^\nu}{\rho^4}\left[a+\frac{(a\cdot R)}{\rho}v\right]\\ &\quad\cdot\partial\partial_\nu\phi d^4x,\quad \text{for } \phi\in\mathscr{D}(\Omega)\,.\end{aligned} \tag{A2}$$

By Lebesgue monotone convergence theorem:

$$\int_\Omega\frac{R^\mu R^\nu}{\rho^4}\partial^2\partial_\nu\phi\, d^4x=\lim_{\epsilon\to 0}\int_{\bar\Omega(\epsilon)}\frac{R^\mu R^\nu}{\rho^4}\partial^2\partial_\nu\phi d^4x\,. \tag{A3}$$

Integrating by parts, it follows that:

$$\int_{\bar{\Omega}(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}\partial_\nu \partial^2\phi\, d^4x = -\int_{B(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}\partial^2\phi\, dB_\nu\,, \tag{A4}$$

where $dB_\nu = \epsilon^2 d\tau_r d\Omega(u_\nu + (a\cdot u)R_\nu)$.

Let us rewrite the integral over the segment of the Bhabha tube:

$$\begin{aligned}\int_{B(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}\psi(x)dB_\nu &= \int_{B(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}\psi(\epsilon,\tau_r)dB_\nu \\ &\quad + \int_{B(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}[\psi(x)-\psi(\epsilon,\tau_r)]dB_\nu\,,\end{aligned} \tag{A5}$$

where $\psi(x) \equiv \partial^2\phi(x)$ and $\psi(\epsilon,\tau_r)$ is given by (A1).

Then, after some manipulations in (A5), we obtain

$$\begin{aligned}\int_{B(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}\psi(x)dB_\nu &= 4\pi\int_{\tau_1}^{\tau_2} v^\mu(\tau)\psi(\epsilon,\tau)d\tau \\ &\quad + \int_{B(\epsilon)} (u^\mu + v^\mu)[\psi(x)-\psi(\epsilon,\tau)]d\tau\, d\Omega\,.\end{aligned} \tag{A6}$$

Let us define

$$\Delta\psi(\epsilon,\tau) \equiv \int_\Sigma (u^\mu + v^\mu)[\psi(x)-\psi(\epsilon,\tau)]d\Omega\,. \tag{A7}$$

Then we have that

$$|\Delta\psi(\epsilon,\tau)| \leqslant \sup_{x\in\Sigma}|\psi(x)-\psi(\epsilon,\tau)|\int_\Sigma |(u^\mu + v^\mu)|d\Omega\,, \tag{A8}$$

therefore:

$$\lim_{\epsilon\to 0}|\Delta\psi(\epsilon,\tau)| = 0\,, \tag{A9}$$

since Σ is compact and $\psi(x)\in\mathscr{D}(\Omega)$. From (A6), (A7), and (A9) it follows that:

$$\lim_{\epsilon\to 0}\int_{B(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}\psi(x)dB_\nu = 4\pi\int_{\tau_1}^{\tau_2} d\tau\, v^\mu(\tau)\partial^2\phi[z(\tau)]\,. \tag{A10}$$

Similarly:

$$\begin{aligned}\lim_{\epsilon\to 0}\int_{B(\epsilon)} \frac{R^\mu R^\nu}{\rho^4}\left[a + \frac{(a\cdot R)}{\rho}v\right]\cdot\partial\phi\, dB_\nu &= 4\pi\int_{\tau_1}^{\tau_2} d\tau\{v^\mu(\tau)a\cdot\partial\phi[z(\tau)] \\ &\quad + \tfrac{1}{3}a^\mu(\tau)v\cdot\partial\phi[z(\tau)]\}\,.\end{aligned} \tag{A11}$$

From (A2)–(A4), (A10), and (A11), the following result is obtained[15]

$$\begin{aligned}(\partial_\nu\theta^{\mu\nu}_{\rm reg},\phi) &= \frac{e^2}{4}\int_{\tau_1}^{\tau_2} d\tau\, v^\mu(\tau)\partial^2\phi[z(\tau)] \\ &\quad - e^2\int_{\tau_1}^{\tau_2} d\tau\{v^\mu(\tau)a\cdot\partial\phi[z(\tau)] \\ &\quad + \tfrac{1}{3}a^\mu(\tau)v\cdot\partial\phi[z(\tau)]\}\,.\end{aligned} \tag{A12}$$

Finally, for the sake of completeness, let us include a lemma[30] that is an important tool for extracting the equation of motion.

Lemma (Tulczyjew[30]): Let $B^{\mu\nu}(\tau)$ and $C^{\mu\nu\lambda}(\tau)$ be such that

$$\begin{aligned}&B^{\mu\nu}(\tau)v_\nu(\tau) = 0;\quad C^{\mu\nu\lambda}(\tau) = C^{\mu\lambda\nu}(\tau);\\ &C^{\mu\nu\lambda}(\tau)v_\lambda(\tau) = 0\quad \text{for all } \tau\in(\tau_1,\tau_2)\,.\end{aligned} \tag{A13}$$

If

$$\begin{aligned}\int_{\tau_1}^{\tau_2} d\tau\{A^\mu(\tau)\phi[z(\tau)] + B^{\mu\nu}(\tau)\partial_\nu\phi[z(\tau)] \\ + C^{\mu\nu\lambda}(\tau)\partial_\nu\partial_\lambda\phi[z(\tau)]\} = 0\quad \text{for all } \phi\in\mathscr{D}(\Omega)\,,\end{aligned} \tag{A14}$$

then

$$\begin{aligned}&A^\mu(\tau) = 0,\quad B^{\mu\nu}(\tau) = 0 \text{ and } C^{\mu\nu\lambda}(\tau) = 0\\ &\quad\text{for all } \tau\in(\tau_1,\tau_2)\,.\end{aligned} \tag{A15}$$

[1]P. A. M. Dirac, Proc. R. Soc. London Ser. A **167**, 148 (1938).
[2]W. B. Bonnor, Proc. R. Soc. London Ser. A **337**, 591 (1974).
[3]H. I. Ringermacher, Phys. Lett. A **74**, 381 (1979).
[4]E. Honig and G. Szamosi, Phys. Lett. A **82**, 384 (1981); **93**, 319 (1983); have derived a general form of the equation of motion of a point charge.
[5]E. Marx, Int. J. Theor. Phys. **14**, 55 (1975).
[6]R. Tabensky, Phys. Rev. D **13**, 267 (1976).
[7]A. Lozada and P. L. Torres, Phys. Rev. D **35**, 2064 (1987).
[8]A. Lozada and P. L. Torres, J. Math. Phys. **30**, 345 (1989).
[9]C. Teitelboim, Phys. Rev. D **1**, 1572 (1970).
[10]E. G. P. Rowe, Phys. Rev. D **12**, 1576 (1975); **18**, 3639 (1978).
[11]C. A. López and D. Villarroel, Phys. Rev. D **11**, 2724 (1975).
[12]R. Tabensky and D. Villarroel, J. Math. Phys. **16**, 1380 (1975).
[13]C. Teitelboim, D. Villarroel and Ch. G. Van Weert, Riv. Nuovo Cimento **3**, 1 (1980).
[14]J. G. Taylor, Proc. Cambridge Philos. Soc. **52**, 119 (1956).
[15]E. G. P. Rowe, Nuovo Cimento B **73**, 226 (1983).
[16]H. Poincaré, Comptes Rend. **140**, 1504 (1905); Rend. Circ. Mat. Palermo **21**, 129 (1906).
[17]M. Von Laue, Ann. Phys. (Leipzig) **35**, 524 (1911).
[18]J. Schwinger, Found. Phys. **13**, 373 (1983).
[19]E. G. P. Rowe and G. T. Rowe, Phys. Rep. **149**, 287 (1987).
[20]K. Yosida, *Functional Analysis* (Springer, New York, 1978).
[21]V. S. Vladimirov, *Generalized Functions in Mathematical Physics* (Mir, Moscow, 1979).
[22]K. Hepp, Commun. Math. Phys. **2**, 301 (1966); this paper contains a rigorous development of the theory. Renormalization in this framework is an extension of certain continuous linear functionals.
[23]P. Nyborg, Nuovo Cimento **23**, 47 (1962).
[24]A. Lozada and P. L. Torres (unpublished).
[25]P. Nyborg, Nuovo Cimento **26**, 821 (1962).
[26]H. J. Bhabha and H. C. Corben, Proc. R. Soc. London Ser. A **178**, 273 (1941); see especially, Eqs. (101) and (103) with $k = 0, g_2 = 0$.
[27]H. C. Corben, Nuovo Cimento **20**, 529 (1961); Phys. Rev. **121**, 1833 (1961).
[28]I. Bialynicki-Birula, in *Nonlinear Problems in Theoretical Physics, Lecture Notes in Physics*, Vol. 98, edited by A. F. Rañada (Springer, Berlin, 1979), p. 16.
[29]H. Grotch, E. Kazes, F. Rohrlich, and D. H. Sharp, Acta Phys. Austriaca **54**, 31 (1982).
[30]W. Tulczyjew, Acta Phys. Pol. **18**, 393 (1959).

Null path model for classical interactions

Roderick I. Sutherland
Department of Applied Sciences, University of Western Sydney, Hawkesbury, Richmond, New South Wales, 2753 Australia

(Received 19 May 1988; accepted for publication 15 March 1989)

It is shown that certain simple field interactions which are known to be reexpressible in direct interaction form can also be given an "action at no distance" interpretation. The solutions necessary for this interpretation are obtained by generalizing the mode of interaction in space-time from null lines (straight) to null paths (with corners).

I. INTRODUCTION

From the investigations of Schwarzschild,[1] Tetrode,[2] Fokker,[3] and Wheeler and Feynman,[4] it is well known that classical electrodynamics can be understood in terms of direct interaction between pairs of particles, without the need for the additional concept of an independent mediating field. Such a formulation is reasonable because the interactions in electromagnetism are via null lines in space-time (i.e., lines along which the interval $ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$ is zero), so that we do not need the field picture to explain "action at a distance": we actually have "action at *no* distance" from a four-dimensional point of view. (The momentum and energy exchanged between world lines can then be thought of as occurring via direct contact.)

A direct interaction reformulation (i.e., one which does away with the field as an independent physical entity) is also possible for other field theories, in addition to electromagnetism.[5] However, such a reformulation is perhaps less appealing in the case of interactions that propagate through three-dimensional space at speeds other than that of light, since the action at no distance picture is apparently no longer available. (One is left with the feeling that some additional physical entity is needed to carry the interaction across the gap from one particle to another.) The aim of the present paper is to show that an action at no distance picture is possible in the general case of spatial propagation at any speed if one dispenses with the unnecessary restriction that the space-time paths along which the interaction occurs be straight (i.e., without corners). The discussion will be limited to the simple cases of scalar and vector fields. Also, we will restrict ourselves to classical field theories and to flat (i.e., Minkowskian) space-time.

II. ELECTRODYNAMICS

In this section, the relevant parts of the direct interaction formulation of electrodynamics will be summarized in order both to facilitate the subsequent extensions to other types of field and to establish the notation.

The field equation for the electromagnetic four-potential, $A^\mu(x)$, has the form

$$\frac{\partial}{\partial x^\nu}\left(\frac{\partial A^\mu}{\partial x_\nu} - \frac{\partial A^\nu}{\partial x_\mu}\right) = \frac{4\pi}{c} J^\mu , \qquad (1)$$

where $J^\mu(x)$ is the four-current density. (As usual, μ and ν take the values 0, 1, 2, 3; a summation over repeated greek indices is implied; x stands for x^0, x^1, x^2, x^3; and c is the speed of light.) If one imposes the Lorentz gauge condition

$$\frac{\partial A^\mu}{\partial x^\mu} = 0 , \qquad (2)$$

Eq. (1) reduces to

$$\frac{\partial^2 A^\mu}{\partial x_\nu \partial x^\nu} = \frac{4\pi}{c} J^\mu . \qquad (3)$$

We now wish to consider a system of charged point particles[6] and to determine the potential on the ith particle's world line arising from the charges on the other particles.[7] In this case, the form we require for the source current density in (3) is the delta function expression

$$J^\mu(\mathbf{r}_i,t) = \sum_{j\neq i} e_j \delta(\mathbf{r}_i - \mathbf{r}_j)\frac{dx_j^\mu}{dt} .$$

Here $\mathbf{r}_i \equiv (x_i^1, x_i^2, x_i^3)$ is the position of the ith particle at time $t = x^0/c$ and e_i is its charge. This expression for the current density can be written in the manifestly covariant form

$$J^\mu(x_i) = \sum_{j\neq i} e_j c \int \delta^4(x_i - x_j) v_j^\mu \, d\tau_j , \qquad (4)$$

where

$$\delta^4(x_i - x_j) \equiv \delta(x_i^0 - x_j^0)\delta(x_i^1 - x_j^1)\delta(x_i^2 - x_j^2)\delta(x_i^3 - x_j^3) ,$$

$$v_j^\mu \equiv \frac{dx_j^\mu}{d\tau_j} ,$$

and τ_j is the proper time along the jth particle's world line (the integral being from $-\infty$ to $+\infty$). With (4) inserted in (3), a solution for the four-potential is

$$A^\mu(x_i) = \sum_{j\neq i} e_j \int \delta(s_{ij}^2) v_j^\mu \, d\tau_j , \qquad (5)$$

where

$$s_{ij}^2 \equiv (x_{i\mu} - x_{j\mu})(x_i^\mu - x_j^\mu) \qquad (6)$$

(i.e., s_{ij} is the interval between the points x_i and x_j in space-time). This expression satisfies (3) on account of the identity

$$\frac{\partial^2}{\partial x_{i\nu}\, \partial x_i^\nu} \delta(s_{ij}^2) = 4\pi \delta^4(x_i - x_j) . \qquad (7)$$

Equation (5) expresses the potential at point x_i on the ith particle's world line as a sum of contributions from all the points x_j along every other particle's world line. Now, from the form of (6) we have the further identity

 0022-2488/89/081721-06$02.50

$$\frac{\partial}{\partial x_i^\mu}\delta(s_{ij}^2) = -\frac{\partial}{\partial x_j^\mu}\delta(s_{ij}^2), \tag{8}$$

from which it follows that the potential solution (5) satisfies the gauge condition (2):

$$\begin{aligned}\frac{\partial}{\partial x_i^\mu}A^\mu(x_i) &= \sum_{j\neq i} e_j \int \frac{\partial}{\partial x_i^\mu}\delta(s_{ij}^2)v_j^\mu\, d\tau_j \\ &= -\sum_{j\neq i} e_j \int v_j^\mu \frac{\partial}{\partial x_j^\mu}\delta(s_{ij}^2)\, d\tau_j \\ &= -\sum_{j\neq i} e_j \int \frac{d}{d\tau_j}\delta(s_{ij}^2)\, d\tau_j \\ &= -\sum_{j\neq i} e_j [\delta(s_{ij}^2)]_{\tau_j=-\infty}^{\tau_j=+\infty} \\ &= 0,\end{aligned}$$

for any finite x_i. Hence this solution of (3) is also a solution of Eq. (1).

In the usual formulation of electrodynamics, the basic equations are taken to be the field equation (1) [to which (5) is merely one of many solutions], together with the equations of motion of the charged particles:

$$\frac{d}{d\tau_i}(m_i v_{i\mu}) = \frac{e_i}{c}\left(\frac{\partial A_\nu}{\partial x_i^\mu} - \frac{\partial A_\mu}{\partial x_i^\nu}\right)v_i^\nu, \tag{9}$$

m_i being the mass of the ith particle. The direct interaction formulation, on the other hand, is based on the equations of motion (9) alone, in which the symbol A^μ stands for the quantity (5). The field equation is then viewed as merely an identity that A^μ satisfies. In this alternative picture, the field is no more than a mathematically convenient way of describing interactions between world lines: fields do not exist free of particles and other solutions of the field equation are not physically permitted. In particular, the homogeneous equation

$$\frac{\partial}{\partial x^\nu}\left(\frac{\partial A^\mu}{\partial x_\nu} - \frac{\partial A^\nu}{\partial x_\mu}\right) = 0$$

(which in the usual formulation describes field propagation in free space) now has no meaning and so arbitrary solutions of it cannot be added to the expression (5) for A^μ. Hence one advantage of this formulation is that it does away with unnecessary and unwanted solutions.

The direct interaction formulation can be conveniently summarized by the action function

$$\begin{aligned}S = &-\sum_i m_i c\int\left(\frac{dx_{i\mu}}{d\tau_i}\frac{dx_i^\mu}{d\tau_i}\right)^{1/2} d\tau_i \\ &-\sum_{i<j}\sum \frac{e_i e_j}{c}\iint \delta(s_{ij}^2) \\ &\times\frac{dx_{i\mu}}{d\tau_i}\frac{dx_j^\mu}{d\tau_j}\, d\tau_i\, d\tau_j .\end{aligned}$$

Variation of the ith particle's world line then yields the ith equation of motion.

In constructing a solution for A^μ, such as the one discussed above, it is useful to proceed via an appropriate Green's function (as will be done for the fields considered in Secs. IV and V). For the electromagnetic case, one has the Green's function equation

$$\frac{\partial^2}{\partial x_{i\nu}\,\partial x_i^\nu}G(x_i,x) = \delta^4(x_i - x). \tag{10}$$

The point of this equation is, of course, that multiplying it on both sides by $(4\pi/c)J^\mu(x)$ and integrating over space-time gives

$$\frac{\partial^2}{\partial x_{i\nu}\,\partial x_i^\nu}\int G(x_i,x)\frac{4\pi}{c}J^\mu(x)d^4x = \frac{4\pi}{c}J^\mu(x_i),$$

which corresponds in form to (3), so that each solution $G(x_i,x)$ obtained for Eq. (10) yields a four-potential solution

$$A^\mu(x_i) = \int G(x_i,x)\frac{4\pi}{c}J^\mu(x)\, d^4x \tag{11}$$

to (3). Now, to find a solution suitable as a basis for a direct interaction formulation, we simply substitute the point particle current density

$$J^\mu(x) = \sum_{j\neq i} e_j c\int \delta^4(x - x_j)v_j^\mu\, d\tau_j$$

into (11), thus obtaining the desired class of solution:

$$A^\mu(x_i) = \sum_{j\neq i} 4\pi e_j \int G(x_i,x_j)v_j^\mu\, d\tau_j .$$

Hence the problem of deducing an appropriate solution for A^μ reduces to obtaining an appropriate Green's function solution.

Three important solutions of (10) are the so-called retarded, advanced, and symmetric Green's functions

$$\begin{aligned}G_{\rm ret} &= (4\pi|\mathbf{r}_i - \mathbf{r}_j|)^{-1}\delta(|\mathbf{r}_i - \mathbf{r}_j| - c[t_i - t_j]), \\ G_{\rm adv} &= (4\pi|\mathbf{r}_i - \mathbf{r}_j|)^{-1}\delta(|\mathbf{r}_i - \mathbf{r}_j| + c[t_i - t_j]), \\ G_{\rm sym} &= (4\pi)^{-1}\delta(s_{ij}^2),\end{aligned} \tag{12}$$

which are related via

$$G_{\rm sym} = \tfrac{1}{2}G_{\rm ret} + \tfrac{1}{2}G_{\rm adv},$$

and

$$G_{\rm ret}(x_i,x_j) = G_{\rm adv}(x_j,x_i). \tag{13}$$

Now, it is easily shown that equality of action and reaction for two interacting particles i and j (and thereby conservation of momentum and energy) is ensured if we impose the symmetry requirement

$$G(x_i,x_j) = G(x_j,x_i). \tag{14}$$

Of the three possibilities in (12), only $G_{\rm sym}$ possesses this property, which is why it is the one adopted in the direct interaction formulation.[8]

The choice of $G_{\rm sym}$ necessarily entails the existence of backward-in-time effects, which raises the question of causal loop paradoxes. A general "continuity" argument for resolving such paradoxes was given by Wheeler and Feynman[4] (see also Schulman[9]).

We end this section by reiterating that the above formulation [as embodied in the potential expression (5)] achieves *two* aims. In addition to providing us with a direct interaction interpretation for electrodynamics (i.e., one which eliminates the need for the field as an additional physical entity), it also provides an action at no distance interpretation, since the delta function $\delta(s_{ij}^2)$ indicates that the direct interaction occurs only along null lines in space-time (i.e., lines of zero four-dimensional length).

III. NULL PATH INTERACTIONS

Having looked at the electromagnetic case, our aim is now to produce a similar reformulation for interparticle actions that are not restricted to the light cone. Again, mediation via fields will be replaced by direct interaction via paths of zero length in space-time. To this end, we introduce the idea of a "null path," this being a path in space-time made up of a succession of null lines. Three such paths between a pair of points x_i and x_j are shown in Fig. 1. These paths can be thought of as having zero length in the sense that the interval along each separate straight segment is 0.

Just as a single straight line in the electromagnetic case is described mathematically by a delta function $\delta(s_{ij}^2)$, a path containing corners will be described by a product of delta functions. For example, the middle path in Fig. 1 will correspond to a term of the form

$$\delta(s_{i1}^2)\delta(s_{12}^2)\delta(s_{23}^2)\delta(s_{34}^2)\delta(s_{4j}^2)\,.$$

The total interaction between a pair of points is then given by the sum over all possible null paths connecting them (this sum being over all different numbers of corners and all possible positions for these corners).

IV. SCALAR INTERACTIONS

We will now apply the null path idea to the case of scalar interactions. The field equation for a scalar potential $\phi(x)$ is

$$\frac{\partial^2\phi}{\partial x_\nu\,\partial x^\nu} + \alpha^2\phi = \frac{4\pi}{c}\rho_0\,, \tag{15}$$

where $\rho_0(x)$ is the rest density of the "charge" which gives rise to the interaction. The presence of the mass term α^2 means that the interaction is not restricted to the light cone. Equation (15) has solutions

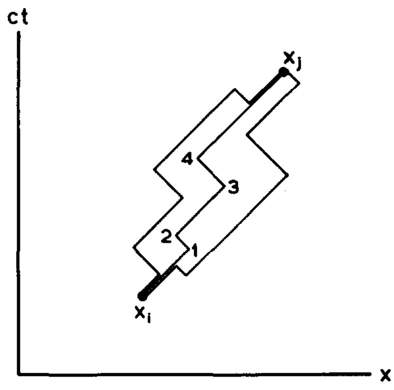


FIG. 1. Examples of null paths connecting the space-time points x_i and x_j.

$$\phi(x_i) = \int G(x_i,x)\,\frac{4\pi}{c}\rho_0(x)d^4x\,, \tag{16}$$

where the Green's function is a solution of the equation

$$\left(\frac{\partial^2}{\partial x_{i\nu}\,\partial x_i^\nu} + \alpha^2\right)G\,(x_i,x_j) = \delta^4(x_i - x_j)\,. \tag{17}$$

For a system of point particles, ρ_0 takes the form

$$\rho_0(x) = \sum_{j\neq i} e_j c \int \delta^4(x - x_j)d\tau_j\,,$$

e_j being the scalar interaction charge on the jth particle. This expression reduces (16) to

$$\phi(x_i) = \sum_{j\neq i} 4\pi e_j \int G(x_i,x_j)d\tau_j. \tag{18}$$

The equations of motion for particles in a scalar field can be found from the action

$$S = -\sum_i m_i c \int \left(\frac{dx_{i\mu}}{d\tau_i}\frac{dx_i^\mu}{d\tau_i}\right)^{1/2} d\tau_i - \sum_i e_i \int \left(\frac{dx_{i\mu}}{d\tau_i}\frac{dx_i^\mu}{d\tau_i}\right)^{1/2} \phi(x_i)\,d\tau_i\,.$$

Variation of the ith world line yields[10]

$$\frac{d}{d\tau_i}\left[\left(m_i + \frac{e_i}{c}\phi\right)v_{i\mu}\right] = e_i c\,\frac{\partial\phi}{\partial x_i^\mu}. \tag{19}$$

Equations (19), together with the expression (18), provide a direct interaction formulation for scalar interactions once an appropriate solution for $G(x_i,x_j)$ is found.

To obtain a solution that also allows a null path interpretation, we will look for a Green's function of the form

$$\begin{aligned} G(x_i,x_j) &= k_0\delta(s_{ij}^2) \\ &\quad + k_1 \int \delta(s_{i1}^2)\delta(s_{1j}^2)\,d^4x_1 \\ &\quad + k_2 \iint \delta(s_{i1}^2)\delta(s_{12}^2)\delta(s_{2j}^2)d^4x_1\,d^4x_2 \\ &\quad + \cdots, \end{aligned} \tag{20}$$

where the constants k are to be determined. These successive terms correspond to paths having zero corners, one corner, two corners, etc. It is easily seen that expression (20) satisfies the symmetry requirement (14).

To simplify the reasoning below, it is more convenient to rewrite (20) in the equivalent form

$$\begin{aligned} G(x_i,x_j) &= k_0\delta(s_{ij}^2) \\ &\quad + k_1 \int \delta(s_{j1}^2)\delta(s_{1i}^2)d^4x_1 \\ &\quad + k_2 \iint \delta(s_{j1}^2)\delta(s_{12}^2)\delta(s_{2i}^2)d^4x_1\,d^4x_2 \\ &\quad + \cdots. \end{aligned} \tag{21}$$

Now, the identity (7) allows us to deduce the following result for the Green's function in (21):

$$\frac{\partial^2}{\partial x_{i\nu}\,\partial x_i^\nu} G(x_i,x_j)$$
$$= k_0 4\pi\delta^4(x_i - x_j) + k_1 \int \delta(s_{j1}^2) 4\pi\delta^4(x_1 - x_i) d^4x_1 + k_2 \iint \delta(s_{j1}^2)\delta(s_{12}^2) 4\pi\delta^4(x_2 - x_i) d^4x_1\, d^4x_2 + \cdots$$
$$= 4\pi k_0 \delta^4(x_i - x_j) + 4\pi k_1 \delta(s_{ij}^2) + 4\pi k_2 \int \delta(s_{j1}^2)\delta(s_{1i}^2) d^4x_1 + \cdots . \tag{22}$$

Also, from (17) we require this expression to be equal to

$$\delta^4(x_i - x_j) - \alpha^2 G(x_i,x_j) = \delta^4(x_i - x_j) - \alpha^2 k_0 \delta(s_{ij}^2) - \alpha^2 k_1 \int \delta(s_{j1}^2)\delta(s_{1i}^2) d^4x_1 - \cdots . \tag{23}$$

Comparing (22) and (23), we see that equality is obtained if we put

$$k_0 = 1/4\pi,$$
$$k_1 = -\alpha^2 k_0/4\pi = (1/4\pi)(-\alpha^2/4\pi),$$
$$k_2 = -\alpha^2 k_1/4\pi = (1/4\pi)(-\alpha^2/4\pi)^2,$$

etc., i.e.,

$$k_n = (1/4\pi)(-\alpha^2/4\pi)^n .$$

Hence, substituting these values into (20), we obtain the Green's function we are seeking:

$$G(x_i,x_j) = \frac{1}{4\pi}\left[\delta(s_{ij}^2) + \sum_{n=1}^{\infty}\left(-\frac{\alpha^2}{4\pi}\right)^n \iint \cdots \int \delta(s_{i1}^2)\delta(s_{12}^2) \cdots \delta(s_{nj}^2) d^4x_1\, d^4x_2 \cdots d^4x_n\right]. \tag{24}$$

By choosing this particular Green's function in our direct interaction equations (18) and (19), the resulting theory will have the desired "action at no distance" structure.

Having achieved our aim, we can also specify this formulation by the action function

$$S = -\sum_i m_i c \int \left(\frac{dx_{i\mu}}{d\tau_i}\frac{dx_i^\mu}{d\tau_i}\right)^{1/2} d\tau_i - \sum_{i<j}\sum \frac{4\pi e_i e_j}{c} \iint G(x_i,x_j)\left(\frac{dx_{i\mu}}{d\tau_i}\frac{dx_i^\mu}{d\tau_i}\right)^{1/2} \times \left(\frac{dx_{j\nu}}{d\tau_j}\frac{dx_j^\nu}{d\tau_j}\right)^{1/2} d\tau_i\, d\tau_j ,$$

in which $G(x_i,x_j)$ is defined by (24) (the equations of motion then being obtainable by variation of each particle's world line).

V. VECTOR INTERACTIONS

We will now return to vector interactions and consider a null path interpretation for the more general case where the interaction is not restricted to the light cone.

The generalized field equation is

$$\frac{\partial}{\partial x^\nu}\left(\frac{\partial A^\mu}{\partial x_\nu} - \frac{\partial A^\nu}{\partial x_\mu}\right) + \alpha^2 A^\mu = \frac{4\pi}{c} J^\mu . \tag{25}$$

Taking $\partial/\partial x^\mu$ of both sides of (25), and using the continuity equation

$$\frac{\partial J^\mu}{\partial x^\mu} = 0,$$

we obtain

$$\frac{\partial A^\mu}{\partial x^\mu} = 0 . \tag{26}$$

Substituting this restriction into (25) yields the reduced field equation

$$\frac{\partial^2}{\partial x_\nu\,\partial x^\nu} A^\mu + \alpha^2 A^\mu = \frac{4\pi}{c} J^\mu . \tag{27}$$

[Any solution of (25) automatically satisfies both (26) and (27). However, a solution of (27) is not necessarily a solution of (25) or (26).]

For a system of point particles, Eq. (27) has solutions

$$A^\mu(x_i) = \sum_{j\neq i} 4\pi e_j \int G(x_i,x_j) v_j^\mu\, d\tau_j , \tag{28}$$

where $G(x_i,x_j)$ is a solution of

$$\left(\frac{\partial^2}{\partial x_{i\nu}\,\partial x_i^\nu} + \alpha^2\right) G(x_i,x_j) = \delta^4(x_i - x_j) . \tag{29}$$

The equations of motion for the particles are the same as in the electromagnetic case:

$$\frac{d}{d\tau_i}(m_i v_{i\mu}) = \frac{e_i}{c}\left(\frac{\partial A_\nu}{\partial x_i^\mu} - \frac{\partial A_\mu}{\partial x_i^\nu}\right) v_i^\nu . \tag{30}$$

Now, (29) is the same Green's function equation as for the scalar field in the previous section, so the same Green's function derived there will also yield a null path solution for *vector* interactions. Thus, combining (24) and (28), our null path vector potential is

$$A^\mu(x_i) = \sum_{j\neq i} e_j \int \delta(s_{ij}^2) v_j^\mu\, d\tau_j + \sum_{j\neq i}\sum_{n=i}^{\infty} e_j\left(-\frac{\alpha^2}{4\pi}\right)^n \iint \cdots \int \delta(s_{i1}^2)\delta(s_{12}^2) \cdots \delta(s_{nj}^2) v_j^\mu\, d\tau_j\, d^4x_1\, d^4x_2 \cdots d^4x_n . \tag{31}$$

This potential satisfies the reduced field equation (27). However, in order to establish that it is also a solution of the field equation (25), it is necessary to prove that it conforms to the gauge condition (26). The fact that the first term of (31) has zero four-divergence, in accordance with (26), has been shown in Sec. II. It is now a simple matter to demonstrate that the remaining terms in (31) also have vanishing four-divergence, i.e., to show that the expression

$$\frac{\partial}{\partial x_i^\mu}\int\int\cdots\int \delta(s_{i1}^2)\delta(s_{12}^2)\cdots\delta(s_{nj}^2)v_j^\mu\, d\tau_j\, d^4x_1\, d^4x_2\cdots d^4x_n$$

is zero. Using (8), this expression becomes

$$-\int\int\cdots\int \frac{\partial\delta(s_{i1}^2)}{\partial x_1^\mu}\,\delta(s_{12}^2)\cdots\delta(s_{nj}^2)v_j^\mu\, d\tau_j\, d^4x_1\, d^4x_2\cdots d^4x_n\,,$$

and integrating by parts then yields

$$\int\int\cdots\int \delta(s_{i1}^2)\,\frac{\partial\delta(s_{12}^2)}{\partial x_1^\mu}\cdots\delta(s_{nj}^2)v_j^\mu\, d\tau_j\, d^4x_1\, d^4x_2\cdots d^4x_n\,.$$

Thus, alternately using (8) and then integrating by parts, we eventually obtain

$$-\int\int\cdots\int \delta(s_{i1}^2)\delta(s_{12}^2)\cdots\frac{\partial\delta(s_{nj}^2)}{\partial x_j^\mu}\,v_j^\mu d\tau_j\, d^4x_1\, d^4x_2\cdots d^4x_n\,,$$

and referring again to the proof in Sec. II, this expression is seen to be zero. We have therefore established that our proposed null path potential is, in fact, a solution of the vector field equation (25).

In conclusion, our null path direct interaction formulation for vector fields is given by the equations of motion (30), with A^μ defined by (31). As usual, the formulation can alternatively be expressed in terms of an action function:

$$S=-\sum_i m_i c\int\left(\frac{dx_{i\mu}}{d\tau_i}\frac{dx_i^\mu}{d\tau_i}\right)^{1/2}d\tau_i-\sum_{i<j}\sum\frac{4\pi e_i e_j}{c}\int\int G(x_i,x_j)\left(\frac{dx_{i\mu}}{d\tau_i}\frac{dx_j^\mu}{d\tau_j}\right)d\tau_i\, d\tau_j\,,$$

where $G(x_i,x_j)$ is given by (24).

VI. OTHER SUITABLE GREEN'S FUNCTIONS

As a final point, it should be noted that more than one null path Green's function is available for the interactions in Secs. IV and V. For comparison with the expression given below, we will rewrite the Green's function we have been considering in the form

$$G(x_i,x_j)=G_{\rm sym}(ij)+\sum_{n=1}^{\infty}(-\alpha^2)^n\int\int\cdots\int G_{\rm sym}(i1)G_{\rm sym}(12)\cdots G_{\rm sym}(nj)d^4x_1\, d^4x_2\cdots d^4x_n\,, \tag{32}$$

where $G_{\rm sym}(ij)$ is the symmetric electromagnetic Green's function defined in (12). Now, using the retarded and advanced electromagnetic Green's functions $G_{\rm ret}$ and $G_{\rm adv}$, another solution of (17) that satisfies our requirements is

$$G(x_i,x_j)=G_{\rm sym}(ij)+\sum_{n=1}^{\infty}(-\alpha^2)^n\int\int\cdots\int\frac{1}{2}[G_{\rm ret}(i1)G_{\rm ret}(12)\cdots G_{\rm ret}(nj)+G_{\rm adv}(i1)G_{\rm adv}(12)\cdots G_{\rm adv}(nj)]\,d^4x_1\, d^4x_2\cdots d^4x_n\,, \tag{33}$$

the delta functions in $G_{\rm ret}$ and $G_{\rm adv}$ providing the desired null path structure. This solution satisfies the symmetry requirement (14) because of the identity (13).

Expression (33) differs from (32) in being nonzero only for points x_i, x_j that are timelike separated. There does not seem to be any simple way of formulating an analogous null path solution that is purely spacelike.

ACKNOWLEDGMENT

The author wishes to thank John R. Shepanski for many helpful discussions.

[1]K. Schwarzschild, Nachr. Ges. Wiss. Gottingen Math. Phys. Kl. Fachgruppe 2 **128**, 132 (1903).

[2]H. Tetrode, Z. Phys. **10**, 317 (1922).

[3]A. D. Fokker, Z. Phys. **58**, 386 (1929); Physica **9**, 33 (1929); **12**, 145 (1932).

[4]J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **21**, 425 (1949).

[5]For example, H. Van Dam and E. P. Wigner, Phys. Rev. **138** B1576 (1965) and **142**, 838 (1966) discuss more general vector interactions; A. Katz, J. Math. Phys. **10**, 1929 (1969) considers the scalar interaction case; G. J. H. Burgers and H. Van Dam, J. Math. Phys. **28**, 677 (1987) examine a certain class of vector and tensor interactions; and J. V. Narlikar, Proc. Cambridge Philos. Soc. **64**, 1071 (1968) considers the generalization to arbitrary tensor and spinor interactions.

[6]In writing down Eqs. (1) and (3), J^μ is, of course, thought of as representing a continuous distribution of charge. Ultimately, however, any classical charge distribution is made up of discrete particles.

[7]To avoid particles having infinite self-energy, it is usually assumed in a direct interaction formulation that the potential acting on a particular particle is due only to particles other than itself. For this reason, the case $i=j$ is excluded from all our summations.

[8]The direct interaction formulation is often considered in conjunction with

the separate notion of absorber theory [J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **17**, 157 (1945)], in which it is postulated that the universe acts as a complete absorber of radiation. This idea provides an explanation for why only retarded fields are observed in nature as well as predicting the correct expression for the radiative reaction force. We will not be concerned with the absorber proposal here.

[9] L. S. Schulman, Am. J. Phys. **39**, 481 (1971).

[10] Here, $m + (e/c)\,\phi$ may be regarded as the "generalized rest mass" [analogous to the generalized 4-momentum $mv^\mu + (e/c)A^\mu$ for vector interactions].

Monopoles and instantons from Berry's phase

M. G. Benedict
Department of Theoretical Physics, JATE, P.O. Box 428, Szeged, H-6701, Hungary

L. Gy. Fehér[a)]
Bolyai Institute, JATE, P.O. Box 656, Szeged, H-6701, Hungary

Z. Horváth
Institute for Theoretical Physics, Roland Eötvös University, Budapest, H-1088, Hungary

(Received 20 October 1988; accepted for publication 22 March 1989)

The topological phase picked up by an arbitrary spin undergoing a cyclic series of quantum jumps is deduced by geometric considerations. Generalizing the Hamiltonian of a spin in a magnetic field, higher-dimensional Dirac monopoles are derived from the non-Abelian adiabatic phase. These monopoles appear in the momentum space description of zero mass particles in higher dimensions.

I. INTRODUCTION

Berry's phase[1,2] and its non-Abelian generalization, first studied by Wilczek and Zee,[3] allows for realizing and measuring interesting Yang–Mills fields in adiabatic quantum systems. An important example is a spin in a slowly varying external magnetic field $x = (x^i)$. The quantum adiabatic phase belonging to the Hamiltonian (proportional to the Hamiltonian of the spin),

$$H(x) = x^i\sigma_i \quad (i = 1,2,3)\,, \quad x\in\mathbb{R}^3\{0\}\,, \tag{1.1}$$

simulates the effect of a Dirac monopole of minimal magnetic charge. On the other hand, the Dirac monopole has interesting higher-dimensional generalizations,[4–6] which had been extensively studied in the 1970's, e.g., in connection with spontaneous compactification of extra dimensions.[6,7] The question arises: How could one obtain them from Berry's phase? To see this observe that there are two ways of looking at the Pauli matrices in (1). First, they are the spin-$\frac{1}{2}$ generators of SO(3). Replacing them by SO(3) generators in various other representations, one obtains Dirac monopoles of arbitrary charge. But the Pauli matrices also generate Cl(3), the Clifford algebra of the three-dimensional Euclidean space. From this viewpoint, a natural generalization of (1) is the Hamiltonian

$$H(x) = \Gamma(x) = x^i\Gamma_i\,,$$
$$i = 1,\ldots,2n+1\,, \quad (x^i)\in\mathbb{R}^{2n+1}\setminus\{0\}\,, \tag{1.2}$$

where Γ_i,

$$\{\Gamma_i,\Gamma_k\} = 2\delta^{ik}\,, \quad \Gamma_i{}^+ = \Gamma_i\,, \tag{1.3}$$

are the generators of Cl$(2n+1)$. We shall show that this Hamiltonian leads to the higher-dimensional Dirac monopoles.

In order to fix notations, first we recall the formalism of the quantum adiabatic phase. Consider a quantum system whose Hamiltonian $H(x)$ depends smoothly on some external parameter x varying in a manifold M. Let us suppose that for any $x\in$M, $H(x)$ has an isolated eigenvalue $E(x)$ giving rise to a smooth function on M. Denote V the state space of the quantum system and $\mathrm{V}(x)\subset\mathrm{V}$ the eigensubspace of energy $E(x)$. We assume that $N = \dim \mathrm{V}(x)$ does not depend on x, and that $V_E(\mathrm{M}) = \cup_x \mathrm{V}(x) \subset \mathrm{M}\times\mathrm{V}$ is a smooth vector bundle over M. Let us now cover M by contractible coordinate patches $(U_{(\alpha)}, x^i_{(\alpha)})$ ($\alpha\in I$, $i = 1,\ldots,\dim$ M), and on every such coordinate neighborhood choose a frame of eigenstates $|n_a^{(\alpha)}(x)\rangle$, fulfilling the relations

$$H(x)|n_a^{(\alpha)}(x)\rangle = E(x)|n_a^{(\alpha)}(x)\rangle\,,$$
$$\langle n_a^{(\alpha)}(x)|n_b^{(\alpha)}(x)\rangle = \delta_{ab}\,, \tag{1.4}$$
$$|n_b^{(\beta)}(x)\rangle = |n_a^{(\alpha)}(x)\rangle g_{ab}^{(\alpha\beta)}(x) \quad (a,b = 1,\ldots,N)\,.$$

Here $g^{(\alpha\beta)}: U_{(\alpha)}\cap U_{(\beta)} \to \mathrm{U}(N)$ (the group of $N\times N$ unitary matrices) are the transition functions of the bundle $V_E(M)$. Now consider an adiabatic process,[8] during which the parameter is varied slowly along a curve $\gamma = x(t)$, $0\leqslant t\leqslant T$. Let us choose $|\psi(0)\rangle\in\mathrm{V}(x(0))$ as the initial value for the time-dependent Schrödinger equation and, for simplicity, suppose that γ lies entirely in $U_{(\alpha)}$. The quantum adiabatic theorem states, that (for large times) $|\psi(t)\rangle\in\mathrm{V}(x(t))$ and thus the evolution $|\psi(0)\rangle\to|\psi(t)\rangle$ associates a unitary matrix $\mathrm{U}_{ab}^{(\alpha)}(x(t))$ to $x(t)$. This matrix describes the (approximate) evolution of the initial states $|n_b^{(\alpha)}(x(0))\rangle$,

$$|n_b^{(\alpha)}(x(0))\rangle \to |\psi_b^{(\alpha)}(x(t))\rangle = |n_a^{(\alpha)}(x(t))\rangle \mathrm{U}_{ab}^{(\alpha)}(x(t))\,. \tag{1.5}$$

The interesting part of $\mathrm{U}_{ab}^{(\alpha)}(x(t))$ is the "phase factor" defined by the Yang–Mills potential (a skew-Hermitian matrix in our convention),

$$A_{ab}^{(\alpha)}(x) = \left\langle n_a^{(\alpha)}(x)\left|\frac{\partial}{\partial x^i}\right|n_b^{(\alpha)}(x)\right\rangle dx^i\,. \tag{1.6}$$

That is, in the adiabatic limit,

$$\mathrm{U}^{(\alpha)}(x(t)) = \exp\left(-i\int_0^t E(x(\tau))\,d\tau\right)$$
$$\times\mathscr{P}\left[\exp\left(-\int_0^t A_i^{(\alpha)}(x(\tau))\frac{dx^i}{d\tau}\,d\tau\right)\right]. \tag{1.7}$$

[a)] Present address: Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland.

 0022-2488/89/081727-05$02.50

On $U_{(\alpha)} \cap U_{(\beta)}$, with $g = g^{(\beta\alpha)} = [g^{(\alpha\beta)}]^{-1}$, the relation

$$A^{(\beta)} = gA^{(\alpha)}g^{-1} + g\,dg^{-1} \tag{1.8}$$

holds. This ensures that the local representatives $A^{(\alpha)}$, $A^{(\beta)}$,... give us a connection[9] on the vector bundle $V_E(M)$. This connection is sometimes called the Berry connection. For a closed curve γ, the path ordered integral in (1.7) yields the Berry–Wilczek–Zee phase.[1-3]

In Sec. II we review the example of a spin in a slowly varying external magnetic field.[1,10] We shall pay particular attention to the explanation of the use of geodesic polygons in interpreting the experiments[11,12] based on cyclic series of quantum jumps. Section III contains our main result; we prove that the Hamiltonian (2) leads to the generalized Dirac monopoles (instantons when restricted to S^{2n}) as the source of the associated non-Abelian Berry's phase.

II. SPIN IN MAGNETIC FIELD

Here we study the Abelian Berry's phase generated by the Hamiltonian

$$H(x) = x^i J_i \quad (i = 1,2,3)\,, \quad (x^i) \in \mathbb{R}^3\{0\}\,, \tag{2.1}$$

where J_i $(i = 1,2,3)$ are the spin j generators of SO(3). We are interested in the adiabatic connection carried by the eigenstate bundle of energy $E(x) = rm$ for some $-j \leqslant m \leqslant j$, $r = |x|$. Let $\widetilde{U}_N(\widetilde{U}_S)$ be the contractible region obtained from $\mathbb{R}^3 \setminus \{0\}$ by removing the negative (positive) part of the x^3 axis. The eigenstate bundle under study is $V_m = \cup_x |j,m(x)\rangle$, where

$$(x^i/r)J_i|j,m(x)\rangle = m|j,m(x)\rangle\,, \quad r = |x|\,. \tag{2.2}$$

V_m allows for the following local sections

$$|j,m(x)\rangle_M = \{d^j_{l,m}(\theta)\exp[i(\lambda m - l)\varphi]\}^j_{l=-j} \tag{2.3}$$

on $\widetilde{U}_M$. Here $d^j_{l,m}(\theta) = \langle l,j|\exp(-i\theta J_2)|j,m\rangle$ are the Wigner functions,[13] $\lambda = +1, -1$ according to $M = N,S$, and we use spherical polar coordinates $\hat{x} = x/r(\sin\theta \times \cos\varphi, \sin\theta\sin\varphi, \cos\theta)$. On the intersection $\widetilde{U}_N \cap \widetilde{U}_S$, one has the relation

$$|j,m(x)\rangle_N = |j,m(x)\rangle_S \exp(i2m\varphi)\,. \tag{2.4}$$

It is well known that there is a unique spherically symmetric connection[14,15] on this bundle: the one describing a Dirac monopole of charge $2m$ (for $m = 1$ this is the Levi-Civita connection on the tangent bundle of S^2). The adiabatic gauge field we are after is clearly spherically symmetric and therefore it is the gauge field of a Dirac monopole of charge $2m$. In fact, the formulas (1.4)–(1.6) yield the standard form of the monopole potential

$$A_M = im(\lambda - \cos\theta)\,d\varphi\,. \tag{2.5}$$

This abstract monopole has already been observed[10,16] experimentally. The explanation of the optical fiber experiment of Tomita and Chiao[16] is based on the simple fact that the spin vector of a positive helicity photon is always parallel to its instantaneous wave vector $\hat{k}$ determined by the fiber. In other words, the varying direction of the optical fiber acts on the spin state of the photon like an adiabatically changing magnetic field. When transported along a closed curve C on S^2, the positive and negative helicity components of a linearly polarized photon state pick up opposite phases that are equal to the nonintegrable phase factors $\Phi_\pm$ $(\Phi_- = -\Phi_+)$ of a monopole of charge $m = \pm 1$, respectively. This implies[17] that the plane of linear polarization will be rotated by an angle Φ_+, which has been observed in the experiment.

In a modified version of this experiment[12] the spin state of the photon has been subjected to a cyclic series of quantum jumps

$$\begin{aligned}&|j=1,m=1(\hat{k}_1)\rangle\\ &\quad\to|j=1,m=1(\hat{k}_2)\rangle\\ &\qquad\times\frac{\langle(\hat{k}_2)m=1,j=1|j=1,m=1(\hat{k}_1)\rangle}{|\langle(\hat{k}_2)m=1,j=1|j=1,m=1(\hat{k}_1)\rangle|}\,.\end{aligned} \tag{2.6}$$

The resulting phase shift turned out to be equal to the area of the geodesic polygon spanned by $\hat{k}_1, \hat{k}_2, \ldots$, on the sphere of spin directions.[18,19]

More generally, let us consider the phase factor $e^{i\Phi}$ induced by a cyclic sequence of quantum jumps,

$$\begin{aligned}&|j,m(\hat{x}_i)\rangle\to|j,m(\hat{x}_{i+1})\rangle\\ &\qquad\times\frac{\langle m(\hat{x}_{i+1}),j|j,m(\hat{x}_i)\rangle}{|\langle m(\hat{x}_{i+1}),j|j,m(\hat{x}_i)\rangle|}\,,\quad i=1,\ldots,3\,,\end{aligned} \tag{2.7}$$

where $\hat{x}_4 = \hat{x}_1$. Obviously,

$$e^{i\Phi} = \frac{\langle m(\hat{x}_1),j|j,m(\hat{x}_3)\rangle\langle m(\hat{x}_3),j|j,m(\hat{x}_2)\rangle\langle m(\hat{x}_2),j|j,m(\hat{x}_1)\rangle}{|\langle m(\hat{x}_1),j|j,m(\hat{x}_3)\rangle\langle m(\hat{x}_3),j|j,m(\hat{x}_2)\rangle\langle m(\hat{x}_2),j|j,m(\hat{x}_1)\rangle|}\,. \tag{2.8}$$

Choosing $\hat{x}_1 = (0,0,1)$, $\hat{x}_2 = (\sin a, 0, \cos a)$, and $x_3 = (\sin b\cos\gamma, \sin b\sin\gamma, \cos b)$, Eq. (2.8) gives

$$\begin{aligned}\Phi &= \arg\sum_l [d^j_{lm}(a)d^j_{lm}(b)e^{i(l-m)\gamma}]\\ &= \arg\sum_l D^j_{ml}(0,-a,-\gamma)D^j_{lm}(0,b,\gamma)\\ &= \arg(d^j_{mm}(\beta)e^{-im\alpha})\,,\end{aligned}$$

with some α and β. Since $d^j_{mm}(\beta)$ is real, we only have to determine α. To this it is enough to perform the calculation for the $j = \frac{1}{2}, m = \frac{1}{2}$ case. Using the standard $j = \frac{1}{2}$ rotation matrices,[13] one obtains that

$$-\frac{\alpha}{2} = \arg\left(\cos\frac{a}{2}\cos\frac{b}{2} + \sin\frac{a}{2}\sin\frac{b}{2}e^{-i\gamma}\right) = -\frac{\Omega}{2}\,,$$

where Ω is the solid angle of the geodesic triangle spanned by the points $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$ (we adopt the sign convention of

Ref. 16). It follows that Φ (mod 2π) is $-m$ times the area of the geodesic triangle spanned by $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$. By spherical symmetry, the result is independent of the specific choice of the points. This geodesic rule clearly generalizes to processes consisting of more than three steps.

Let us now suppose that $\hat{x}_i$ and $\hat{x}_{i+1}$ are not antipodal points on S^2 and thus there is a shortest geodesic C_i connecting them. The final state of the quantum jump (2.7) can be recovered by parallelly transporting the initial state along C_i by using the monopole connection (2.5). This statement (which has been proved by Berry[20] for $j=\frac{1}{2}$) is easy to check by using formulas (2.3) and (2.5) if C_i lies on the equator $\theta=\pi/2$. Then the general case follows by spherical symmetry and gauge invariance. From our parallel transport rule we get, once more, that the phase factor induced by a cyclic sequence of quantum jumps of type (2.7) is the holonomy of connection (2.5) along the associated geodesic polygon on S^2.

III. GENERALIZED DIRAC MONOPOLES FROM BERRY'S PHASE

In this section we study the adiabatic gauge field generated by the Hamiltonian (1.2). Here $H(x)$ acts on the spin space V of dimension 2^n ($n=1,2,\ldots$). For $|x|=r$, the eigenvalues of $H(x)$ are $\pm r$ and V decomposes accordingly as $\mathrm{V}=\mathrm{V}_+(x)\oplus\mathrm{V}_-(x)$. One can choose the eigenvectors in $\mathrm{V}_+(x)$ and $\mathrm{V}_-(x)$ in such a way that they depend only on $\hat{x}=x/r$. This implies that the adiabatic Yang–Mills potentials (and also the transition functions) are independent of r and have vanishing radial components, so they can be recovered from their restrictions to the unit sphere S^{2n} by a pullback operation. For this reason, we restrict our considerations below to S^{2n}. First, we shall clarify the structure of the eigenspace bundles.

Let $\mathrm{Spin}(2n)\subset\mathrm{Spin}(2n+1)$ be the isotropy subgroup of the north pole $N\in S^{2n}$ with respect to the usual left action of $\mathrm{Spin}(2n+1)$ on S^{2n}. Here $S^{2n}=\mathrm{Spin}(2n+1)/\mathrm{Spin}(2n)$ and thus $\mathrm{Spin}(2n+1)$ is a principal fiber bundle[9] over S^{2n} with structure group $\mathrm{Spin}(2n)$. It will be useful to consider the associated vector bundle[9] $V(S^{2n})=\mathrm{Spin}(2n+1)x_{\mathrm{Spin}(2n)}\mathrm{V}$. Now we define a map f: $V(S^{2n})\to S^{2n}x\mathrm{V}$ by the formula

$$[(g,v)]\xrightarrow{f}(\pi(g),gv)\,. \tag{3.1}$$

Here $g\in\mathrm{Spin}(2n+1)$, $v\in\mathrm{V}$, π: $\mathrm{Spin}(2n+1)\to S^{2n}$ is the bundle projection, and the square bracket denotes the equivalence relation defining the elements of $V(S^{2n})$. $\mathrm{Spin}(2n+1)$ and $\mathrm{Cl}(2n+1)$ act in a natural way on the bundles $S^{2n}x\mathrm{V}$ and $V(S^{2n})$. The actions in question are defined by the following formulas:

$$S^{2n}x\mathrm{V}\ni(\pi(g),v)\xrightarrow{g_0\in\mathrm{Spin}(2n+1)}(\pi(g_0g),g_0v)\in S^{2n}x\mathrm{V}\,,$$

$$V(S^{2n})\ni[(g,v)]\xrightarrow{g_0\in\mathrm{Spin}(2n+1)}[(g_0g,v)]\in V(S^{2n})\,,$$

$$S^{2n}x\mathrm{V}\ni(\pi(g),v)\xrightarrow{\Gamma\in\mathrm{Cl}(2n+1)}(\pi(g),\Gamma v)\in S^{2n}x\mathrm{V}\,,$$

$$V(S^{2n})\ni[(g,v)]\xrightarrow{\Gamma\in\mathrm{Cl}(2n+1)}[(g,(g^{-1}\Gamma g)v)]\in V(S^{2n})\,. \tag{3.2}$$

It is easy to verify that f is a bundle isomorphism intertwining the respective actions of $\mathrm{Spin}(2n+1)$ and $\mathrm{Cl}(2n+1)$. Using this isomorphism, from now on we identify $V(S^{2n})$ and $S^{2n}x\mathrm{V}$.

We proceed by splitting V into a direct sum $\mathrm{V}=\mathrm{V}_+(N)\oplus\mathrm{V}_-(N)$ of representations of $\mathrm{Spin}(2n)$ of positive and negative chirality. Here $\mathrm{V}_\pm(N)$ consists of eigenvectors of $H(N)$ of energy (chirality) ± 1. Correspondingly, $V(S^{2n})$ decomposes into a direct sum of subbundles, $V(S^{2n})=V_+(S^{2n})\oplus V_-(S^{2n})$ with $V_\pm(S^{2n})=\mathrm{Spin}(2n+1)x_{\mathrm{Spin}(2n)}\mathrm{V}_\pm(N)$. The important point is that the fiber of $V_\pm(S^{2n})$ over $x\in S^{2n}$ is carried by f into $V_\pm(x)$, the eigensubspace of $H(x)$ of energy ± 1. This means that the eigenspace bundles $\cup_x\mathrm{V}_\pm(x)$ are isomorphic to $V_\pm(S^{2n})$, respectively. Note that the subbundles $V_\pm(S^{2n})$ are invariant under the action of $\mathrm{Spin}(2n+1)$.

Now we turn to the adiabatic connection carried by $V_\pm(S^{2n})$. The relation

$$gH(x)g^{-1}=H(gx) \tag{3.3}$$

implies that this connection will be invariant under the action of $\mathrm{Spin}(2n+1)$ on $V_\pm(S^{2n})$. On the other hand, the bundle $V_\pm(S^{2n})$ is already endowed with an invariant connection of geometric origin. In fact, the Levi-Civita spin connection operating on the bundle of spin frames $\mathrm{Spin}(2n+1)\xrightarrow{\pi}S^{2n}$ induces[9] a connection invariant under $\mathrm{Spin}(2n+1)$ also on the associated vector bundle $V_\pm(S^{2n})$. We shall show that the adiabatic and the Levi-Civita connections on $V_\pm(S^{2n})$ are identical. There could be two ways of proving this. First, one could prove that there is only one connection on $V_\pm(S^{2n})$ that is invariant[9,14] under $\mathrm{Spin}(2n+1)$. Second, one could attempt to verify the equality of the two connections by a direct calculation. While convinced about the validity of the first argument, here we shall follow the second route, which we think is more instructive.

First, we cover S^{2n} by the usual contractible coordinate patches $U_N=S^{2n}\backslash\{S\}$, $U_S=S^{2n}\backslash\{N\}$. Then we choose two sections $g_M:U_M\to\mathrm{Spin}(2n+1)$ $(M=N,S)$ $[g_N(N)=1]$ of the bundle $\mathrm{Spin}(2n+1)\to S^{2n}$. By construction, these satisfy $g_M(x)[N]=x$ for any $x\in U_M$ $(M=N,S)$. Here $\mathrm{Spin}(2n+1)$ acts on S^{2n} via the homomorphism $\mathrm{Spin}(2n+1)\to\mathrm{SO}(2n+1)$. We also fix an orthonormal basis $\{|v_a\rangle$, $a=1,\ldots,2^n\}$ of V, such that the first 2^{n-1} elements are from $\mathrm{V}_+(N)$, the others from $\mathrm{V}_-(N)$. Now we restrict ourselves to $V_+(S^{2n})$ and introduce eigenstate frames [as in (1.4)] by the definition

$$|v_a^M(x)\rangle=g_M(x)|v_a\rangle\,,\quad a=1,\ldots,2^{n-1}\quad(M=N,S)\,. \tag{3.4}$$

Using these frames, we can calculate the Berry connection from (1.6). The result is

$$A^M_{ab}(x) = \langle v_a | g_M^{-1}\, dg_M | v_b \rangle = (g_M^{-1}\, dg_M)_{ab} \qquad (M = N,S), \tag{3.5}$$

on U_N and U_S, respectively. The Lie algebra valued one-form $g_M^{-1}\, dg_M \in \mathrm{spin}(2n+1)$ can be uniquely decomposed as

$$g_M^{-1}\, dg_M = \frac{1}{2} \sum_{\mu\nu} \mathscr{A}^{\mu\nu}_M \sigma_{\mu\nu} + \sum_{\mu} \mathscr{A}^{\mu,2n+1}_M \sigma_{\mu,2n+1}, \tag{3.6}$$

where $\sigma_{\mu,\nu}$ $(\mu,\nu = 1,...,2n)$ generate spin$(2n)$, the isotropy algebra of the north pole. Let P: spin$(2n+1) \to$ spin$(2n)$ denote the natural projection [defined, e.g., with the aid of the decomposition (3.6)]. For further reference, we rewrite (3.5) as

$$A^M = P(g_M^{-1}\, dg_M) \quad (M = N,S). \tag{3.7}$$

Of course, here $P(g_M^{-1}\, dg_M) \in \mathrm{spin}(2n)$ is taken in the positive chirality representation. On $V_-(S^{2n})$ the adiabatic connection is given by the same expression (3.7), but one has to use the negative chirality representation of spin$(2n)$.

In the gauge chosen, the transition function of the bundle $V_+(S^{2n})$ is $g = g_{NS} = g_N g_S^{-1}: U_N \cap U_S \to \mathrm{Spin}(2n)$. In fact, the following relations hold on the intersection $U_N \cap U_S$

$$|v_a^N(x)\rangle = g(x)|v_a^S\rangle = |v_b^S(x)\rangle [g^{-1}(x)]_{ba}, \qquad (a,b = 1,..,2^{n-1})$$

$$A^N(x) = g(x) A^S(x) g^{-1}(x) + g(x)\, dg^{-1}(x). \tag{3.8}$$

Now we determine the local formula of the Levi-Civita spin connection one-forms ω^{ij}_M. To this the starting point is to choose fields of $2n$-beins $e^i_M(x)$ $(i = 1,...,2n)$ on U_M $(M = N,S)$. Let e^I $(I = 1,...,2n+1)$ be the Cartesian coframe of $\mathbb{R}^{2n+1}$. The first $2n$ elements of this basis provide us with a $2n$-bein at the north pole. A convenient moving frame is obtained by rotating the one at N,

$$e^i_M(x) = \bar{g}_M(x)[e^i] \quad (i = 1,...,2n) \quad (M = N,S), \tag{3.9}$$

where $\bar{g}_M(x) \in \mathrm{SO}(2n+1)$ is the homomorphic image of $g_M(x) \in \mathrm{Spin}(2n+1)$. Cartan's structure equations,[9]

$$de^i_M(x) + \omega^{ij}_M(x) \wedge e^j_M(x) = 0, \quad \omega^{ij}_M = -\omega^{ji}_M, \tag{3.10}$$

then yield the explicit formula

$$\omega^{ij}_M = [(\bar{g}_M)^{-1}\, d\bar{g}_M]^{ij}. \tag{3.11}$$

In other words,

$$\omega_M = \bar{P}[(\bar{g}_M)^{-1}\, d\bar{g}_M], \tag{3.12}$$

where $\bar{P}$: so$(2n+1) \to$ so$(2n)$ is the natural projection. Comparing the local representatives (3.7) and (3.12), and taking into account the isomorphism o$(2n) \simeq$ spin$(2n)$ we see that the Berry connection we are interested in is in fact identical to the Levi-Civita connection operating on the bundle $V_\pm(S^{2n})$.

Now we shall derive explicit expressions for the adiabatic connection. To fix the representation for the generators of Cl$(2n+1)$, we apply the following recursive procedure:

$$\Gamma_k^{(2n+1)} = \sigma_1 \otimes \Gamma_k^{(2n-1)}, \quad k = 1,2,...,2n-1,$$
$$\Gamma_1^{(1)} = I^{(1)},$$
$$\Gamma_{2n}^{(2n+1)} = \sigma_2 \otimes I^{(2n-1)}, \quad \Gamma_{2n}^{(2n+1)} = \sigma_3 \otimes I^{(2n-1)}. \tag{3.13}$$

The generators $\Gamma^{(2n+1)}$ are represented by $2^n \times 2^n$ matrices; $I^{(2n-1)}$ denotes the 2^{n-1}-dimensional unit matrix and $\sigma_i - s$ are the standard Pauli matrices. Using Cartesian coordinates in $\mathbb{R}^{2n+1}$, the eigenfunctions of the Hamiltonian (1.2) corresponding to $\pm r$ can be chosen to have the following form:

$$|n^M_a(x)\rangle_\pm = [\sqrt{2r(r + \lambda x^{2n+1})}]^{-1} \times \{[\pm x^i \Gamma_i]_{j,a+|\lambda \mp 1|2^{n-2}} + r\delta_{j,a+|\lambda \mp 1|2^{n-2}}\}_{j=1}^{2^n}, \tag{3.14}$$

respectively, where $\Gamma_i = \Gamma_i^{(2n+1)}$, $\lambda = 1$ for $M = N$, $\lambda = -1$ for $M = S$, and $a = 1,2,...,2^{n-1}$. The Yang–Mills potential calculated by combining (1.6) and (3.14) is

$$A^M_\mu = [-i\lambda / r(r + \lambda x^{2n+1})]\sigma_{\mu\nu} x^\nu, \quad A^M_{2n+1} = 0, \tag{3.15}$$

with $\sigma_{\mu\nu} = (1/4i)[\Gamma_\mu, \Gamma_\nu]$, $\mu\nu = 1,...,2n$. This is the usual expression[5,6] of the generalized Dirac monopole in local coordinates. By restricting this monopole to S^{2n}_r, one recovers the BPST (Belavin–Polyakov–Schwarz–Tyupkin) instanton[4] for $n = 2$ and its higher-dimensional analogs for $n > 2$. The size of the instanton is r, the radius of the sphere.

It is known[21] that a Dirac monopole of charge $eg = s$ appears in the momentum space description of a massless particle of helicity s $(s = 0, \pm\frac{1}{2}, \pm 1,...)$. (The phase shift observed in the optical fiber experiment[16] can be interpreted[17] as a manifestation of the monopole in momentum space.) The relationship between Dirac monopoles and zero mass particles has been recently established[22] also in higher dimensions. Here we give an argument shedding light on this relationship from the viewpoint of Berry's phase. The wave equation of a zero mass particle of minimal spin (left- or right-handed "neutrino") moving in Minkowski space $\mathbb{R}^{1,2n+1}$ can be written as follows:

$$\frac{\partial \Psi}{\partial x^0} = \pm \Gamma_k^{(2n+1)} \frac{\partial \Psi}{\partial x_k}. \tag{3.16}$$

Here Ψ is a 2^n-component Weyl spinor in $\mathbb{R}^{1,2n+1}$ and we use the representation of the Γ matrices given by (3.13). The $\pm$ signs refer to left- and right-handed particles, respectively. Now let us switch to the momentum representation. By substituting $\psi(p)\exp[-i(p^0x^0 - p^kx^k)]$ $[p^0 = |p|,\ p = (p^k)]$ for Ψ, (3.16) is converted into

$$p^k \Gamma_k^{(2n+1)} \psi(p) = \pm |p| \psi(p). \tag{3.17}$$

The Hamiltonian on the left-hand side of this equation is identical in form to the Hamiltonian (1.2) from which we derived the generalized Dirac monopoles. It is for this reason that Dirac monopoles turn up in the momentum space description of zero mass particles.

IV. CONCLUDING REMARKS

We have shown that the Hamiltonian (1.2) yields the generalized Dirac monopoles as the source of the associated Berry's phase. From a mathematical point of view, our result amounts to a kind of linearization of the Levi-Civita connection on S^{2n}. By choosing Hamiltonians with appropriate symmetries, other invariant connections[9,14] could be generated by the adiabatic construction. This could provide an effective mathematical tool for constructing symmetric Yang–Mills fields. Physically, of course, the most interesting would be to find a real system governed by the Hamiltonian (1.2) and to study "quasi-instantons" experimentally.

Recent generalizations[23,24,19] of Berry's phase also enriched our knowledge of the mathematical structure of quantum mechanics. The Aharonov–Anandan phase[23] is due to the fact that if $|\psi(t)\rangle$ is a normalized solution of the time-dependent Schrödinger equation then

$$|\tilde{\psi}(t)\rangle = \exp\{i\langle\psi(t)|H(t)|\psi(t)\rangle\}|\psi(t)\rangle$$

is parallelly transported along the curve $|\psi(t)\rangle\langle\psi(t)|$ in the projective Hilbert space, with respect to the natural connection on the Hopf bundle.[9] Note that $|\psi(t)\rangle\langle\psi(t)|$ itself is an integral curve of the Hamiltonian system defined by the energy function and the canonical symplectic form of the projective Hilbert space.[25]

It turned out[24,19] that quantum jumps can also be described in terms of the same geometry as the Schrödinger equation. Namely, the final state of a quantum jump is recovered by parallelly transporting the initial state along the shortest geodesic connecting the projectors specified by the initial and final states in the projective Hilbert space. Thus the phase induced by a cyclic evolution is a manifestation of the curved geometry of the projective Hilbert space and the Hopf bundle over it.

In an adiabatic situation the map $x \to V_E(x)$ embeds the parameter space into the Grassmannian of the Hilbert space and the Berry–Wilczek–Zee phase arises[2,15] because of the nontrivial geometry of the canonical vector bundle over the Grassmannian.[9]

Note added in proof: After this work was completed, there appeared a paper[26] deriving the BPST instanton from a Hamiltonian possibly allowing for experimental study, which is mathematically equivalent to Hamiltonian (1.2) for $n = 2$. Two other related papers[27,28] have also come to our attention.

ACKNOWLEDGMENT

We are grateful to P. Forgács, P. A. Horváthy, and L. Palla for discussions and comments.

[1] M. V. Berry, Proc. R. Soc. London **392**, 45 (1984).
[2] B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).
[3] F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).
[4] A. Belavin, A. Polyakov, A. Schwarz, and Y. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[5] T. Yoneya, J. Math. Phys. **18**, 1759 (1977); C. N. Yang, *ibid.* **19**, 320 (1978).
[6] Z. Horváth and L. Palla, Nucl. Phys. B **142**, 327 (1978).
[7] E. Cremmer and J. Scherk, Nucl. Phys. B **118**, 61 (1977).
[8] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1962), Vol. II.
[9] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vols. I and II; T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).
[10] T. Bitter and D. Dubbers, Phys. Rev. Lett. **59**, 251 (1987); D. Suter, K. T. Mueller, and A. Pines, *ibid.* **60**, 1218 (1988).
[11] R. Bhandari and J. Samuel, Phys. Rev. Lett. **60**, 1211 (1988).
[12] R. Y. Chiao, A. Antaramian, K. M. Ganga, H. Jiao, S. R. Wilkinson, and H. Nathel, Phys. Rev. Lett. **60**, 1214 (1988); M. Kitano, T. Yabuzaki, and T. Ogawa, *ibid.* **58**, 523 (1987).
[13] E. Wigner, *Group Theory* (Academic, New York, 1959); L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, MA, 1981).
[14] P. Forgács and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).
[15] J. Segert, Phys. Rev. A **36**, 10 (1987); J. Math. Phys. **28**, 2102 (1987).
[16] A. Tomita and R. Y. Chiao, Phys. Rev. Lett. **57**, 937 (1986).
[17] R. Y. Chiao and Yong-Shi Wu, Phys. Rev. Lett. **57**, 933 (1986); I. Białynicki-Birula and Z. Białynicka-Birula, Phys. Rev. D **35**, 2383 (1987).
[18] T. F. Jordan, J. Math. Phys. **28**, 1759 (1987); Phys. Rev. Lett. **60**, 1584 (1988); Phys. Rev. A **38**, 1590 (1988).
[19] M. G. Benedict and L. Gy. Fehér, Phys. Rev. D **39**, 3194 (1989).
[20] M. V. Berry, J. Mod. Opt. **34**, 1401 (1987).
[21] A. O. Barut and A. J. Bracken, Lett. Math. Phys. **7**, 407 (1983).
[22] M. V. Atre, A. P. Balachandran, and T. R. Govindarajan, J. Mod. Phys. A **2**, 453 (1987).
[23] Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987); D. N. Page, Phys. Rev. A **36**, 3479 (1987).
[24] J. Samuel and R. Bhandari, Phys. Rev. Lett. **60**, 2339 (1988); J. Anandan and Y. Aharonov, Phys. Rev. D **38**, 1863 (1988).
[25] G. M. Tuynman and W. A. Wiegerinck, J. Geom. Phys. **4**, 207 (1987).
[26] J. E. Avron, L. Sadun, J. Segert, and B. Simon, Phys. Rev. Lett. **61**, 1329 (1988).
[27] S. Giler, P. Kosiński, and L. Szymanowski, Int. J. Mod. Phys. A **4**, 1453 (1989).
[28] G. M. O'Brien and D. H. Tchrakian, J. Math. Phys. **29**, 1212 (1988).

On generalized coherent states with maximal symmetry for the harmonic oscillator

J. Beckers and N. Debergh
Theoretical and Mathematical Physics, Institute of Physics, B.5, University of Liège, B-4000 Liege 1, Belgium

(Received 31 January 1989; accepted for publication 5 April 1989)

Generalized coherent states for the one-dimensional harmonic oscillator with maximal symmetry, i.e., admitting the semidirect sum so(2,1) □ h(2) as the largest invariance Lie algebra pointed out by Niederer are constructed. The normalization of such states as well as their completeness property are determined and discussed. They are also analyzed in the subcontexts of the so(2,1) algebra and of the Heisenberg h(2) algebra. General considerations on Heisenberg relations, on minimal dispersions, and on quantum mechanical entropies are presented in connection with the uncertainty principle.

I. INTRODUCTION

In connection with our work[1,2] on *supersymmetric* properties of quantum harmonic oscillators, as well as with other recent contributions,[3,4] we planned to search for new supersymmetric characteristics of the so-called coherent states[5–7] originally developed in this harmonic oscillator context. Then we were very surprised to note that, to the best of our knowledge, until now nobody has used the Niederer result[8] for the construction of such coherent states.

In 1973, Niederer[8] pointed out the *largest* invariance Lie group (or algebra) of the quantum harmonic oscillator. In fact, if we insist on the Lie algebra content of such a study, let us recall that this largest invariance algebra refers to *kinematical*[9] symmetries for the n-dimensional harmonic oscillator. By limiting ourselves to the one-dimensional case (for simplicity), this algebra has the nonsemisimple semidirect structure so(2,1) □ h(2). Here the noncompact simple part so(2,1) $\sim$ sp(2) $\sim$ su(1,1) is generated by three operators (hereafter called K_+, K_-, and K_0) while the nonsemisimple part h(2) is the Heisenberg algebra ($[x,p] = i$) with a central extension, generated by two operators (called P_+ and P_-) supplemented by the identity. So the largest invariance algebra, so(2,1) □ h(2), we are dealing with is a six-dimensional algebra whose characteristics will be recalled in the following section.

Historically the algebra h(2) has subtended the original developments on coherent states[5,6] while the algebra so(2,1) has played the prominent role for the construction of the so-called generalized coherent states.[10,11] Now from Niederer's result it seems natural to combine and superpose the respective characteristics for h(2) *and* so(2,1) in order to construct generalized coherent states *with maximal symmetry* (when harmonic oscillators are concerned). This is not a trivial task and it is the main purpose of the present paper.

Here we want to apply Perelomov's method[11,12] to one-dimensional harmonic oscillators described by a Schrödinger equation with the bosonic Hamiltonian

$$H_B = \tfrac{1}{2}(p^2 + x^2), \quad p = -i\left(\frac{d}{dx}\right), \tag{1.1}$$

where we choose as units $\hbar = 1$ as well as the angular frequency $\omega = 1$ and the mass $m = 1$. The symmetries of such an equation are summarized by the generators $[\tfrac{1}{2}H_B \equiv K_0, K_\pm]$ for so(2,1) and $(P_\pm, I)$ for h(2), associated with coordinate transformations explicitly written down by Niederer.[8] In fact, the operators $K_\pm$ deal with "conformal" transformations such as dilations and expansions[9,13] inside the extended Galilean symmetry, while the Heisenberg generators $P_\pm$ are essentially annihilation (a) and creation ($a^\dagger$) operators of the corresponding oscillator when

$$\tfrac{1}{2}H_B \equiv K_0 = \tfrac{1}{2}(a^\dagger a + \tfrac{1}{2}), \quad [a, a^\dagger] = 1. \tag{1.2}$$

The content of this paper is distributed as follows. In Sec. II we successively reexamine the information contained in the algebras h(2), h_4 (introduced by Miller[14]), so(2,1), and so(2,1) □ h(2) from different points of view[12,15] and, particularly, in connection with their (quadratic) Casimir operator content.[15,16] Section III is devoted to the construction of the generalized coherent states with maximal symmetry and to their explicit test for being the set of closest states to the classical ones through the study of the Heisenberg uncertainty relations. In Sec. IV we come back on the uncertainty, but by considering it through the minimal invariant dispersion[12,17] of the Casimir operators associated with our specific structures, as well as through its discussion in terms of quantum mechanical entropies[18] associated with some specific observables such as the position and momentum. As coherent states have to be normalized states inside a well-defined scalar product (as it will be discussed in Sec. III), they also have to satisfy the so-called completeness[7] property: Thus Section V will contain the corresponding discussion on our generalized coherent states with maximal symmetry. There we show how to construct a measure for these states inside the so(2,1) □ h(2) symmetry and its consistency with the particular contexts of the Heisenberg and so(2,1) subalgebras.

II. SYMMETRY LIE ALGEBRAS OF ONE-DIMENSIONAL HARMONIC OSCILLATORS

In order to apply Perelomov's method[11] in the different contexts related to symmetry Lie algebras for the one-di-

 0022-2488/89/081732-07$02.50

mensional harmonic oscillator, let us summarize the necessary information here on the four following cases.

A. The Heisenberg algebra h(2)

As already mentioned in the Introduction, this algebra has subtended the original developments on coherent states.[5,6] It has three generators (P_+, P_-, I), which can be realized in terms of the usual bosonic creation $(a^\dagger)$ and annihilation (a) operators

$$P_+ = ia^\dagger, \quad P_- = -ia, \quad I, \tag{2.1}$$

chosen in this way for consistency with the other algebras we want to consider, according to Niederer's realization.[8] The only nonzero commutation relation reads

$$[P_+, P_-] = -I \tag{2.2}$$

and it is easy to convince ourselves[15] that h(2) is a noncompact, nonsemisimple, solvable, and nilpotent Lie algebra. According to the theorem giving the number of Casimir operators,[16] we easily get that h(2) admits only one, which is the identity operator belonging to the algebra. Then applying Perelomov's considerations[11] we notice that the stationary subalgebra B of the fundamental state $|\Psi_0\rangle$ and its adjoint $\overline{B}$ are given by

$$B \equiv \{P_-, I\}, \quad \overline{B} \equiv \{P_+, I\} \tag{2.3}$$

and we conclude that

$$|\Psi_0\rangle \equiv |0\rangle, \tag{2.4}$$

where $|0\rangle$ is nothing else than the vacuum state. Here the only generator of $\overline{B}$, which does not belong to $B\cap\overline{B}$, is P_+ and, consequently, the generalized (Perelomov) coherent states are defined by

$$|\beta\rangle = Me^{\beta P_+}|0\rangle, \quad \beta \text{ arbitrary}, \tag{2.5}$$

where M is an *ad hoc* normalization factor.

B. The Miller algebra h_4

As a larger algebra, with respect to h(2), Miller[14] has pointed out the structure h_4 generated by the four operators (N, P_+, P_-, I), where, in addition to the realization (2.1), we define the occupation number operator

$$N = a^\dagger a. \tag{2.6}$$

The only nonzero commutation relations are now

$$[N,P_+] = P_+, \quad [N,P_-] = -P_-, \quad [P_+,P_-] = -I, \tag{2.7}$$

and h_4 appears as a noncompact, nonsemisimple, solvable but non-nilpotent Lie algebra evidently containing the Heisenberg algebra. Here we find[16] *two* Casimir operators already obtained by Miller: the identity operator and a quadratic operator

$$C_2^M = P_-P_+ - N. \tag{2.8}$$

The subalgebras B and $\overline{B}$ are $B \equiv \{P_-, N, I\}$, $\overline{B} \equiv (P_+, N, I)$ because of the hermiticity of N and the fundamental state is once again $|\Psi_0\rangle \equiv |0\rangle$, so that the definition of the generalized coherent states for h_4 is *identical* with that for h(2) [see Eq. (2.5)], P_+ also being the only generator of $\overline{B}$ that does not belong to $B\cap\overline{B}$.

C. The algebra so(2,1)

Also playing an interesting role,[15] in connection with the harmonic oscillator, the algebra so(2,1) has been the first example[10] of Lie algebras examined for the construction of the so-called generalized coherent states.[11] Generated by the three operators (K_+, K_-, K_0) it can be realized through the explicit forms

$$K_+ = -(i/2)(a^\dagger)^2, \quad K_- = (i/2)(a)^2, \\ K_0 = \tfrac{1}{2}(a^\dagger a + \tfrac{1}{2}) \equiv (1.2). \tag{2.9}$$

Its nonzero commutation relations are

$$[K_+,K_-] = -2K_0, \quad [K_0,K_+] = K_+, \\ [K_0,K_-] = -K_-, \tag{2.10}$$

and it is a noncompact, simple, nonsolvable, and non-nilpotent algebra. Here we evidently confirm easily[16] that it has only one (quadratic) Casimir operator given by

$$C_2^{so(2,1)} = K_0^2 - \tfrac{1}{2}(K_+K_- + K_-K_+), \tag{2.11}$$

with eigenvalues $k(k-1)$, where $k = \frac{1}{4}$ *or* $\frac{3}{4}$. Its subalgebras B and $\overline{B}$ are

$$B \equiv \{K_-,K_0\}, \quad \overline{B} \equiv \{K_+,K_0\}, \tag{2.12}$$

and the fundamental state can be chosen as (2.4), as above. The only generator belonging to $\overline{B}$ but not to $B\cap\overline{B}$ is K_+ and here we get the Perelomov generalized states on the form

$$|\alpha\rangle = M'e^{\alpha K_+}|0\rangle, \tag{2.13}$$

where α will be constrained on the unit disk $|\alpha| < 1$ in the following[11] and M' is an *ad hoc* normalization factor.

D. The largest kinematical algebra so(2,1) $\square$ h(2)

After the historical facts presented in the Introduction, the algebra so(2,1) $\square$ h(2) appears as the *largest*[8] invariance algebra for the one-dimensional harmonic oscillator. It is generated by the six operators (K_+, K_-, K_0) and (P_+, P_-, I), realized[8] according to Eqs. (2.1) and (2.9). Its nonzero commutation relations evidently collect the relations (2.2) and (2.10), but also the following ones:

$$[K_0,P_+] = \tfrac{1}{2}P_+, \quad [K_0,P_-] = -\tfrac{1}{2}P_-, \\ [K_+,P_-] = -iP_+, \quad [K_-,P_+] = -iP_-. \tag{2.14}$$

Such a semidirect sum is a noncompact, nonsemisimple, nonsolvable, and non-nilpotent algebra. Moreover, here the Beltrametti–Blasi theorem[16] leads to two Casimir operators; the identity operator and a second one, which is not of quadratic order so that for our purpose we do not need its explicit form.

The subalgebras B and $\overline{B}$ are readily determined as

$$B \equiv \{K_-,K_0,P_-,I\}, \quad \overline{B} \equiv \{K_+,K_0,P_+,I\} \tag{2.15}$$

and once again the fundamental state appears as given by the vacuum state (2.4). Applying Perelomov's considerations, we get here, as generalized coherent states, the set $\{|\alpha,\beta\rangle\}$, defined by

$$|\alpha,\beta\rangle = Qe^{\alpha K_+ + \beta P_+}|0\rangle, \tag{2.16}$$

where Q is the normalization factor. Indeed the operators K_+ and P_+ of $\overline{B}$ do not belong to $B\cap\overline{B}$. Let us also mention that the states (2.16) can immediately be written as

$$|\alpha,\beta\rangle = Qe^{\alpha K_+}e^{\beta P_+}|0\rangle, \tag{2.17}$$

as a result of the fact that K_+ and P_+ commute. The consistency of such states $|\alpha,\beta\rangle$ with the corresponding h(2) states $|\beta\rangle$ or (and) so(2,1) states $|\alpha\rangle$ is evident, as are the inclusions so(2,1)□h(2)⊃h(2) and so(2,1)□h(2) ⊃so(2,1).

III. GENERALIZED COHERENT STATES WITH MAXIMAL SYMMETRY

The characterization[7] of coherent states requires the existence of a well-defined (normalized) scalar product, i.e., on our generalized coherent states $|\alpha,\beta\rangle \equiv (2.17)$ with maximal symmetry, we need the relation

$$\langle\alpha,\beta|\alpha\beta\rangle = 1. \tag{3.1}$$

Because of the definition (2.17), let us extract the information on the normalization factor Q, ensuring Eq. (3.1). As already pointed out,[19] this normalization will particularly privilege the so-called modified Bessel functions of the first kind[20]:

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{[\frac{1}{2}z]^{\nu+2m}}{m!\Gamma(\nu+m+1)}. \tag{3.2}$$

By using the current normalized energy states $|n\rangle$ such that

$$\langle n'|n\rangle = \delta_{n'n} \tag{3.3}$$

and by noticing that[8]

$$\begin{aligned} P_+|n\rangle &= i(n+1)^{1/2}|n+1\rangle, \\ K_+|n\rangle &= -i/2[(n+1)(n+2)]^{1/2}|n+2\rangle, \end{aligned} \tag{3.4}$$

the states (2.17) can be rewritten as

$$|\alpha,\beta\rangle = Q\sum_{m,p=0}^{\infty} \frac{(i\beta)^m}{m!}\frac{(-i\alpha)^p}{p!} \times\frac{[(m+2p)!]^{1/2}}{2^p}|m+2p\rangle \tag{3.5}$$

and their scalar product as

$$\langle\alpha,\beta|\alpha,\beta\rangle = |Q|^2(1-|\alpha|^2)^{-1/2}\exp[|\beta|^2/(1-|\alpha|^2)]I_0[|\alpha||\beta|^2/(1-|\alpha|^2)], \tag{3.6}$$

where the $\nu = 0$-modified Bessel function is singled out. Thus we get

$$|Q|^2 = (1-|\alpha|^2)^{1/2}\exp[-|\beta|^2/(1-|\alpha|^2)]\, I_0^{-1}[|\alpha||\beta|^2/(1-|\alpha|^2)]. \tag{3.7}$$

It has to be mentioned that the results (3.6) and (3.7) have been obtained by restricting our summations on bra and ket vectors labeled by the same indices, taking into account the separate actions of the generators P_+ and K_+. Moreover, it is already interesting to notice the Eq. (3.7) corresponds to well-defined results for $\alpha = 0$ *or* $\beta = 0$ because of the property [see Eq. (3.2)]

$$I_0(0) = 1. \tag{3.8}$$

In fact, these particular cases lead back to the original[5,11] normalization factors $|M|^2$ *or* $|M'|^2$ in the h(2) *or* so(2,1) contexts, respectively. In particular, we get for the so(2,1) case

$$Q|_{\beta=0} = M' = (1-|\alpha|^2)^{1/4}, \tag{3.9}$$

which fixes to one-fourth the Perelomov number k [in the eigenvalue of $C_2^{so(2,1)}$, cf. Eq. (2.11)], a typical value for the harmonic oscillator, as everybody knows.

Through the information (3.1), (3.5), and (3.7), it is now possible to evaluate the expectation values of some interesting observables in order to test our generalized coherent states with maximal symmetry as the closest states to classical ones via the Heisenberg uncertainty relations. By using the following further properties:

$$\begin{aligned} x &= \frac{1}{\sqrt{2}}(a^\dagger + a) \\ &= -\frac{i}{\sqrt{2}}(P_+ - P_-), \\ p &= \frac{i}{\sqrt{2}}(a^\dagger - a) = \frac{1}{\sqrt{2}}(P_+ + P_-), \\ P_-|n\rangle &= -i\sqrt{n}|n-1\rangle, \quad I_1(z) = I_{-1}(z), \end{aligned}$$

it is possible to show that[19]

$$\langle x\rangle_{\alpha,\beta} = -\sqrt{2}(\operatorname{Im}\beta)(1-|\alpha|^2)^{-1}(1+|\alpha|(I_1/I_0)) \tag{3.10}$$

and

$$\langle p\rangle_{\alpha,\beta} = \sqrt{2}(\operatorname{Re}\beta)(1-|\alpha|^2)^{-1}(1+|\alpha|(I_1/I_0)), \tag{3.11}$$

where the argument of the modified Bessel functions of the first kind (I_0 and I_1) is once again $|\alpha||\beta|^2(1-|\alpha|^2)^{-1}$ [see Eq. (3.6)]. For brevity we will drop this argument in the following relations. Moreover, the squared expectation values $\langle x\rangle^2_{\alpha,\beta}$, and $\langle p\rangle^2_{\alpha,\beta}$, as well as the expectation values of the squares $\langle x^2\rangle_{\alpha\beta}$ and $\langle p^2\rangle_{\alpha\beta}$, can also be evaluated. After tedious but elegant calculations,[19] we get

$$\langle x\rangle^2_{\alpha,\beta} = 2(\mathrm{Im}^2\beta)(1-|\alpha|^2)^{-2}[1+2|\alpha|(I_1/I_0) + |\alpha|^2(I_1/I_0)^2], \tag{3.12}$$

$$\langle p\rangle^2_{\alpha,\beta} = 2(\mathrm{Re}^2\beta)(1-|\alpha|^2)^{-2}[1+2|\alpha|(I_1/I_0) + |\alpha|^2(I_1/I_0)^2], \tag{3.13}$$

$$\langle x^2\rangle_{\alpha,\beta} = I_0^{-1}A_+, \tag{3.14}$$

$$\langle p^2\rangle_{\alpha,\beta} = I_0^{-1}A_-, \tag{3.15}$$

where

$$A_\pm = \left(\frac{1}{2} + \frac{|\alpha|^2+|\beta|^2}{1-|\alpha|^2} + \frac{2|\alpha|^2|\beta|^2}{(1-|\alpha|^2)^2}\right)I_0 + \frac{2|\alpha|\,|\beta|^2}{(1-|\alpha|^2)^2}I_1 \pm [\mathrm{Im}^2\beta - \mathrm{Re}^2\beta]\left[\frac{(1+|\alpha|^2)}{(1-|\alpha|^2)^2}I_0 + \left(\frac{2|\alpha|}{(1-|\alpha|^2)^2} - \frac{|\alpha|}{|\beta|^2(1-|\alpha|^2)}\right)I_1\right].$$

Evidently, these results do permit us to calculate

$$(\Delta x)^2_{\alpha,\beta} = \langle x^2\rangle_{\alpha,\beta} - \langle x\rangle^2_{\alpha,\beta},\quad (\Delta p)^2_{\alpha,\beta} = \langle p^2\rangle_{\alpha,\beta} - \langle p\rangle^2_{\alpha,\beta}, \tag{3.16}$$

for considering the Heisenberg equality and its conclusions concerning our states. With Eqs. (3.12), (3.14), and (3.16), we easily get

$$(\Delta x)^2_{\alpha,\beta} = \tfrac{1}{2} + \frac{|\alpha|^2}{1-|\alpha|^2} + \frac{2|\alpha|^2}{(1-|\alpha|^2)^2}\mathrm{Im}^2\beta + \frac{|\alpha|(|\beta|^2 - 2\,\mathrm{Im}^2\beta)}{|\beta|^2(1-|\alpha|^2)}\left(\frac{I_1}{I_0}\right) - \frac{2|\alpha|^2\,\mathrm{Im}^2\beta}{(1-|\alpha|^2)^2}\left(\frac{I_1}{I_0}\right)^2, \tag{3.17}$$

while with Eqs. (3.13), (3.15), and (3.16), we obtain

$$(\Delta p)^2_{\alpha,\beta} = \tfrac{1}{2} + \frac{|\alpha|^2}{1-|\alpha|^2} + \frac{2|\alpha|^2}{(1-|\alpha|^2)^2}\mathrm{Re}^2\beta + \frac{|\alpha|(|\beta|^2 - 2\,\mathrm{Re}^2\beta)}{|\beta|^2(1-|\alpha|^2)}\left(\frac{I_1}{I_0}\right) - \frac{2|\alpha|^2\mathrm{Re}^2\beta}{(1-|\alpha|^2)^2}\left(\frac{I_1}{I_0}\right)^2. \tag{3.18}$$

The product of these two expressions can be considered for arbitrary α *and* β. Using the relation[20] connecting the modified Bessel functions I_ν of the first kind to the current Bessel functions J_ν of the first kind, i.e.,

$$I_\nu(z) = \exp(-(i/2)\nu\pi)J_\nu(iz), \tag{3.19}$$

we know from Petiau's book[21] that

$$\frac{I_1(z)}{I_0(z)} = -i\frac{J_1(iz)}{J_0(iz)} = \frac{z}{2} - \frac{z^3}{16} + \frac{z^5}{96}\cdots. \tag{3.20}$$

With the argument $z = |\alpha||\beta|^2(1-|\alpha|^2)^{-1}$, it is thus possible to evaluate $(\Delta x)^2\,(\Delta p)^2$. Let us only mention for simplicity that, in the constrained context[11] of the unit disk $|\alpha| < 1$, by limiting our considerations to first-order terms in $|\alpha|$ (while maintaining β arbitrary), we get from Eq. (3.20)

$$I_1/I_0 \approx \tfrac{1}{2}|\alpha||\beta|^2,\quad (\Delta x)^2 \approx \tfrac{1}{2},\quad (\Delta p)^2 \approx \tfrac{1}{2},$$

so that the (minimal) Heisenberg uncertainty equality is

$$(\Delta x)_{\alpha,\beta}(\Delta p)_{\alpha,\beta} \approx \tfrac{1}{2}, \tag{3.21}$$

showing that our generalized coherent states are the closest states to classical ones.

The above results are also meaningful in the restricted context of the Heisenberg algebra (corresponding here to $\alpha \equiv 0$). From Eqs. (3.10) and (3.11) we deduce

$$\langle x\rangle_{0,\beta} = -\sqrt{2}\,\mathrm{Im}\,\beta,\quad \langle p\rangle_{0,\beta} = \sqrt{2}\,\mathrm{Re}\,\beta, \tag{3.22}$$

according to old results[5,22] within Niederer's realization.[8] All the expressions (3.12)–(3.18) are significant when $\alpha \equiv 0$ and we are led to the well-known result[5]

$$(\Delta x)_{0,\beta}(\Delta p)_{0,\beta} = \tfrac{1}{2}. \tag{3.23}$$

The other restricted context ($\beta \equiv 0$) of the so(2,1) algebra is still meaningful at the level of Eqs. (3.10)–(3.18). We immediately get

$$\begin{aligned}&\langle x\rangle_{\alpha,0} = \langle p\rangle_{\alpha,0} = 0,\quad \langle x\rangle^2_{\alpha,0} = \langle p\rangle^2_{\alpha,0} = 0,\\ &\langle x^2\rangle_{\alpha,0} = (\Delta x)^2_{\alpha,0} = \tfrac{1}{2} + |\alpha|^2/(1-|\alpha|^2)\\ &\qquad = \langle p^2\rangle_{\alpha,0} = (\Delta p)^2_{\alpha,0},\end{aligned} \tag{3.24}$$

so that

$$(\Delta x)_{\alpha,0}(\Delta p)_{\alpha,0} = \tfrac{1}{2},$$

iff $|\alpha|^2 = 0$, a consistent constraint with the one leading to Eq. (3.21). These so(2,1) results are to the best of our knowledge, new ones. They are easy to handle in order to show, after Barut and Girardello,[10] that the so(2,1) results are in correspondence with the Heisenberg ones by remembering that the so(2,1) contracted algebra is isomorphic to the Heisenberg algebra.

IV. UNCERTAINTY, DISPERSION, AND QUANTUM ENTROPY

The uncertainty principle has already been illustrated in Sec. III by considering the Heisenberg relations on the archetypal example of the position (x) and momentum (p) in one dimension. But this is not the only way to express the measure of the uncertainty in the result of a measurement or preparation of an observable A: indeed the most natural measure is the entropy,[18]

$$S_A(|\Psi\rangle) = -\sum_a |\langle a|\Psi\rangle|^2 \ln|\langle a|\Psi\rangle|^2 \tag{4.1}$$

where the summation (integral) is extended to the whole set of eigenvalues a of the operator associated with the discrete (continuous) observable A acting in an appropriate Hilbert space. Let us recall that different quantitative formulations of the uncertainty principle have already been given[23,24] in terms of such quantum mechanical entropies, as it will be summarized hereafter. Let us finally mention that another way to study the indeterminacy or uncertainty principle is to investigate the invariant dispersion[12,17] associated with the quadratic Casimir operator of the Lie algebra subtended by the concerned generalized coherent states. As an example, Perelomov[12] has considered, in the so(2,1) context, the Casimir operator (2.11) [with eigenvalues $k(k-1)$] and its dispersion

$$\Delta C_2^{\mathrm{so}(2,1)} = \langle C_2^{\mathrm{so}(2,1)}\rangle - \langle K_0^2\rangle + \tfrac{1}{2}\langle K_+\rangle\langle K_-\rangle + \tfrac{1}{2}\langle K_-\rangle\langle K_+\rangle, \tag{4.2}$$

which is minimal ($\Delta C_2 = k$) for the vector of lowest weight $|\Psi_0\rangle \equiv |k,m=k\rangle$, m being the eigenvalue of the generator K_0. Parallel considerations can be applied to h_4 but neither to h(2) nor to so(2,1) □ h(2) because of the absence of quadratic Casimir operators in these last two algebras (see Sec. II).

Here let us only point out that, with the information of Sec. II B, it is easy to find that the dispersion of the Miller–Casimir operator of h_4 is minimal,

$$\Delta C_2^M = 1, \tag{4.3}$$

while, with the information of Sec. II C, we immediately recover the Perelomov result on $\Delta C_2^{\mathrm{so}(2,1)}$.

Let us now come back on entropic formulations of the uncertainty through position and momentum as (continuous) observables. We want to evaluate, with the help of Eq. (4.1), the quantity[23,24]

$$\mathfrak{A}(x,p;|\Psi\rangle) = S_x(|\Psi\rangle) + S_p(|\Psi\rangle), \tag{4.4}$$

where $|\Psi\rangle$ is the relative state representing the outcome of the concerned measurement or preparation when our (one-dimensional) harmonic oscillator is described by the generalized coherent states (3.5), admitting a maximal symmetry. We have found

$$\begin{aligned} S_x(|\alpha,\beta\rangle) = {}& I_0^{-1}\exp[-|\alpha|^2|\beta|^2/(1-|\alpha|^2)] \\ &\times \exp[(b^2-ac)/a] \\ &\times\Big[\tfrac{1}{2}+\tfrac{1}{2}\ln\pi + \frac{|\alpha|^2|\beta|^2}{1-|\alpha|^2} - \ln(a^{1/2}I_0^{-1}) \\ &+\Big(\sqrt{2}D - \frac{b}{a}\Big)^2 \\ &-(|\alpha|^2+A_2)(D^2-B_1^2)M^{-1} \\ &-2DM^{-1}A_1B_1\Big] \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} &A_1\equiv \mathrm{Re}\,\alpha,\quad A_2\equiv\mathrm{Im}\,\alpha,\quad B_1\equiv\mathrm{Re}\,\beta,\quad B_2\equiv\mathrm{Im}\,\beta,\\ &a=(1-|\alpha|^2)M^{-1},\quad M=1+2A_2+|\alpha|^2,\\ &b=\sqrt{2}(A_1B_1+A_2B_2+B_2)M^{-1},\\ &c=(|\alpha|^2+A_2)|\beta|^2M^{-1}+b\sqrt{2}B_2, \end{aligned} \tag{4.6}$$

and

$$D=\sqrt{2}ba^{-1}-B_2.$$

The corresponding result on the momentum is readily obtained from Eqs. (4.5) and (4.6) through the following substitutions and correspondences:

$$\begin{aligned} &x\to p,\quad A_1\to A_1,\quad A_2\to -A_2,\\ &B_1\to -B_2,\quad B_2\to -B_1, \end{aligned} \tag{4.7}$$

which are already working in Eqs. (3.10)–(3.18). The sum (4.4) can now be written in final form as

$$\begin{aligned} \mathfrak{A}(x,p;|\alpha,\beta\rangle) = {}& I_0^{-1}\exp\Big(\frac{2A_1B_1B_2+A_2B_2^2-A_2B_1^2}{(1-A_1^2-A_2^2)}\Big)\\ &\cdot\Big[1+\ln\pi+2\ln I_0-\ln(1-A_1^2-A_2^2)+\tfrac{1}{2}\ln(1+2A_2+A_1^2+A_2^2)\\ &+\tfrac{1}{2}\ln(1-2A_2+A_1^2+A_2^2)-\frac{2(2A_1B_1B_2+A_2B_2^2-A_2B_1^2)}{1-A_1^2-A_2^2}\Big]. \end{aligned} \tag{4.8}$$

This result illustrates the recent theorem[24];

$$\mathfrak{A}(x,p;|\alpha,\beta\rangle)\geqslant\mathfrak{B} \tag{4.9}$$

and shows that our generalized coherent states with maximal symmetry lead to the boundedness from below the sum of the two entropies, the lower bound being a constant in the position-momentum considerations, as expected.

The discussion of this approach has already been published[25] for arbitrary α and β as well as its implications for the specific contexts ($\alpha=0,\beta$ arbitrary) or (α arbitrary but $|\alpha|\langle 1, \beta=0$). In each case it is possible to see our generalized coherent states with maximal symmetry as the closest ones to classical states. The minimum value of $\mathfrak{B}$ is determined as $(1+\ln\pi)$ according to general theorems[23] and specific values of our parameters.

V. COMPLETENESS PROPERTY AND MEASURE

In order to ensure the fact that we are dealing with coherent states, we also have to require the existence of a well-defined (strictly positive) measure $\mu(\alpha,\beta)$ so that the so-called completeness property[7] is satisfied, i.e.,

$$\int|\alpha,\beta\rangle\langle\alpha,\beta|\,d\mu(\alpha,\beta)=I. \tag{5.1}$$

This second characteristic of our generalized coherent states $|\alpha,\beta\rangle$, defined by Eqs. (3.5) and (3.7), is to be trivial in the context of the maximal symmetry related to the algebra so(2,1)□h(2).

Let us start by writing

$$d\mu(\alpha,\beta)=I_0 f(|\alpha|,|\beta|)d^2\alpha\, d^2\beta, \tag{5.2}$$

so that the left-hand side of Eq. (5.1) reads

$$\int |\alpha,\beta\rangle\langle\alpha,\beta|\,d\mu(\alpha,\beta) = \sum_{m,p=0}^{\infty} \frac{(m+2p)!}{(m!)^2(p!)^2 2^{2p}} \int (1-|\alpha|^2)^{1/2} \exp\left(-\frac{|\beta|^2}{1-|\alpha|^2}\right) \cdot|\beta|^{2m}|\alpha|^{2p} f(|\alpha|,|\beta|)\,d^2\alpha\, d^2\beta\, |m+2p\rangle\langle m+2p| \equiv (C). \tag{5.3}$$

Our problem consists in fixing $f(|\alpha|,|\beta|)$ such that the right-hand side of Eq. (5.3) reduces to the identity operator. By remembering that $|\alpha|$ is constrained on the unit disk while $|\beta|$ is left arbitrary, we can transform the integrals on α *and* β into integrals on $|\alpha|$ *and* $|\beta|^2$ and rewrite (C) as

$$(C) = \sum_{m,p=0}^{\infty} \frac{(m+2p)!2\pi^2}{(m!)^2(p!)^2 2^{2p}} \int_0^{\infty} |\beta|^{2m}\left[\int_0^1 (1-|\alpha|^2)^{1/2}|\alpha|^{2p+1} \cdot\exp\left(-\frac{|\beta|^2}{1-|\alpha|^2}\right) f(|\alpha|,|\beta|)\,d|\alpha|\right] d|\beta|^2 |m+2p\rangle\langle m+2p| \equiv (C'). \tag{5.4}$$

Let us now put

$$f(|\alpha|,|\beta|) = [|\beta|^{2q}|\alpha|^{2t}/(1-|\alpha|^2)^s]\,g(|\beta|^2) \tag{5.5}$$

and search for fixing the real numbers q, t, and s. Through the change of variable $(1-|\alpha|^2) = v$, we get

$$(C') = \sum_{m,p=0}^{\infty} \frac{(m+2p)!\pi^2}{(m!)^2(p!)^2 2^{2p}} \int_0^{\infty} g(|\beta|^2)|\beta|^{2m+2q} \times\left[\int_0^1 v^{1/2-s}(1-v)^{p+t}\cdot\exp\left(-\frac{|\beta|^2}{v}\right)dv\right] \times d|\beta|^2 |m+2p\rangle\langle m+2p| \equiv (C''). \tag{5.6}$$

Noticing that[26]

$$\int_0^u v^{\nu-1}(u-v)^{\delta-1}\exp(-\gamma v^{-1})\,dv = \gamma^{(\nu-1)/2} u^{(2\delta+\nu-1)/2} \times\exp\left(-\frac{\gamma}{2u}\right)\Gamma(\delta)\, W_{(1-2\delta-\nu)/2,\nu/2}(\gamma u^{-1}), \tag{5.7}$$

when $u>0$, $\operatorname{Re}\delta>0$, and $\operatorname{Re}\gamma>0$, we deduce that such a relation requires $t>-1$ and $|\beta|^2\neq 0$ in our context as it immediately follows from the identifications $u=1$, $\delta = p+t+1$, and $\gamma = |\beta|^2$. The integral on v inside (C'') is then obtained and leads to

$$\int_0^{\infty} |\beta|^{2m+2q+1/2-s} g(|\beta|^2)\exp\left(-\frac{|\beta|^2}{2}\right) W_{s/2-p-t-5/4,3/4-s/2}(|\beta|^2)\,d|\beta|^2 = \frac{[\Gamma(m+1)]^2[\Gamma(p+1)]^2 2^{2p}}{\pi^2\Gamma(m+2p+1)\Gamma(p+t+1)}, \quad \forall m,p = 0,1,\ldots,\infty, \tag{5.8}$$

a condition ensuring our final demand (5.1), according to

$$\sum_{m,p=0}^{\infty} |m+2p\rangle\langle m+2p| = I. \tag{5.9}$$

In Eqs. (5.7) and (5.8) we notice the appearance of the so-called Whittaker and gamma functions, whose properties are well known and quoted in the literature.[20]

Our last step is now the study of the condition (5.8). Choosing $t=0$ and

$$g(|\beta|^2) = |\beta|^{s-2q-1/2} h(|\beta|^2), \tag{5.10}$$

we also multiply each member of Eq. (5.8) by $(iy)^m[\Gamma(m+1)]^{-1}$ and sum over all m, y being a real variable. Then we obtain

$$\int_0^{\infty} \exp(i|\beta|^2 y)\exp\left(\frac{-|\beta|^2}{2}\right) h(|\beta|^2)\, W_{s/2-p-5/4,3/4-s/2} \times(|\beta|^2)\,d|\beta|^2 = \frac{\Gamma(p+1)2^{2p}}{\pi^2}\sum_{m=0}^{\infty}\frac{\Gamma(m+1)(iy)^m}{\Gamma(m+2p+1)} = M(y), \quad \forall p. \tag{5.11}$$

By noticing that the series $M(y)$ can be expressed in terms of the hypergeometric function[20] ${}_2F_1(1,1;2p+1,iy)$ as

$$M(y) = [\Gamma(\tfrac{1}{2})/\pi^2\Gamma(p+\tfrac{1}{2})]\,{}_2F_1(1,1;2p+1;iy), \tag{5.12}$$

we know that it absolutely converges for $|y|<1$ while it has to be analytically continued for $|y|>1$. Let us also point out that for $|y|=1$ we have to consider[20] the quantity $1+1-(2p+1) = 1-2p$ and see that *for* $p=0$, *the series diverges*, while for $p\neq 0$, i.e., for $p\geqslant 1$, it absolutely converges. This singles out all the values of p in our summation, except the $p=0$ one.

Coming back on the information (5.12) inside Eq. (5.11), we get

$$h(|\beta|^2) = \frac{1}{2\pi}\exp\left(\frac{|\beta|^2}{2}\right)[W_{s/2-p-5/4,3/4-s/2}(|\beta|^2)]^{-1} \times\int_{-\infty}^{+\infty}\exp(-i|\beta|^2 y)M(y)\,dy. \tag{5.13}$$

This shows that for each p, we are dealing with a well-defined measure referred to

$$d\mu_p(\alpha,\beta) = \frac{I_0}{2\pi}\exp\left(\frac{|\beta|^2}{2}\right)\frac{|\beta|^{s-1/2}}{(1-|\alpha|^2)^s} \times[W_{s/2-p-5/4,3/4-s/2}(|\beta|^2)]^{-1} \times\left[\int_{-\infty}^{+\infty}\exp(-i|\beta|^2 y)M(y)\,dy\right] d^2\alpha\, d^2\beta. \tag{5.14}$$

The integral in Eq. (5.14) can now be evaluated through the information (5.12), the change of variable $z = (1/2) - iy$, and an inverse Laplace transform.[27] Indeed we have

$$\int_{1/2 - i\infty}^{1/2 + i\infty} \exp(|\beta|^2 z)\, {}_2F_1\!\left(1,1;2p+1;\frac{1}{2} - z\right) dz$$
$$= 2i\pi\Gamma(2p+1)\,\frac{1}{|\beta|}\,W_{1/2-2p,0}(|\beta|^2). \qquad (5.15)$$

We finally get from Eqs. (5.12), (5.14), and (5.15), that

$$d\mu_p = \frac{I_0}{\pi^2}\,\frac{(|\beta|^2)^{s/2-3/4}}{(1-|\alpha|^2)^s}\,\Gamma(p+1)2^{2p} \times \frac{W_{1/2-2p,0}(|\beta|^2)}{W_{s/2-p-5/4,3/4-s/2}(|\beta|^2)}\,d^2\alpha\, d^2\beta, \qquad (5.16)$$

where only the real number s is still free.

We can go further in order to fix s by noticing that the measure we are searching for has to be consistent, in particular, with the well-defined measure characteristic of the special case $\alpha = 0$, i.e., the Heisenberg context and algebra, where[5]

$$d\mu_{h_{(2)}} \sim (1/\pi) d^2\beta. \qquad (5.17)$$

If Eqs. (5.16) and (5.17) have to be consistent at the limit $\alpha \to 0$, we immediately notice that the ratio of the Whittaker functions contained in Eq. (5.16) cannot be maintained. So, with the property,[20] the *only* one giving an equality between two W's,

$$W_{a,b}(|\beta|^2) = W_{a,-b}(|\beta|^2),$$

we point out that

$$b = 0 \Rightarrow -b = 0 \Rightarrow \tfrac{3}{4} - s/2 = 0 \Rightarrow s = \tfrac{3}{2}, \qquad (5.18a)$$

leading (with $s = \frac{3}{2}$) to

$$(\tfrac{1}{2}) - 2p = (\tfrac{3}{4}) - p - \tfrac{5}{4} \Rightarrow p = 1. \qquad (5.18b)$$

These values (5.18), introduced in Eq. (5.16), restricted to the only variable β (when $\alpha = 0$), lead to a perfect agreement with Eq. (5.17), the argument of the modified Bessel function I_0 being evidently equal to zero. A complementary argument leading to the measure (5.17) is the cancellation of the $|\beta|^2$ exponent so that the values (5.18) are once again recovered. As a further comment, let us also recall that the other particular context $\beta = 0$ is forbidden since $|\beta|^2$ has to be strictly positive [cf. Eq. (5.7)]: this is in complete agreement with the Perelomov result,[12] which corresponds here to $k = \frac{1}{4}$ as already noticed [see Eq. (3.9)].

Let us end this section by pointing out that the value $p = 1$, leading to the expected result (5.17) in the particular context $\alpha = 0$, is precisely a remarkable value ensuring the absolute convergence of our preceding developments for $|y| < 1$, $|y| > 1$, *and* $|y| = 1$ [see our discussion after Eq. (5.12)].

Note added in proof: The restriction on summations leading to the results (3.6) and (3.7) can be relaxed in order to get the general normalization factor Q. This leads to the following substitution in our main relations:

$$I_0\!\left(\frac{|\alpha|\,|\beta|^2}{1-|\alpha|^2}\right) \to \exp\!\left[\frac{\mathrm{Re}(i\alpha\beta^{*2})}{1-|\alpha|^2}\right].$$

The results become independent of the modified Bessel functions in such a general context. The new Gaussian factor has no effect in expectation values nor in entropies. The main role is left to the α-dependent factor issued from the SO(2,1) considerations.

[1] J. Beckers and V. Hussin, Phys. Lett. A **118**, 319 (1986).
[2] J. Beckers, D. Dehin, and V. Hussin, J. Phys. A: Math. Gen. **20**, 1137 (1987); **21**, 651 (1988); J. Math. Phys. **29**, 1705 (1988).
[3] C. Aragone and F. Zypman, J. Phys. A: Math. Gen. **19**, 2267 (1986).
[4] A. B. Balantekin, H. A. Schmitt, and B. R. Barrett, J. Math. Phys. **29**, 1634 (1988).
[5] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[6] J. R. Klauder, J. Math. Phys. **4**, 1055 (1963).
[7] J. R. Klauder and B. S. Skagerstam, *Coherent States—Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1985).
[8] U. Niederer, Helv. Phys. Acta **46**, 191 (1973).
[9] U. Niederer, Helv. Phys. Acta **45**, 802 (1972).
[10] A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[11] A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972); *Generalized Coherent States and Their Applications*, Texts and Monographs in Physics (Springer, Berlin, 1986).
[12] A. M. Perelomov, Sov. Phys. Usp. **20**, 703 (1977); Sov. J. Nucl. Phys. **29**, 867 (1979).
[13] C. R. Hagen, Phys. Rev. D **5**, 377 (1972).
[14] W. Miller, Jr., *Symmetry Groups and Their Applications* (Academic, New York, 1972).
[15] B. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[16] E. G. Beltrametti and A. Blasi, Phys. Lett. **20**, 62 (1966).
[17] R. Delbourgo and J. R. Fox, J. Phys. A: Math. Gen. **10**, L233 (1977).
[18] J. von Neumann, *Mathematical Foundations of Quantum Mechanics—Investigations in Physics* (Princeton U.P., Princeton, NJ, 1955), Vol. 2.
[19] J. Beckers and N. Debergh, Bull. Soc. R. Sc. Liège **57**, 637 (1988).
[20] M. Abramowitz and I. A. Stegun, *Handbook for Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (National Bureau of Standards Applied Mathematics, Washington, 1965), 4th ed.
[21] G. Petiau, *La théorie des fonctions de Bessel* (C.N.R.S., Paris, 1955).
[22] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Mécanique quantique*, 1 (Hermann, Paris, 1973).
[23] I. Bialynicki-Birula and J. Mycielski, Commun. Math. Phys. **44**, 129 (1975); W. Beckner, Proc. Nat. Acad. Sci. U.S.A. **72**, 638 (1975).
[24] D. Deutsch, Phys. Rev. Lett. **50**, 631 (1983); M. H. Partovi, *ibid.* **50**, 1883 (1983).
[25] J. Beckers and N. Debergh, Bull. Soc. Roy. Sc. Liège **58**, 111 (1989).
[26] I. S. Gradshteyn and J. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965), 4th ed.
[27] H. Bateman, *Tables of Integral Transforms*, edited by A. Erdelyi (McGraw-Hill, New York, 1954), Vol. I.

On quantum integrable models related to nonlinear quantum optics. An algebraic Bethe ansatz approach

Branislav Jurčo
Joint Laboratory of Optics, Czechoslovak Academy of Sciences and Palacký University, Gottwaldova 15, CS-771 46 Olomouc, Czechoslovakia

(Received 10 November 1988; accepted for publication 22 February 1989)

A unified approach based on Bethe ansatz in a large variety of integrable models in quantum optics is given. Second harmonics generation, three-boson interaction, the Dicke model, and some cases of four-boson interaction as special cases of su(2) ⊕ su(1,1)-Gaudin models are included.

I. INTRODUCTION

In 1973, Gaudin introduced a new class of integrable quantum models,[1] however only recently have they attracted increased attention.[2,3] As it is now known, these models can be associated with solutions of classical Yang–Baxter equations. As a matter of fact, these models are better tractable than those associated with quantum Yang–Baxter equations.[4] To be more concrete, we assume a finite-dimensional Lie algebra g and its representation $T = \otimes T_j$. Let X^j_a represent the generators X_a in T_j and coefficients w^{jl}_a are such that the operators (special elements of corresponding representation of enveloping algebra)

$$H_j = \sum_{\substack{l=1\\ l\neq j}}^{M} \sum_a w^{jl}_a X^j_a X^l_a \tag{1.1}$$

are mutually commuting. In the case $g = \mathrm{su}(2)$, Gaudin diagonalized such models first with the help of the coordinate Bethe ansatz,[1,5] and later by algebraic Bethe ansatz[6] in the sense of quantum inverse scattering method [in his book[6] he also diagonalized the rational u(n) models]. Sklyanin shows the possibility of functional Bethe ansatz approach for the rational su(2)-model.[2] Recently, the algebraic Bethe ansatz (using only classical Yang–Baxter equations) was applied to a wide class of these models associated with classical Lie algebras.[3] In this form, algebraic Bethe ansatz can be viewed as a simpler variant of the quantum inverse scattering method.

Some of these models (or their limiting cases) related to the su(2) ⊕ su(1,1) case are of great interest in nonlinear quantum optics. In this paper we show how to apply algebraic Bethe ansatz to the models of second harmonics generation, three-boson interaction, the Dicke model, and some degenerated cases of four-boson interaction.[7–9] In this way algebraic Bethe ansatz as developed in Ref. 3 gives a unified approach to a large variety of models in quantum optics.

In Sec. II, we fix notations and describe the algebraic Bethe ansatz in a form suitable to discuss the above mentioned quantum optical models in Sec. III. Section IV contains some additional comments.

II. ALGEBRAIC BETHE ANSATZ

To make this paper self-contained, we review the algebraic Bethe ansatz in detail for later use.

Let $L(\lambda)$ be a 2×2 matrix

$$L(\lambda) = \begin{pmatrix} A(\lambda) & B(\lambda) \\ C(\lambda) & D(\lambda) \end{pmatrix}, \tag{2.1}$$

its elements being quantum operators depending on complex parameter λ, acting in some Hilbert space $\mathscr{H}$. We assume that the commutation relations for these matrix elements can be written in the following compact form,

$$[L(\lambda)\otimes I, I\otimes L(\mu)]_0 + [r(\lambda-\mu), L(\lambda)\otimes I + I\otimes L(\mu)]_0 = 0, \tag{2.2}$$

where I is the unit 2×2 matrix and $r(\lambda-\mu)$ is a C-number 4×4 matrix of the form

$$r(\lambda) = \begin{pmatrix} f(\lambda) & 0 & 0 & 0 \\ 0 & 0 & g(\lambda) & 0 \\ 0 & g(\lambda) & 0 & 0 \\ 0 & 0 & 0 & f(\lambda) \end{pmatrix}. \tag{2.3}$$

Here the functions $f(\lambda)$, $g(\lambda)$ are (i) $f(\lambda) = g(\lambda) = 1/\lambda$ in the "rational case" and (ii) $f(\lambda) = \cot g\lambda$, $g(\lambda) = 1/\sin\lambda$ in the "trigonometric case." Note that $[\ ,\ ]_0$ is a matrix commutator respecting the operator nature of matrix elements of L. Further, we introduce operators

$$T(\lambda) = \tfrac{1}{2}\,\mathrm{Tr}_0\, L^2(\lambda), \tag{2.4}$$

where Tr_0 is the "2×2 matrix trace." The family of operators $T(\lambda)$ has properties[3]

$$[T(\lambda), T(\mu)] = 0, \tag{2.5}$$

$$[T(\lambda), L(\mu)] + [\mathrm{Tr}_0(r(\lambda-\mu)(L(\lambda)\otimes I)), L(\mu)]_0 = 0, \tag{2.6}$$

where $[\ ,\ ]$ is an operator-commutator.

From (2.2) we obtain

$$\begin{aligned} [A(\lambda), B(\mu)] &= -f(\lambda-\mu)B(\mu) + g(\lambda-\mu)B(\lambda), \\ [D(\lambda), B(\mu)] &= -g(\lambda-\mu)B(\lambda) + f(\lambda-\mu)B(\mu), \\ [B(\lambda), B(\mu)] &= 0, \quad [B(\lambda), C(\lambda)] = A'(\lambda) - D'(\lambda), \end{aligned} \tag{2.7}$$

and from (2.6)

$$\begin{aligned} [T(\lambda), B(\mu)] &= g(\lambda-\mu)A(\mu)B(\lambda) + f(\lambda-\mu)B(\mu)D(\lambda) \\ &\quad - f(\lambda-\mu)A(\lambda)B(\mu) - g(\lambda-\mu)B(\lambda)D(\mu). \end{aligned} \tag{2.8}$$

Further we assume that there is the so-called pseudovacuum, i.e., a vector $|0\rangle \in \mathscr{H}$, such that

 0022-2488/89/081739-05$02.50

$$C(\lambda)|0\rangle = 0,$$
$$A(\lambda)|0\rangle = a(\lambda)|0\rangle, \quad D(\lambda)|0\rangle = d(\lambda)|0\rangle,$$
$$T(\lambda)|0\rangle = t(\lambda)|0\rangle, \tag{2.9}$$

holds $[t(\lambda) = \frac{1}{2}(a^2(\lambda) + d^2(\lambda) + a'(\lambda) - d'(\lambda))$ from (2.7)].

Now, having all necessary ingredients of algebraic Bethe ansatz, we construct the family of vectors

$$|\lambda_1,...,\lambda_K\rangle = B(\lambda_1)\cdots B(\lambda_K)|0\rangle . \tag{2.10}$$

Using relations (2.7), (2.8), and definition (2.9) we shall show that (2.10) are eigenvectors of $T(\lambda)$ if some conditions on λ_i are satisfied. By the straightforward computation,

$$\begin{aligned}
&T(\lambda)B(\lambda_1)\cdots B(\lambda_K)|0\rangle \\
&= \sum_i g(\lambda-\lambda_i)B(\lambda_1)\cdots B(\lambda_{i-1})A(\lambda_i)B(\lambda)B(\lambda_{i+1})\cdots B(\lambda_K)|0\rangle \\
&\quad + \sum_i f(\lambda-\lambda_i)B(\lambda_1)\cdots B(\lambda_i)D(\lambda)B(\lambda_{i+1})\cdots B(\lambda_K)|0\rangle \\
&\quad - \sum_i f(\lambda-\lambda_i)B(\lambda_1)\cdots B(\lambda_{i-1})A(\lambda)B(\lambda_i)\cdots B(\lambda_K)|0\rangle \\
&\quad - \sum_i g(\lambda-\lambda_i)B(\lambda_1)\cdots B(\lambda_{i-1})B(\lambda)D(\lambda_i)B(\lambda_{i+1})\cdots B(\lambda_K)|0\rangle + B(\lambda_1)\cdots B(\lambda_K)T(\lambda)|0\rangle \\
&= \left\{\sum_i (f^2(\lambda-\lambda_i) - g^2(\lambda-\lambda_i)) + t(\lambda)\right\}|\lambda_1,...,\lambda_K\rangle \\
&\quad + \sum_i g(\lambda-\lambda_i)B(\lambda_1)\cdots B(\lambda)(A(\lambda_i) - D(\lambda_i))B(\lambda_{i+1})\cdots B(\lambda_K)|0\rangle \\
&\quad + \sum_i f(\lambda-\lambda_i)B(\lambda_1)\cdots B(\lambda_i)(D(\lambda) - A(\lambda))B(\lambda_{i+1})\cdots B(\lambda_K)|0\rangle \\
&= \left\{\sum_i \left(f^2(\lambda-\lambda_i) - g^2(\lambda-\lambda_i) + \sum_{j\neq i} f(\lambda-\lambda_i)f(\lambda-\lambda_j)\right) + t(\lambda) + \sum_i f(\lambda-\lambda_i)(d(\lambda)-a(\lambda))\right\}|\lambda_1,...,\lambda_K\rangle \\
&\quad + \sum_i g(\lambda-\lambda_i)\left[(a(\lambda_i) - d(\lambda_i)) - 2\sum_{j\neq i} f(\lambda_i-\lambda_j)\right]|\lambda_1,...,\lambda_{i-1},\lambda,\lambda_{i+1},...,\lambda_K\rangle ,
\end{aligned}$$

we obtained the result that each vector (2.10) is an eigenvector of $T(\lambda)$ with the eigenvalue

$$\sum_i \left[f^2(\lambda-\lambda_i) - g^2(\lambda-\lambda_i) + f(\lambda-\lambda_i)(d(\lambda)-a(\lambda)) + \sum_{j\neq i} f(\lambda-\lambda_i)f(\lambda-\lambda_j)\right] + t(\lambda) \tag{2.11}$$

if and only if complex numbers $\{\lambda_1,...,\lambda_K\}$ satisfy the equations

$$\varphi_i = a(\lambda_i) - d(\lambda_i) - 2\sum_{j\neq i} f(\lambda_i - \lambda_j) = 0. \tag{2.12}$$

We omit the proof that the norm squared is equal to

$$\langle\lambda_1,...,\lambda_K|\lambda_1,...,\lambda_K\rangle = \det\left|\frac{\partial\varphi_i}{\partial\lambda_k}\right| \tag{2.13}$$

(see Refs. 5, 6, 10).

The algebraic Bethe ansatz is just given by relations (2.10)–(2.12).

III. INTEGRABLE MODELS IN NONLINEAR QUANTUM OPTICS

Let S_α, K_α, $\alpha = 1, 2, 3$ be generators of Lie algebras su(2), su(1,1), respectively,

$$[S_1,S_2] = iS_3, \quad [S_2,S_3] = iS_1, \quad [S_3,S_1] = iS_2, \tag{3.1}$$

$$[K_1,K_2] = -iK_3, \quad [K_2,K_3] = iK_1, \quad [K_3,K_1] = iK_2. \tag{3.2}$$

Denote as usual

$$S_\pm = S_1 \pm iS_2, \tag{3.3}$$

$$K_\pm = \pm i(K_1 \pm iK_2). \tag{3.4}$$

Consider a chain of M_1 sites with su(2) spins and M_2 sites with su(1,1) spins and representation $T = \bigotimes_{i_1=1}^{M_1} (T_{i_1} \oplus 0) \otimes \bigotimes_{i_2=M_1+1}^{M_1+M_2} (0 \oplus T_{i_2})$ of su(2) $\oplus$ su(1,1) in a tensor product Hilbert space, where T_{i_1} and T_{i_2} are representations of su(2) and su(1,1), respectively. If elements of $L(\lambda)$ are constructed as operators,

$$\begin{aligned}
A(\lambda) &= -\sum_{i_1=1}^{M_1} f(\lambda-\varepsilon_{i_1})S_3^{i_1} \\
&\quad - \sum_{i_2=M_1+1}^{M_1+M_2} f(\lambda-\varepsilon_{i_2})K_3^{i_2} , \\
B(\lambda) &= \sum_{i_1=1}^{M_1} g(\lambda-\varepsilon_{i_1})S_+^{i_1} \\
&\quad + \sum_{i_2=M_1+1}^{M_1+M_2} g(\lambda-\varepsilon_{i_2})K_+^{i_2} , \\
C(\lambda) &= \sum_{i_1=1}^{M_1} g(\lambda-\varepsilon_{i_1})S_-^{i_1}
\end{aligned}$$

$$-\sum_{i_1=M_1+1}^{M_1+M_2} g(\lambda-\varepsilon_{i_2})K^{i_2}_{-}\,, \tag{3.5}$$

$$D(\lambda)=-A(\lambda),$$

then $L(\lambda)$ fulfills relation (2.2).

The Hamiltonians of type (1.1) are obtained from $T(\lambda)$,

$$H_j=\operatorname*{res}_{\lambda=\varepsilon_j} T(\lambda), \tag{3.6}$$

e.g.,

$$\begin{aligned} H_1=\sum_{i_1=2}^{M_1}&\{2f(\varepsilon_1-\varepsilon_{i_1})S^1_3S^{i_1}_3\\ &+g(\varepsilon_1-\varepsilon_{i_1})(S^1_+S^{i_1}_-+S^1_-S^{i_1}_+)\}\\ &+\sum_{i_2=M_1+1}^{M_1}\{2f(\varepsilon_1-\varepsilon_{i_1})S^1_3K^{i_2}_3\\ &+g(\varepsilon_1-\varepsilon_{i_1})(-S^1_+K^{i_2}_-+S^1_-K^{i_2}_+)\}. \end{aligned} \tag{3.7}$$

(Note that the last term is non-Hermitian if T_{i_1}, T_{i_2} are Hermitian representations.)

Taking the limits $\mu\to\infty$ in the rational case and $\mu\to i\infty$ in the trigonometric one, we derive from (2.6)

$$[T(\lambda),N_i]=0,\quad i=1,2,3 \tag{3.8}$$

in the rational case and

$$[T(\lambda),N_3]=0 \tag{3.9}$$

in the trigonometric case, where we denoted "total spin operators"

$$\begin{aligned} N_1&=\sum_{i_1}S^{i_1}_1+\sum_{i_2}iK^{i_2}_1,\\ N_2&=\sum_{i_1}S^{i_1}_2+\sum_{i_2}iK^{i_2}_2,\\ N_3&=\sum_{i_1}S^{i_1}_3+\sum_{i_2}K^{i_2}_3. \end{aligned} \tag{3.10}$$

In this same way from (2.7) (taking limits in λ),

$$[N_3,B(\mu)]=B(\mu). \tag{3.11}$$

Now we shall specify representations T_{i_1}, T_{i_2} and take some appropriate limits.

A. Four-boson interaction

1. Let $M_1=2$, $M_2=0$ and we construct the familiar boson realizations of su(2) in the Fock space of four kinds of bosons

$$\begin{array}{ll} \text{site 1:} & \text{site 2:}\\ S^1_3=\tfrac12(a_1^+a_1-a_2^+a_2), & S^2_3=\tfrac12(a_3^+a_3-a_4^+a_4),\\ S^1_+=a_1^+a_2, & S^2_+=a_3^+a_4,\\ S^1_-=a_2^+a_1, & S^2_-=a_4^+a_3. \end{array} \tag{3.12}$$

We also have

$$(S^1)^2=\frac{n_2}{2}\Big(\frac{n_2}{2}+1\Big),\quad (S^2)^2=\frac{n_4}{2}\Big(\frac{n_4}{2}+1\Big), \tag{3.13}$$

with non-negative integers n_2, n_4 being the eigenvalues of Casimir operators

$$n_2=a_1^+a_1+a_2^+a_2,\quad n_4=a_3^+a_3+a_4^+a_4.$$

The Hamiltonian (3.7) in the trigonometric case (with $\varepsilon_1=0$, $\varepsilon_2=-\pi/2$ for simplicity) takes the form

$$H=a_1^+a_4^+a_2a_3+a_2^+a_3^+a_1a_4. \tag{3.14}$$

For fixed numbers n_2, n_4 we obtain for the pseudovacuum the Fock state

$$|0\rangle=|0,n_2,0,n_4\rangle\,. \tag{3.15}$$

We have, as a simple consequence of (2.10)–(2.12), (3.5), (3.6), and (3.11),

$$B(\lambda)=(a_1^+a_2)/\sin\lambda+(a_3^+a_4)/\cos\lambda, \tag{3.16}$$

$$N_3|\lambda_1,...,\lambda_K\rangle=(K-(n_2/2)-(n_4/2))|\lambda_i,...,\lambda_K\rangle\,, \tag{3.17}$$

$$h=n_2\sum_i\cot\lambda_i=n_4\sum_i\tan\lambda_i \tag{3.18}$$

$$\varphi_i=n_2\cot\lambda_i-n_4\tan\lambda_i-2\sum_{j\neq i}\cot(\lambda_i-\lambda_j)=0. \tag{3.19}$$

The second equality in (3.18) follows from $\Sigma\varphi_i=0$. According to (3.16), $K\leqslant n_2+n_4$. Further we note that the addition of free Hamiltonian $\Sigma_i\omega_ia_i^+a_i$, $\omega_1+\omega_4=\omega_2+\omega_3$ (which is a linear function of N_3, n_2, and n_4) leads only to an additional shift in h. We can also add some other fourth-order terms taking second-order polynomials in N_3, n_2, n_4 or taking ε_1, ε_2 such that $f(\varepsilon_1-\varepsilon_2)\neq0$. The corresponding modifications in (3.16)–(3.19) are obvious.

2. Let $M_1=1$, $M_2=1$, and

$$\begin{array}{ll} S_3=\tfrac12(a_1^+a_1-a_2^+a_2), & K_3=\tfrac12(a_3^+a_3+a_4^+a_4+1),\\ S_+=a_1^+a_2, & K_+=a_3^+a_4^+,\\ S_-=a_2^+a_1, & K_-=a_3a_4,\\ S^2=\dfrac{n_2}{2}\Big(\dfrac{n_2}{2}+1\Big), & k=\tfrac12(1+|n_0|),\quad n_0=a_3^+a_3-a_4^+a_4,\\ n_2=a_1^+a_1+a_2^+a_2, & K_3^2-\tfrac12(K_+K_-+K_-K_+)=k(k-1). \end{array} \tag{3.20}$$

Now fixing the integers $n_2\geqslant0$, n_0 we can write for the Hermitian Hamiltonian (again with $\varepsilon_1=0$, $\varepsilon_2=-\pi/2$)

$$\tilde H=iH_1=i(a_2^+a_3^+a_4^+a_1-a_1^+a_2a_3a_4) \tag{3.21}$$

(free Hamiltonian $\Sigma_i\omega_ia_i^+a_i$, $\omega_1=\omega_2+\omega_3+\omega_4$ is again a linear function of N_3, n_2, n_0). So we have for the eigenstates and eigenvalues of $\tilde H$ following formulae,

$$|0\rangle = |0,n_2,n_0,0\rangle \quad \text{for } n_0 \geqslant 0,$$
$$|0\rangle = |0,n_2,0,|n_0|\rangle \quad \text{for } n_0 < 0, \tag{3.22}$$

$$B(\lambda) = (a_1^+ a_2)/\sin\lambda + (a_3^+ a_4^+)/\cos\lambda, \tag{3.23}$$

$$N_3|\lambda_1,\dots,\lambda_K\rangle = (K + \tfrac{1}{2}(|n_0| + 1 - n_2))|\lambda_1,\dots,\lambda_K\rangle, \tag{3.24}$$

$$\tilde{h} = in_2 \sum_i \cot\lambda_i = -i(|n_0| + 1)\sum_i \tan\lambda_i, \tag{3.25}$$

$$\varphi_i = n_2 \cot\lambda_i + (|n_0| + 1)\tan\lambda_i - 2\sum_{j\neq i}\cot(\lambda_i - \lambda_j) = 0. \tag{3.26}$$

B. The Dicke model, three-boson interaction, second harmonics generation

1. Here we describe a limiting case[5] of the trigonometric Gaudin's model ($M_2 = 0$), which leads to the Dicke model. First we note that in (2.2) the L and r matrices can be multiplied by an arbitrary common factor. Now let $S_1 \to \infty$ in such a manner that

$$\sqrt{2S_1}\cot(\varepsilon_1 - \varepsilon_i) \to \tilde{\varepsilon}_i, \quad i \neq 1, \tag{3.27}$$

hence

$$\sin(\varepsilon_1 - \varepsilon_i) \to 1. \tag{3.28}$$

We denote $\sqrt{2S_1}\cot(\lambda - \varepsilon_1) \to -E$ in this limit. So we have

$$\frac{1}{\sqrt{2S_1}} L(\lambda) \to \tilde{L}(E) = \begin{pmatrix} -\dfrac{E}{2} - \sum \dfrac{S_3^i}{E - \tilde{\varepsilon}_i} & a^+ + \sum \dfrac{S_+^i}{E - \tilde{\varepsilon}_i} \\ a + \sum \dfrac{S_-^i}{E - \tilde{\varepsilon}_i} & \dfrac{E}{2} + \sum \dfrac{S_3^i}{E - \tilde{\varepsilon}_i} \end{pmatrix}, \tag{3.29}$$

$$\frac{1}{\sqrt{2S_1}} r(\lambda - \mu) \to \frac{1}{E - E'} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.30}$$

$$\tilde{H} = -\operatorname*{res}_{E=\infty} \frac{1}{2}\mathrm{Tr}_0 \tilde{L}^2(E) = \sum_i (aS_+^i + a^+ S_-^i + \tilde{\varepsilon}_i S_3^i). \tag{3.31}$$

From N_3 (3.10) we find (after appropriate shift renormalization) the integral of motion

$$N = a^+ a + \sum_i (S_3^i + S^i). \tag{3.32}$$

Pseudovacuum is now a state $|0\rangle$ such that

$$a|0\rangle = 0, \quad S_-^i|0\rangle = 0, \quad S_3^i|0\rangle = -S^i|0\rangle, \tag{3.33}$$

and by straightforward calculation we obtain

$$\tilde{h} = -\sum_i \tilde{\varepsilon}_i S^i + \sum_\alpha E_\alpha, \tag{3.34}$$

$$\varphi_\alpha = -E_\alpha + 2\sum_i \frac{S^i}{E - \tilde{\varepsilon}_i} - 2\sum_{\beta\neq\alpha}\frac{1}{E_\alpha - E_\beta} = 0, \tag{3.35}$$

$$N|E_1,\dots,E_K\rangle = K|E_1,\dots,E_K\rangle. \tag{3.36}$$

2. If we assume only one spin S and take its boson representation (3.12) (and $\tilde{\varepsilon} = 0$) we get from $\tilde{H}$ just the three-boson interaction Hamiltonian (now again the free Hamiltonian can be added as a linear function of S and N). So we can write ($a \to a_3$)

$$H = a_1^+ a_2 a_3 + a_3^+ a_2^+ a_1, \tag{3.37}$$

$$B(E) = a_3^+ + (a_1^+ a_2)/E, \tag{3.38}$$

$$|0\rangle = |0,n_2,0\rangle, \tag{3.39}$$

$$N = a_3^+ a_3 + \tfrac{1}{2}(a_1^+ a_1 - a_2^+ a_2 + n_2),$$
$$N|E_1,\dots,E_K\rangle = K|E_1,\dots,E_K\rangle, \tag{3.40}$$

$$h = \sum_\alpha E_\alpha, \tag{3.41}$$

$$\varphi_\alpha = -E_\alpha + \frac{n_2}{E_\alpha} - 2\sum_{\beta\neq\alpha}\frac{1}{E_\alpha - E_\beta} = 0. \tag{3.42}$$

The same limit can be adapted in the case $M_1 = 0$ and we obtain a Hamiltonian for the interaction of one-boson mode with su(1,1)-atoms. Then taking one atom ($\tilde{\varepsilon} = 0$) and boson representation (3.20) of su(1,1) we get another form of Bethe ansatz for the three-boson interaction (3.37) ($n_0 = a_2^+ a_2 - a_3^+ a_3$),

$$B(E) = a_1^+ + (a_2^+ a_3^+)/E, \tag{3.43}$$

$$|0\rangle = |0,n_0,0\rangle \quad \text{for } n_0 \geqslant 0,$$
$$|0\rangle = |0,0,|n_0|\rangle \quad \text{for } n_0 < 0, \tag{3.44}$$

$$N = a_1^+ a_1 + \tfrac{1}{2}(a_2^+ a_2 + a_3^+ a_3 - n_0),$$
$$N|E_1,\dots,E_K\rangle = K|E_1,\dots,E_K\rangle, \tag{3.45}$$

$$h = \sum_\alpha E_\alpha, \tag{3.46}$$

$$\varphi_\alpha = E_\alpha - \frac{1 + |n_0|}{E_\alpha} - 2\sum_{\beta\neq\alpha}\frac{1}{E_\alpha - E_\beta} = 0. \tag{3.47}$$

3. If we take another boson realization of su(1,1)

$$K_+ = \frac{a_2^{+2}}{2}, \quad K_- = \frac{a_2^2}{2}, \quad K_3 = \frac{a_2^+ a_2 + a_2 a_2^+}{4}, \tag{3.48}$$

with $k = \frac{1}{4}$ and $k = \frac{3}{4}$, we have the Hamiltonian for the second harmonics generation

$$H = a_1^+ a_2^2 + a_1 a_2^{+2}. \tag{3.49}$$

Now we have to take the algebraic Bethe ansatz with two pseudovacua $|0,0\rangle$ and $|0,1\rangle$ for $k = 1/4$ and $k = 3/4$, respectively. After rescaling $E \to E/2$ we can write

$$B(E) = a_1^+ + a_2^{+2}/E, \tag{3.50}$$

$$N = a_1^+ a_1 + (a_2^+ a_2)/2$$
$$= k + \tfrac{1}{4}, \quad N|E_1,\dots,E_K\rangle = K|E_1,\dots,E_K\rangle, \tag{3.51}$$

$$h = \sum_\alpha E_\alpha, \tag{3.52}$$

$$\varphi_\alpha = \frac{1}{2}E_\alpha - \frac{4k}{E_\alpha} - 4\sum_{\beta\neq\alpha}\frac{1}{E_\alpha - E_\beta}. \tag{3.53}$$

So we can see that the three-boson interaction and the second harmonics generation can be considered as special cases of one type of interaction of one-boson mode with su(1,1) atom if different boson representations of su(1,1) algebra are taken. The formulae (3.43) and (3.50) give a quantum analog of reduction $a_2 = a_3$ for classical solutions.

IV. ADDITIONAL REMARKS

The system of equations (2.12) in the trigonometric case (3.5) after substitution,[5]

$$z_i = \exp(2i\lambda_i), \quad \vartheta_i = \exp(2i\varepsilon_i), \tag{4.1}$$

gives

$$\frac{-\Sigma S^{i_1} + \Sigma k^{i_2} + M - 1}{z_i} + 2\sum_{i_1} \frac{S^{i_1}}{z_i - \vartheta_{i_1}} - 2\sum_{i_2} \frac{k^{i_2}}{z_i - \vartheta_{i_2}} - 2\sum_{j \neq i} \frac{1}{z_i - z_j} = 0. \tag{4.2}$$

So the completeness of the set of Bethe eigenvectors can be related to counting the number of distinct solutions of (4.2). In the pure su(2) case ($M_2 = 0$) the completeness follows from Gaudin's results.[5] In a pure su(1,1) case, completeness is a consequence of the Heine–Stieltjes theorem.[11] All quantum optical models described in Sec. III except one (3.21) are just of the above mentioned type (or its limiting cases). So the set of Bethe eigenvectors in these cases is complete. We conjecture that this is also true in the remaining case (3.21).

We note also that to all our models the functional Bethe ansatz (as developed in Refs. 2 and 12) can be applied.

The form of algebraic Bethe ansatz adapted in Sec. II can be obtained also as a limiting case[6] from quantum inverse scattering method for models associated with the 4×4 quantum R-matrix.[10] But it is an advantage of this form that it can be directly generalized to other Lie algebras as shown in Ref. 3.

What remains as an unsolved problem in quantum optics and what is of great interest from an experimental point of view is the time evolution of correlation functions with the initial coherent or squeezed state.[7] In general, only short time solutions are known. It seems to us that the systematic exploitation of integrability can lead to further development in this field.

ACKNOWLEDGMENT

I would like to thank J. Tolar for several useful discussions.

[1]M. Gaudin, Note CEA **1559**, 1 (1972); CEN-Saclay, **1559**, 2 (1973).
[2]E. K. Sklyanin, Zap. Nauchn. Sem. LOMI **164**, 151 (1987).
[3]B. Jurčo, J. Math. Phys. **30**, 1289 (1989).
[4]P. P. Kulish and E. K. Sklyanin, Lect. Notes Phys. **15**, 61 (1982).
[5]M. Gaudin, J. Phys. (Paris) **37**, 1087 (1976).
[6]M. Gaudin, *La Fonction d'Onde de Bethe* (Commissariat à l'Energie Atomique, Paris, 1983).
[7]J. Peřina, *Quantum Statistics of Linear and Nonlinear Optical Phenomena* (Reidel, Dordrecht, 1984).
[8]S. Stenholm, Phys. Rep. **60**, 1 (1973).
[9]H. P. Yuen and J. H. Shapiro, Opt. Lett. **4**, 334 (1979).
[10]V. E. Korepin, Commun. Math. Phys. **86**, 391 (1982).
[11]G. Szegö, *Orthogonal Polynomials* (Am. Math. Soc., New York, 1959).
[12]E. K. Sklyanin, Lect. Notes Phys. **226**, 196 (1985).

Gauge transformations for the quadratic bundle

Y. S. Vaklev
Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Boul. Lenin 72, Sofia 1784, Bulgaria

(Received 22 July 1988; accepted for publication 15 March 1989)

Gauge transformations constructed by means of the Jost solutions related to two similar cases of the quadratic bundle of general form are studied. Those transformations are taken at arbitrary fixed point $\lambda = \lambda_0$ of the continuous spectrum of the problem. The entire class of gauge-equivalent nonlinear evolution equations (NLEE) and their Hamiltonian structures are obtained. An interesting case of reduction of the potential is examined. It was shown that the entire class of NLEE splits into three different (in terms of coefficient functions) subclasses of gauge-equivalent NLEE. The gauge equivalence between the derivative nonlinear Schrödinger equation and some modifications of the derivative Landau–Lifshitz equation (DLLE) is demonstrated. The simplest soliton solutions of the DLLE and its higher analogs are obtained.

I. INTRODUCTION

The idea of gauge equivalence of certain nonlinear evolution equations (NLEE) was originally introduced in the pioneer papers.[1] In these papers the covariant form of the Lax compatibility condition

$$[U,V]_- + U_t - V_x = 0, \tag{1.1}$$

under gauge transformations,

$$U \to \mathrm{Ad}_g(U - g_x g^{-1}), \quad V \to \mathrm{Ad}_g(V - g_t g^{-1}), \tag{1.2}$$

was used. The operator U was given by the Zakharov–Shabat system and g was taken to be a Jost solution of the problem at the point $\lambda = 0$.

The lax approach, however, allows one to examine only one NLEE at a time. In Ref. 2 a way was found to obtain and examine the entire class of NLEE related to the Zakharov–Shabat system. That class was generated by an integrodifferential operator Λ. This operator took a very important place in that approach. A deeper understanding of its spectral decomposition has led to the so-called "expansion over 'squared solutions' method" (EOSSM).[3] An explicitly gauge-covariant formulation of the results of Ref. 3 was proposed in Ref. 4. This gauge-covariant formulation was extended to include some natural generalizations of the Zakharov–Shabat system[5] and the corresponding discrete case.[6]

The EOSSM was further developed and applied to the quadratic bundle of general form.[7–9] Here $L(q_0,\lambda q_1,\lambda^2)v = 0$ [see Eq. (1.1) below]. The corresponding class of NLEE in this case is larger than the previous one and includes the derivative nonlinear Schrödinger equation (DNSE), the relativistic Mikhailov model (which was shown to be equivalent to the massive Thirring model, etc. A covariant formulation of those results is given in Ref. 10; there, however, only transformations (1.2) with g being the Jost solution $L(q_0,\lambda_0 q_1,\lambda_0^2)g = 0$ at the point $\lambda_0 = 0$ were used. This fact obviously simplifies the matter, since one of the potentials q_1 does not contribute to g.

In the present paper we study a more general class of gauge transformations. We take g (1.2) to be a Jost solution of the quadratic bundle $L(p_0,\lambda_0,p_1,\lambda_0^2)g = 0$, where p_0 and p_1 could be different from q_0 and q_1 [see (2.1) below] and λ_0 is an arbitrary fixed point from the continuous spectrum of the problem. (We would like to note that when $q_1 = 0$ the quadratic bundle reduces to the Zakharov–Shabat system. There is no need, in this case, to study gauge transformations at points $\lambda_0 \neq 0$ of the spectrum separately, because the results can be derived by the simple change $\lambda^2 \to \lambda^2 + \lambda_0^2$. This is not true if $q_1 \neq 0$ and this case requires examination.) By means of an appropriate covariant formulation of the EOSSM we examine in a uniform manner the entire class of NLEE related to the quadratic bundle. The investigations are carried out on two levels.

The first level is to find in a covariant form the gauge equivalent to the initial[8,10] class of NLEE and its Hamiltonian structures. These results are written in terms of the transformed Λ operator and they are in some sense formal if we need their explicit form in terms of coefficient functions. This form depends on the choice of the used gauge transformation and thus the examined class splits into different (in terms of coefficient functions) subclasses of gauge-equivalent NLEE. (As an illustration we give the covariant formulation[4] of the class of NLEE containing the gauge-equivalent[1] nonlinear Schrödinger equation and the Heisenberg ferromagnet equation.)

The second level corresponds to the classification of those subclasses of NLEE. One can see that the subclasses of NLEE obtained by a gauge transformation taken in $\lambda_0 = 0$ and $\lambda_0 \neq 0$ points, are, in general, different.

Some interesting cases of reductions of the quadratic bundles are examined. Imposing $q_0 = i\omega\sigma_3 q_1$ we obtain three different subclasses of gauge-equivalent NLEE. A large part of these NLEE can be given either by a transformation at a point $\lambda_0 \neq 0$ or at $\lambda_0 = 0$, $\omega \neq 0$ as well. That extends the area of validity of the results of Ref. 10. Some examples of appropriate particular choices of the dispersion law are given. They form a subclass of NLEE and some of them have been examined in Refs. 8, 11, and 12.

The simplest soliton solutions for the gauge-transformed case with the examined reductions imposed are obtained.

 0022-2488/89/081744-12$02.50

II. PRELIMINARIES

Before we go to the basic problem of the paper, let us give some necessary results of the EOSSM for the quadratic bundle of the general form[7-10]

$$\{2D_x + q_0(x,t) + \lambda q_1(x,t) + r(x,t) - \lambda^2\}v(x,t,\lambda) = 0,$$

$$q_\alpha(x,t) = \begin{pmatrix} 0, & q_\alpha^+ \\ q_\alpha^-, & 0 \end{pmatrix}(x,t),$$

$$\lim_{|x| \to \infty} q_\alpha^\pm(x,t) = 0, \quad \alpha = 0,1, \tag{2.1}$$

$$r(x,t) = -\frac{1}{2}\langle q_1, q_1\rangle, \quad D_x := \frac{i}{2}\sigma_3 \frac{d}{dx}.$$

In the present paper we shall use the standard notation for the Pauli matrices σ_μ, $\mu = 0,...,3$, and for

$$\langle X,Y\rangle := \tfrac{1}{2}\,\mathrm{tr}\, XY, X,Y \in \mathrm{gl}\,(2,C).$$

The dependence on t in (2.1) is introduced as an exterior parameter, as usual. Following the notations of Refs. 8 and 10 we define the Jost solutions, the transition matrix, and the scattering data for (2.1) by

$$\lim_{x \to \infty} \psi(x,t,\lambda)\exp(i\lambda^2 x\sigma_3)$$

$$= \lim_{x \to -\infty} \varphi(x,t,\lambda)\exp(i\lambda^2 x\sigma_3) = \sigma_0,$$

$$T(\lambda,t) = \begin{pmatrix} a^+, & -b^- \\ b^+, & a^- \end{pmatrix}(\lambda,t) = \psi^{-1}(x,t,\lambda)\varphi(x,t,\lambda),$$

$$I_m \lambda^2 = 0, \quad \det T = 1, \tag{2.2}$$

$$\rho^\pm(\lambda,t) = \frac{b^\pm(\lambda,t)}{a^\pm(\lambda,t)}, \quad C_j^\pm := \frac{b_j^\pm}{\dot{a}_j^\pm},$$

$$b_j^\pm := b^\pm(\lambda_j^\pm,t), \quad \dot{a}_j^\pm := \frac{d}{d\lambda}a^\pm(\lambda)|_{\lambda = \lambda_j^\pm}.$$

The generating operator Λ, related to (2.1), is a central point of the EOSSM and can be defined by[8-10]

$$\Lambda := \begin{pmatrix} z_{10}, & 1 + z_1 \\ \Lambda_0 + z_0, & z_{01} \end{pmatrix}, \quad \Lambda_0 := D_x + r, \quad z_\alpha := z_{\alpha\alpha},$$

$$z_{\alpha\beta} := \frac{i}{2} q_\alpha(x,t)\left(\int_x^\infty dy + \int_x^{-\infty} dy\right)\langle q_\beta(y,t),[\sigma_3,\cdot]_-\rangle,$$

$$\alpha,\beta = 0,1, \tag{2.3}$$

$$[X,Y]_\pm := XY + YX, \quad X,Y \in \mathrm{gl}(n,c).$$

Let the scattering data (2.2) satisfy the conditions

$$\rho_t^\pm(\lambda,t) = \mp iF(\lambda)\rho^\pm(\lambda,t),$$

$$C_{j,t}^\pm = \mp iF(\lambda_j^\pm)C_j^\pm, \quad \frac{d\lambda_j^\pm}{dt} = 0, \tag{2.4}$$

$$F(\lambda) := \sum_{n=-K}^{l} f_n \lambda^n, \quad f_n = \mathrm{const},$$

where $F(\lambda)$ is the dispersion law associated with the problem (2.1) (Hereafter we shall denote the derivatives by subscripts.) One of the main results of EOSSM is the equivalence of the class of linear equations (2.4) and the class of NLEE,

$$i\sum \mathbf{q}_t(x,t) + F(\Lambda)\mathbf{q}(x,t) = 0,$$

$$\mathbf{q}(x,t) = \begin{pmatrix} q_1 \\ q_0 \end{pmatrix}(x,t), \quad \sum := \sigma_0 \times \sigma_3, \tag{2.5}$$

related to (2.1). By "$\times$" we denote the familiar Kronecker product

$$(X \times Y)_{ij,kl} := X_{ik}Y_{jl}, \quad i,j,k,l = 1,...,n, \quad X,Y \in \mathrm{gl}(n,c).$$

Any of the NLEE of the class (2.5) has an infinite series of conservation quantities,[8]

$$A_0 = -\frac{1}{2}\ln\frac{a^+(0)}{a^-(0)} = \frac{i}{2}\int_\Gamma \frac{d\lambda}{\lambda}\pi(\lambda) - \sum_{j=1}^{N}\ln\frac{\lambda_j^+}{\lambda_j^-},$$

$$A_p = \frac{i}{2}\int_\Gamma \frac{d\lambda}{\lambda}\lambda^p\pi(\lambda) - \frac{1}{p}\sum_{j=1}^{N}\{(\lambda_j^+)^p - (\lambda_j^-)^p\}$$

$$= -\frac{1}{p}\left\{2\int_{-\infty}^{\infty} dx\left(\int_x^\infty dy + \int_x^{-\infty} dy\right)\langle \mathbf{q}^R\Sigma_1\Lambda^{p+1}\mathbf{q}\rangle\right.$$

$$\left. + \left[\frac{i}{2}\begin{pmatrix} 0 \\ q_1 \end{pmatrix}\Sigma_1\Lambda^p\mathbf{q}\right]\right\},$$

$$\Sigma_1 := \sigma_1 \times \sigma_3, \quad p = \pm 1, \pm 2,...,$$

$$\delta A_p = -\frac{i}{2}[\![\Sigma\delta\mathbf{q}, \Lambda^{p-1}\mathbf{q}]\!], \quad p = 0, \pm 1,..., \tag{2.6}$$

generated by the functional

$$A(\lambda) = \ln a^+(\lambda), \quad \mathrm{Im}\,\lambda^2 > 0,$$

$$A(\lambda) = -\ln a^-(\lambda), \quad \mathrm{Im}\,\lambda^2 < 0,$$

$$A(\lambda) = \tfrac{1}{2}\ln[a^+(\lambda)/a^-(\lambda)] \quad \mathrm{Im}\,\lambda^2 = 0, \tag{2.7}$$

$$A(\lambda) = \sum_{k=1}^{\infty}\lambda^{-k}A_k \quad |\lambda| \gg 1,$$

$$A(\lambda) = -\sum_{k=0}^{\infty}\lambda^k A_{-k}, \quad |\lambda| \ll 1.$$

The integration contour Γ in the complex plane $\mathbb{C}$ (see Fig. 1) coincides with the continuous spectrum of (2.1) and the simple roots $\lambda_j^\pm \in \Gamma$, $a^\pm(\lambda_j^\pm) = 0$ give the discrete spec-

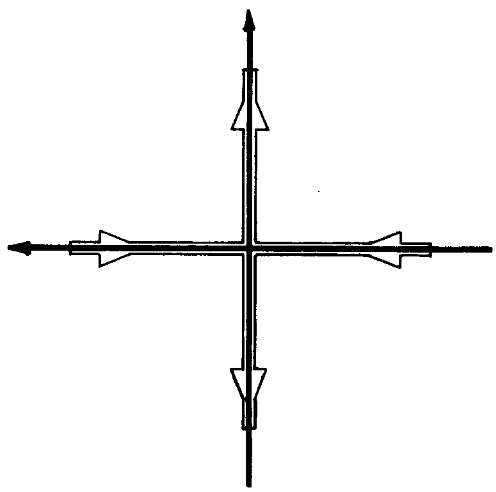


FIG. 1. The contour Γ.

trum of the problem (2.1). The action-angle variables π and κ are defined by

$$\pi(\lambda,t) := (1/\pi)\ln(1+\rho^+\rho^-), \quad \kappa(\lambda,t) := \tfrac{1}{2} b^+/b^-,$$
$$\pi_j^{\pm} := \pm 2\lambda_j^{\pm}, \quad \kappa_j^{\pm} := \pm i \ln b_j^{\pm}. \tag{2.8}$$

In a space $\mathscr{T}$, defined as

$$\mathscr{T} \ni X(x,t) := \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}(x,t),$$
$$X_i \in \mathrm{gl}\,(2,c), \ \lim_{|x|\to\infty} X_i = 0, \quad i = 1,2,$$

we define the automorphism $\underline{\mathrm{Ad}}_M$, R, D, and A operations and the skew-scalar product $[\ ,\ ]$, as follows [see (2.6)]:

$$\underline{\mathrm{Ad}}_M X := \begin{pmatrix} \mathrm{Ad}_M X_1 \\ \mathrm{Ad}_M X_2 \end{pmatrix}, M \in \mathrm{gl}(2,c), \quad X^R := (X_1, X_2),$$
$$X_D := \tfrac{1}{2}(X + \underline{\mathrm{Ad}}_{\sigma_3} X), \quad X_A := \tfrac{1}{2}(X - \underline{\mathrm{Ad}}_{\sigma_3} X), \tag{2.9}$$
$$[X,Y] := \int_{-\infty}^{\infty} dx \,\mathrm{tr}(X_A^R \Sigma_1 Y_A),$$

where $\mathrm{Ad}_A B$ is the familiar notation $A^{-1}BA, A, B \in \mathrm{gl}(n,C)$. The generating operator Λ (2.3) is self-conjugated with respect to the skew-scalar product (2.9),

$$[\Lambda X, Y] = [X, \Lambda Y], \quad X, Y \in \mathscr{T}. \tag{2.10}$$

The Hamiltonian structures, generating the class (2.5), are given by

$$\Omega^{(m)} := \frac{i}{2}\int_\Gamma d\lambda\, \lambda^{(m)}\, \delta\pi(\lambda)\wedge\delta\kappa(\lambda) + \frac{i}{2}\sum_{\substack{j=1\\ \epsilon=\pm}}^{N} \delta\pi_j^\epsilon \wedge \delta\kappa_j^\epsilon$$
$$= \frac{i}{4}[\Sigma\delta\mathbf{q} \overset{\wedge}{,} \Lambda^m \Sigma\delta\mathbf{q}],$$
$$H_F^{(m)} := i \sum_{p=-k}^{l} f_p A_{p+m+1} \tag{2.11}$$
$$= -\frac{1}{2}\int_\Gamma d\lambda\, F(\lambda)\lambda^m \pi(\lambda)$$
$$- i \sum_{j=1}^{N} \int_{\lambda_j^-}^{\lambda_j^+} d\mu\, \mu^m F(\mu).$$

Hereafter the notation $\wedge$ denotes the familiar exterior product of the corresponding quantities. The definition (2.11) justifies the name "action-angle variables" for the quantities given in (2.8).

We would not dwell, in detail, on EOSSM as a generalized Fourier transformation, etc. We would only like to note that with the help of the Jost solutions (2.2), the so-called "extended squared solutions[8,10] can be introduced. The extended squared solutions are eigenfunctions of the Λ operator (2.3). It turns out that the potential $\mathbf{q}$ (2.5) can be expanded over them and these expansions have all the properties of the familiar Fourier expansion of a function, with coefficients of the scattering data $\rho^{\pm}$ and $C_j^{\pm}$ (2.2).

One of the ways to solve the inverse scattering problem for (2.1) was proposed in Ref. 8. Using the analytic properties of the Jost solutions, the following representations for $\underline{\psi}(x,\lambda)$ are obtained:

$$\underline{\psi}(x,\lambda) := (\psi^-, \psi^+)(x,\lambda)\exp i\lambda^2 x\sigma_3,$$
$$\lim_{x\to\infty} \underline{\psi}(x,\lambda) = \sigma_0,$$
$$\underline{\psi}^+_{(x,\lambda)} = e^+ + \sum_{j=1}^{N} \frac{\zeta_j^-}{\lambda_j^- - \lambda}\underline{\psi}_j^-(x)$$
$$+ \frac{1}{2\pi i}\int_\Gamma \frac{d\mu}{\mu-\lambda}\rho^-(\mu)\underline{\psi}^-(x,\mu)\exp - 2i\mu^2 x,$$
$$\mathrm{Im}\,\lambda^2 > 0,$$
$$\underline{\psi}(x,\lambda) = e^- - \sum_{j=1}^{N} \frac{\zeta_j^+}{\lambda_j^+ - \lambda}\underline{\psi}_j^+(x)$$
$$+ \frac{1}{2\pi i}\int_\Gamma \frac{d\mu}{\mu-\lambda}\rho^+(\mu)\underline{\psi}^+(x,\mu)\exp 2i\mu^2 x,$$
$$\mathrm{Im}\,\lambda^2 < 0. \tag{2.12}$$

For the reflectionless ($\rho^{\pm} = 0$) case, (2.12) is reduced to an algebraic system of equations for any N—in particular, for $N = 1$ we get

$$\phi(x,\lambda)\exp i\lambda^2 x\sigma_3$$
$$= \sigma_0 + \alpha_-\sigma_3\{\zeta + (\lambda_+ - \lambda_-)\sigma\}(l - \lambda)^{-1}. \tag{2.13}$$

In (2.12) and (2.13) we used the notations [see (2.2)]

$$\zeta_j^+ := C_j^{\pm}\exp \pm 2i(\lambda_j^{\pm})^2 x, \quad \zeta^{\pm} := \zeta_1^{\pm},$$
$$\lambda^{\pm} := \lambda_1^{\pm}, \quad \zeta := \zeta^+\zeta^-,$$
$$\alpha_{\pm} := (\lambda_+ - \lambda_-)/[(\lambda_+ - \lambda_-)^2 \pm \zeta]$$
$$l := \begin{pmatrix} \lambda_+, 0 \\ 0, \lambda_- \end{pmatrix}, \quad \sigma := \begin{pmatrix} 0, \zeta^- \\ \zeta^+ 0 \end{pmatrix},$$
$$e^+ := \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad e^- := \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \underline{\psi}_j^{\pm}(x) := \underline{\psi}^{\pm}(x,\lambda_j^{\pm}). \tag{2.14}$$

Then using the expansion of (2.12) for $|\lambda| \gg 1$,

$$\underline{\psi}^d(x,\lambda) := \frac{1}{2}(\underline{\psi} + \mathrm{Ad}_{\sigma_3}\underline{\psi})(x,\lambda)$$
$$= \sigma_0 - \frac{i}{\lambda}\sigma_3\int_x^\infty dy\langle q_1, q_0\rangle + O\Big(\frac{1}{\lambda^2}\Big),$$
$$\underline{\psi}^a(x,\lambda) := \frac{1}{2}(\underline{\psi} - \mathrm{Ad}_{\sigma_3}\underline{\psi})(x,\lambda)$$
$$= \frac{1}{2\lambda}q_1 + \frac{1}{2\lambda^2}(q_0 + i\sigma_3 q_1)\int_x^\infty dy\langle q_1, q_0\rangle$$
$$+ O\Big(\frac{1}{\lambda^3}\Big), \tag{2.15}$$

the potentials q_α (2.1),

$$\begin{aligned} q_1 &= -2\alpha_-(\lambda_+ - \lambda_-)\sigma_3\sigma, \\ q_0 &= 2\alpha_-(\lambda_+ - \lambda_-)\sigma(\alpha_-\zeta + \sigma_3 l), \end{aligned} \tag{2.16}$$

were reproduced from (2.13).

This concludes the brief review of the EOSSM and the Zakharov–Shabat dressing method[13] relevant for the application of the inverse scattering method to the problem (2.1).

III. GAUGE-TRANSFORMED CASE

The aim of this section is to obtain and examine the class of NLEE related to an auxiliary linear problem gauge equivalent to (2.1). We shall give the infinite series of conservation quantities related to that class of NLEE as well as the Hamiltonian structures generating it.

Let us consider the auxiliary linear problem:

$$\begin{aligned}&(2D_x + p_0 + \lambda p_1 + r_1 - \lambda^2)W = 0, \quad p_\alpha := K_\alpha q_\alpha,\\ &K_{\alpha,x} = [\sigma_3, K_\alpha]_- = 0, \ \beta_\alpha := \det K_\alpha \neq 0,\\ &\alpha = 0,1, \quad r_1 := \beta_1 r,\end{aligned} \tag{3.1}$$

besides (2.1). The gauge transformations g used are constructed by the Jost solutions u of (3.1) at arbitrary fixed point λ_0 of Γ (see Fig. 1),

$$\lim_{x\to\infty} u(x,\lambda)\exp i\lambda^2 x\sigma_3 = \sigma_0, \quad g := u(x,\lambda_0). \tag{3.2}$$

By that means we obtain that the auxiliary linear problem

$$\begin{aligned}\Big\{&iS\frac{d}{dx} + \frac{i}{2}(1 - B_0^{-1})S_x + (\lambda B_1^{-1} - \lambda_0 B_0^{-1})S'\\ &- (\beta_1 - 1)r + \lambda_0^2 - \lambda^2\Big\} v_g = 0,\end{aligned} \tag{3.3}$$

is gauge equivalent to (2.1), where we denote

$$\begin{aligned}&S(x,t) := \mathrm{Ad}_g\,\sigma_3, \quad S'(x,t) := B_1 \mathrm{Ad}_g\, q_1,\\ &v_g := g^{-1} v C_v, \quad C_{v,x} = [\sigma_3, C_v]_- = 0,\\ &B_\alpha = \mathrm{Ad}_g\, K_\alpha = \langle K_\alpha\rangle\sigma_0 + \langle K_\alpha, \sigma_3\rangle S.\end{aligned} \tag{3.4}$$

We impose on S the natural requirement for the ferromagnets $\lim_{x\to\infty} S = \sigma_3$, which leads to $\rho_1^\pm(\lambda_0) = 0$ for the reflection coefficients (2.2) related to the quadratic bundle (3.1). From (3.2) and (3.4), it directly follows that

$$\mathrm{Ad}_g(p_0 + \lambda_0 p_1) = -(i/2)S_x. \tag{3.5}$$

Before we go straight to the point, we shall make a brief comparison with the well-known Lax approach. If we define the operators U [see (1.1)] by (2.1) and V, so that

$$Uv = v_x, \quad Vv = v_t, \tag{3.6}$$

the familiar Lax condition (1.1) for them is a NLEE from the class (2.5). Thus we have that (2.1) and (1.1) give us the class (2.5) equation by equation. As we mentioned (1.1) has a gauge-covariant form. This fact and (2.3) allow one to obtain the gauge equivalent to (1.1) NLEE—see (1.2). The problem reduces to the explicit expression of the transformed V(1.2) in terms of S and S' (3.4); which, in general, is a different problem for a different choice of V (3.6). This approach was applied in the particular case $q_0 = 0$, $r = 0$ in Ref. 11. In this way the gauge equivalence of DNSE,[7] the derivative Landau–Lifshitz equation (DLLE), and the so-called "higher-order NS-type systems" [8,11] was shown. It is interesting to do that in the general case (not only for $q_0 = 0$, $r = 0$) and for other possible choices of V. This is the aim of this section. To be more precise, we shall express the transformed by (3.2) class (2.5), in terms of S, S', and their derivatives. A more detailed exposition may be found in Appendix A; here we give the final results only. We shall use the gauge transformations at the arbitrary fixed point λ_0 of the spectrum Γ—Fig. 1, mentioned above. In the particular case $\lambda_0 = 0$, $B_0 = B_1 = \sigma_0$, the results reduce to those of Ref. 10. In this case p_1 does not take part in the equation for g [see (3.1)], which simplifies the matter. We shall use the following notations [compare with (2.3); see (3.1), (3.4), and (3.5)]:

$$\begin{aligned}&z := -\frac{i}{8\kappa_0^2} S_x\Big(\int_x^\infty dy + \int_x^{-\infty} dy\Big)\langle S_y, [S,\cdot]_-\rangle,\\ &\tilde z_1 := \frac{i}{2} S'\Big(\int_x^\infty dy + \int_x^{-\infty} dy\Big)\langle S', [S,\cdot]_-\rangle,\\ &\tilde z_{10} := \frac{1}{4} S'\Big(\int_x^\infty dy + \int_x^{-\infty} dy\Big)\langle S_y, [S,\cdot]_-\rangle,\\ &\tilde z_{01} := \frac{1}{4} S_x\Big(\int_x^\infty dy + \int_x^{-\infty} dy\Big)\langle S', [S,\cdot]_-\rangle,\\ &\mathbb{S}^\delta := \begin{pmatrix} -\delta S\\ \tfrac12[S,\delta S']_-\end{pmatrix}, \quad \mathbb{S} := \begin{pmatrix} -\tfrac12[\sigma_3,S]_-\\ \langle\sigma_3,S\rangle S'\end{pmatrix},\\ &K := \begin{pmatrix} K_1, & 0\\ 0, & K_2\end{pmatrix},\\ &B := \mathrm{Ad}_{\sigma_0\times g} K, \quad J := \begin{pmatrix} \sigma_0, & 0\\ \lambda_0\sigma_0, & \sigma_0\end{pmatrix},\\ &\tilde\Lambda_0 := \frac{i}{4}\Big[S, \frac{d}{dx}\cdot\Big]_-, \quad \kappa_0^2 := \lambda_0^2 + \omega^2.\end{aligned} \tag{3.7}$$

For our purposes it will be necessary to give the generating operator M for the problem (3.1) [compare with (2.3)],

$$\begin{aligned}&M = \begin{pmatrix} {}^p z_{10}, & 1 + {}^p z_1\\ M_0 + {}^p z_0, & {}^p z_{01}\end{pmatrix},\\ &M_0 := D_x + r_1, \quad {}^p z_\alpha := {}^p z_{\alpha\alpha},\\ &{}^p z_{\alpha\beta} := \frac{i}{2} p_\alpha\Big(\int_x^\infty dy + \int_x^{-\infty} dy\Big)\langle p_\beta, [\sigma_3,\cdot]_-\rangle,\\ &\alpha,\beta = 0,1.\end{aligned} \tag{3.8}$$

After that we derive the gauge-transformed operator (3.8) and by means of it, using the notations (3.7), we define

$$\begin{aligned}\tilde M &:= \mathrm{Ad}_{\sigma_0\times g}\{K^{-1}(M - \lambda_0)J\}\\ &= B^{-1}J^{-1}\begin{pmatrix} \tilde z_{10}, & 1 + \tilde z_1\\ \tilde\Lambda_0 + \kappa_0^2 z, & \tilde z_{01}\end{pmatrix}.\end{aligned} \tag{3.9}$$

Now we can give the gauge-transformed potential (2.5) and its variation [see (3.7) and (3.9)],

$$\begin{aligned}&\underline{\mathrm{Ad}}_g\mathbf{q} = \tilde M\mathbb{S},\\ &\underline{\mathrm{Ad}}_g\,\Sigma\delta\mathbf{q} = \tilde M\{\mathbb{S}^\delta - \delta A''(\lambda_0)\mathbb{S}\}.\end{aligned} \tag{3.10}$$

The conservation quantity $A''(\lambda_0)$ in (3.10) is given by (2.7) for the problem (3.1) by means of the diagonal elements of the matrix $T_1(\lambda_0) = g^{-1}h$. The Jost solution h of (3.1) has an asymptotics $\lim_{x\to-\infty} h \exp i\lambda_0^2 x\sigma_3 = \sigma_0$. Because of (3.9) and (3.10) we obtain that the class [see (3.7) and (3.8)]

$$\begin{aligned}&i\tilde M\mathbb{S}^t + F(\tilde\Lambda)\tilde M\mathbb{S} = 0 \leftrightarrow i\mathbb{S}^t + F(\Lambda_g)\mathbb{S} = 0,\\ &\tilde\Lambda := \mathrm{Ad}_{\sigma_0\times g}\,\Lambda, \quad \Lambda_g := \mathrm{Ad}_{\tilde M}\,\tilde\Lambda,\end{aligned} \tag{3.11}$$

where [compare with (3.7)]

$$\mathbb{S}^t = \begin{pmatrix} -S_t\\ \tfrac12[S,S'_t]_-\end{pmatrix},$$

is gauge equivalent to (2.5). In spite of its covariant form [compare with (2.5)], (3.11) consists of different (in terms of coefficient functions) kinds of NLEE. In the next section we shall obtain, in a more explicit form, those NLEE imposing some reductions on the potential $\mathfrak{q}$ (2.5).

We can give the transition matrix Tg by analogy with (2.2) by means of the Jost solutions,

$$\lim_{x\to\infty} \psi_g(x,t,\lambda)\exp i(\lambda^2-\lambda_0^2)x\sigma_3$$
$$= \lim_{x\to-\infty} \varphi_g(x,t,\lambda)\exp i(\lambda^2-\lambda_0^2)x\sigma_3 = \sigma_0,$$

of (3.3). One can verify that [see (2.2), (2.6), (2.8), and (3.4)]

$$\begin{aligned} T_g(\lambda,t) &:= \psi_g^{-1}(x,t,\lambda)\varphi_g(x,t,\lambda) \\ &= T(\lambda,t)T_1^{-1}(\lambda_0), \quad \operatorname{Im}\lambda^2 = 0, \\ \rho_g^{\pm}(\lambda,t) &= \rho^{\pm}(\lambda,t), \quad \pi_g(\lambda,t) = \pi(\lambda,t), \\ \kappa_g(\lambda,t) &= \kappa(\lambda,t) - A''(\lambda_0), \\ T_1(\lambda_0) &= \operatorname{diag}(a_1^+,a_1^-)(\lambda_0), \quad a_1^+(\lambda_0)a_1^-(\lambda_0) = 1. \end{aligned} \tag{3.12}$$

Let us define [see (3.4); compare with (2.9)] $\widetilde{D}$ and $\widetilde{A}$ operations,

$$X_{\widetilde{D}} := \tfrac{1}{2}(X + \underline{\mathrm{Ad}}_S X),$$
$$X_{\widetilde{A}} := \tfrac{1}{2}(X - \underline{\mathrm{Ad}}_S X), \quad X\in\mathscr{T},$$

and the skew-scalar product $[\![\ ,\ ]\!]$,

$$[\![X,Y]\!] := \int_{-\infty}^{\infty} dx\, tr(X_{\widetilde{A}}^R\widetilde{\Sigma}_1 Y_{\widetilde{A}}), \quad X,Y\in\mathscr{T}. \tag{3.13}$$

Here and further on we shall use the notations

$$\widetilde{\Sigma} := \sigma_0\times S, \quad \widetilde{\Sigma}_1 := \sigma_1\times S.$$

It is easy to check that [see (2.9), (2.10), and (3.7)]

$$\begin{aligned} &[\![\underline{\mathrm{Ad}}_g X,\underline{\mathrm{Ad}}_g Y]\!] = [\![X,Y]\!], \\ &[\![\widetilde{M}\Lambda_g X,\widetilde{M}Y]\!] = [\![\widetilde{M}X,\widetilde{M}\Lambda_g Y]\!]. \end{aligned} \tag{3.14}$$

All that allows one to give the infinite series of conservation quantities (2.6) in the form [see (2.7) and (3.12)]

$$\begin{aligned} A_0 &= \frac{i}{2}\int\frac{d\lambda}{\lambda}\pi(\lambda) - \sum_{j=1}^{N}\ln\frac{\lambda_j^+}{\lambda_j^-}, \\ A_p &= \frac{i}{2}\int_\Gamma\frac{d\lambda}{\lambda}\lambda^p\pi(\lambda) - \frac{1}{p}\sum_{j=1}^{N}\{(\lambda_j^+)^p - (\lambda_j^-)^p\} \\ &= -\frac{1}{p}\int_{-\infty}^{\infty}dx\Big(\int_x^{\infty}dy + \int_x^{-\infty}dy\Big) \\ &\quad\times\operatorname{tr}\{(\widetilde{M}\mathrm{S})^R\widetilde{\Sigma}_1\widetilde{M}\Lambda_g^{p+1}\mathrm{S}\} \\ &\quad-\frac{i}{2p}\left[\!\!\left[\begin{pmatrix}0\\ B_1^{-1}S'\end{pmatrix}, \widetilde{\Sigma}\widetilde{M}\Lambda_g^p\mathrm{S}\right]\!\!\right]_{\sim}, \quad p=\pm1\pm2,\ldots, \\ \delta A_p &= -(i/2)[\![\widetilde{M}\mathrm{S}^\delta,\widetilde{M}\Lambda_g^{p-1}\mathrm{S}]\!], \quad p = 0,\pm1,\ldots, \end{aligned} \tag{3.15}$$

related to the class (3.11). After that we can give the Hamiltonian structures

$$\begin{aligned} \Omega_g^{(m)} &:= (i/4)[\![\widetilde{M}\mathrm{S}^\delta,\widetilde{M}\Lambda_g^{m-2}\mathrm{S}^\delta]\!], \\ H_{gF}^{(m)} &:= i\sum_{p=-k}^{l} f_p A_{p+m-1}, \end{aligned} \tag{3.16}$$

generating the class (3.11). The comparison with (2.11) gives that

$$\Omega^{(m)} = \Omega_g^{(m+2)} + \delta A_{m+1}\wedge\delta A''(\lambda_0), \quad H_F^{(m)} = H_{gF}^{(m+2)}.$$

To obtain that and (3.16) we have used (2.11), (3.10), (3.14), and (3.15). Letting $B=E_4$ and $\lambda_0 = 0$ in (3.16), one obtains [see (3.9) and (3.11)]

$$\widetilde{M} = \Lambda_{g_0} = \widetilde{\Lambda}, \quad \Omega_{g_0}^{(m)} = (i/4)[\![\mathrm{S}_0^\delta,\widetilde{\Lambda}^m\mathrm{S}_0^\delta]\!].$$

That coincides with the expression given in Ref. 10 and explains the meaning of m as an exponent in (3.16).

Coming to the end of the section, we would like to give (for the one soliton solution) the potentials S, S'—(3.4) of the gauge equivalent to (2.1) problem—(3.3). With the help of (2.13), (2.16), (3.1), and (3.4), we derive

$$\begin{aligned} S &= (1-2m)\sigma_3 - (2\alpha_-/\gamma)(\lambda_+-\lambda_-)(\exp 2i\lambda_0^2x\sigma_3) \\ &\quad\times\{1 - [(\alpha_-\zeta)/\gamma]\sigma_3(l-\lambda_0)\}(l-\lambda_0)\sigma, \\ S' &= -2\alpha_-(\lambda_+-\lambda_-)\sigma_3\{m/\alpha_+ + (\exp 2i\lambda_0^2x\sigma_3)S''\sigma\}, \end{aligned} \tag{3.17}$$

where $\lambda_\pm$ are from the discrete spectrum of the problem (3.1), σ, $\alpha_\pm$, l, and ζ are given by (2.14) and γ, m, and S'' are

$$\begin{aligned} \gamma &:= (\lambda_+-\lambda_0)(\lambda_--\lambda_0), \\ m &:= \alpha_-^2\,\zeta(\lambda_+-\lambda_-)^2/\gamma, \\ S'' &:= \sigma_0 - [(2\alpha_-\zeta)/\gamma]\sigma_3(l-\lambda_0) \\ &\quad + [\alpha_-^2\,\zeta(\lambda_+-\lambda_-)/\alpha_+\gamma^2](l-\lambda_0)^2. \end{aligned}$$

IV. THE REDUCTION CASE

In this section we shall use the previous results imposing some restrictions on the potentials q_α, p_α—(2.1) and (3.1), respectively. Let us impose

$$p_0' = i\omega\sigma_3 p_1', \quad q_0' = i\epsilon\omega\sigma_3 q_1',$$
$$k_0 = \epsilon k_1, \quad \epsilon = \pm1, \tag{4.1a}$$
$$p_\alpha' := (\exp 2i\omega^2x\sigma_3)p_\alpha, \quad q_\alpha' := (\exp 2i\omega^2x\sigma_3)q_\alpha,$$
$$\alpha = 0,1, \quad w\in\Gamma. \tag{4.1b}$$

In this case (2.1) and (3.1) reduce to

$$\{2D_x + (\lambda + i\epsilon\omega\sigma_3)q_1' + r - \lambda^2\}v' = 0 \tag{4.2a}$$
$$\{2D_x + (\lambda + i\omega\sigma_3)p_1' + r_1 - \lambda^2\}W' = 0. \tag{4.2b}$$

One can see (compare with Refs. 10 and 14) that the linear problems

$$\begin{aligned} &(2D_x + \lambda p_1 + r_1 - \lambda^2)w_0 = 0, \\ &\{2D_x + (\mu + i\omega\sigma_3)p_1'' + r_1 - \mu^2\}w_1 = 0, \\ &\mu := \sqrt{\lambda^2-\omega^2}, \\ &p_1'' := \{\exp 2(\eta + i\omega^2x)\sigma_3\}p_1, \quad \eta\in\mathbb{C}, \quad \eta_x = 0, \end{aligned} \tag{4.3}$$

are gauge equivalent because of the transformation

$$w_1 = \tau(\mu,\omega)\{\exp(\eta + i\omega^2x)\sigma_3\}w_0C_W,$$
$$C_{W,x} = [\sigma_3,C_W]_- = 0,$$

$\tau(\mu,\omega)$:

$$= \epsilon_1\begin{pmatrix} \sqrt{\epsilon\sqrt{\mu+i\omega}}/\sqrt[4]{\mu-i\omega}, & 0 \\ 0, & \epsilon\sqrt{\epsilon\sqrt{\mu-i\omega}}/\sqrt[4]{\mu+i\omega} \end{pmatrix},$$

$$\epsilon = \pm 1, \quad \epsilon_1 = \pm 1.$$

This transformation is singular at $\mu = 0, \pm i\omega$; that is why we shall examine the results at these points separately. The notation $\sqrt{\xi}$ used above means the inverse function of $\xi = \lambda^2, \lambda \in \mathbb{C}$, mapping the first Riemanian sheet on the upper half-plane of $\mathbb{C}$ without the ray cut $\operatorname{Im}\lambda = 0$, $\operatorname{Re}\lambda < 0$. All that is given for the scattering data in this case is

$$\rho_1^{\pm}(\mu) = \epsilon(\exp \mp 2\eta)(\sqrt{\mu \mp i\omega}/\sqrt{\mu \pm i\omega})\,\rho_0^{\pm}(\lambda),$$
$$\mu_j^{\pm} := \sqrt{(\lambda_j^{\pm})^2 - \omega^2} \tag{4.4}$$
$$C_{1j}^{\pm} = \epsilon(\exp \mp 2\eta)\sqrt{\mu_j^{+} \mp i\omega}/\sqrt{\mu_j^{\pm} \pm i\omega}.$$

From (4.4) it follows that the dispersion laws associated with the examined linear problems have to be even on λ and are related by

$$F_0(\lambda) = :\mathscr{F}(\lambda^2) = F_1(\sqrt{\lambda^2 - \omega^2}) = :\mathscr{F}_1(\lambda^2 - \omega^2). \tag{4.5}$$

The reduction (4.1) gives for Λ' (2.3),

$$\Lambda' = \begin{pmatrix} -i\epsilon\omega z'\sigma_3, & 1 + z' \\ \Lambda_0 + \omega^2\sigma_3 z'\sigma_3, & i\epsilon\omega\sigma_3 z' \end{pmatrix},$$
$$z' := \frac{i}{2} q_1'\left(\int_x^{\infty} dy + \int_x^{-\infty} dy\right)\langle q_1', [\sigma_3, \cdot\,]\rangle. \tag{4.6}$$

With the help of (4.5) and (4.6) from (2.5) (see Appendix B) one obtains that the class of NLEE,

$$i\sigma_3 q_{1,t}' + F(\Lambda_1')q_1' = 0, \quad \Lambda_1' := (1 + z')(\Lambda_0 + \omega^2), \tag{4.7}$$

generated by (2.1) in the case (4.1), decomposes into equivalent [up to change of variables (4.1a)] subclasses of NLEE corresponding to the different possible choices of ω.

In the case (4.1) we have two different Jost solutions of (4.2b),

$$(2D_x + r_1)g_0 = 0, \quad \kappa_0^2 = 0, \tag{4.8a}$$
$$\{2D_x + (\lambda_0 + i\omega\sigma_3)p_1' + r_1 - \lambda_0^2\}g_1 = 0, \quad \kappa_0^2 \neq 0, \tag{4.8b}$$

giving two ways to construct gauge transformations. In this way we obtain that

$$\{2D_x + (\lambda + i\epsilon\omega\sigma_3)\tilde{q}_1 - (\beta - 1)r - \lambda^2\}\tilde{u} = 0, \tag{4.9a}$$

$$\left\{ iS\frac{d}{dx} - i\frac{(\lambda + i\epsilon\omega s)(\lambda_0 - i\omega s)B_1^{-1} - \kappa_0^2}{2\kappa_0^2}S_x - (\beta_1 - 1)r + \lambda_0^2 - \lambda^2 \right\}\tilde{v} = 0, \quad k_0^2 \pm 0, \tag{4.9b}$$

are gauge equivalent to (4.2a), where [see (4.2a) and (4.8)]

$$\tilde{u} := g_0^{-1}v'c', \quad \tilde{v} := g_1^{-1}v'c, \quad C_x' = C_x = [\sigma_3, C']_- = 0 \tag{4.10}$$

and [compare with (2.4)]

$$\tilde{q}_1 := \left(\exp 2i\beta\sigma_3 \int_x^{\infty} dy\, r\right)q_1', \quad S = \mathrm{Ad}_{g_1}\sigma_3. \tag{4.11}$$

It is easy to see that if $\beta = 0$, (4.9a) coincides with (4.2a). If $\beta \neq 0$, one obtains the gauge equivalent to (4.2a) linear problem (4.9a) obtained by a gauge transformation constructed by (4.8a) with $\beta_1 = \beta$.

In order to obtain the class of gauge equivalent to (4.7) NLEE, we shall use the fact that in the case (4.1) from (3.5), (3.7), and (3.9), we obtain [see (4.11)]

$$\widetilde{M}_0 = K^{-1}\begin{pmatrix} 0, & 1 + \beta K_1 \tilde{z} K_1^{-1} \\ \widetilde{\Lambda}_0, & 0 \end{pmatrix},$$
$$\mathrm{S}_0 = \begin{pmatrix} 0 \\ K_1\tilde{q}_1 \end{pmatrix}, \quad \mathrm{S}_0^{\delta} = \begin{pmatrix} 0 \\ \sigma_3 K_1 \delta\tilde{q} \end{pmatrix},$$
$$S' = -\frac{i}{2\kappa_0^2}(\lambda_0 - i\omega S)S_x,$$
$$\tilde{z} := \frac{i}{2}\tilde{q}_1\left(\int_x^{\infty} dy + \int_x^{-\infty} dy\right)\langle \tilde{q}_1, [\sigma_3, \cdot\,]_-\rangle,$$
$$\mathrm{S} = -\frac{1}{2}\begin{pmatrix} [\sigma_3, S]_- \\ \frac{i}{\kappa_0^2}(\lambda_0 - i\omega S) < \sigma_3, \quad S > S_x \end{pmatrix}, \tag{4.12}$$
$$\widetilde{M} = B^{-1}J^{-1}\begin{pmatrix} (\lambda_0 - i\omega S)z, & 1 + \frac{1}{\kappa_0^2}(\lambda_0 - i\omega s)z(\lambda_0 + i\omega s) \\ \widetilde{\Lambda}_0 + \kappa_0^2 z, & z(\lambda_0 + i\omega s) \end{pmatrix}$$
$$\mathrm{S}^{\delta} = -\begin{pmatrix} \delta S \\ \frac{1}{\kappa_0^2}(\lambda_0 - i\omega S)\widetilde{\Lambda}_0\delta S \end{pmatrix}.$$

Thus from (3.10), (4.11), and (4.12) [see (3.7) and (4.8)] we derive

$$\mathrm{Ad}_{g_{00}}\sigma_3\delta q_1' = (1 + \beta\tilde{z})\sigma_3\delta\tilde{q}_1 + \beta\delta A_0^0\tilde{q}_1,$$
$$g_{00} := \exp - i\beta\sigma_3\int_x^{\infty} dy\, r,$$
$$(\Lambda_\kappa - \kappa_0^2)[\sigma_3, S]_- = iS_x,$$
$$\Lambda_\kappa := (1 + z)(\widetilde{\Lambda}_0 + \kappa_0^2), \tag{4.13}$$
$$\mathrm{Ad}_{g_1}\sigma_3\delta q_1' = -\frac{1}{\kappa_0^2}B_1^{-1}(\lambda_0 - i\omega S)(\Lambda_\kappa - \kappa_0^2) \times\{\delta S - \tfrac{1}{2}\delta A''(\lambda_0)[\sigma_3, S]_-\},$$

where A_0^0 is the lowest conserved quantity for (4.2a) at $q_1' = 0$. With the help of (4.12) and (4.13), one obtains that the classes of NLEE [see (3.10) and Appendix B]

$$i\sigma_3\tilde{q}_{1,t} + \{1 - (\beta - 1)\tilde{z}\}\mathscr{F}(\widetilde{\Lambda}_1)\tilde{q}_1 = 0, \tag{4.14}$$
$$\widetilde{\Lambda}_1 := (\Lambda_0 - \beta r + \omega^2)(1 + \tilde{z}),$$
$$iS_t + \tfrac{1}{2}\mathscr{F}(\Lambda_{\kappa g})S = 0, \tag{4.15}$$
$$\Lambda_{\kappa g} := \mathrm{Ad}_{(\Lambda_\kappa - \kappa_0^2)}\widetilde{\Lambda}_\kappa,$$
$$\widetilde{\Lambda}_\kappa := [1 + (1/\beta_1)z]\{\widetilde{\Lambda}_0 - (\beta_1 - 1)r + \kappa_0^2\}, \quad \kappa_0^2 \neq 0,$$

are gauge equivalent to (4.7). The subclass (4.14) may be considered to be obtained from (4.7) by the change of variables (4.11), since in (4.15) [see (4.10)] the unknown Jost solution g_1 (4.8b) takes part. In addition, we would like to note that (4.15) can be obtained from (4.7) by means of g_1 (4.8b), either at $\lambda_0 = 0$, $\omega \neq 0$, at $\lambda_0 \neq 0, \omega = 0$, or at $\lambda_0 \neq 0, \omega \neq 0$ (λ_0 and ω lie on Γ—Fig. 1). The subclass (4.15) splits into two different (in terms of coefficient functions) subclasses given by $\beta=\beta_1=1$ and $\beta,\beta_1\neq 1$. The same splitting goes for (4.14), but the difference could be compensated by the change of the variables (4.10). In the case $\beta=\beta_1=1$, (4.14) and (4.15) can be considered to be obtained from (4.7) by a gauge transformation constructed by the corresponding Jost solutions of (4.2a) [compare with (4.8)]. As a corollary in this particular case ($\beta=\beta_1=1$) we have that the transformations constructed by the Jost solutions of (4.2a)

$$(2D_x+\lambda_0 q_1+r-\lambda_0^2)g_{01}=0,$$
$$(2D_x+i\omega\sigma_3 q_1+r)g_{02}=0$$

lead to (4.15) by $\kappa_0^2=\lambda_0^2$ or $\kappa_0^2=\omega^2$, as follows. This extends the domain of application of the results of Ref. 10. (In spite of the covariant form (3.11), (4.14) and (4.15) are different in terms of coefficient functions. This could be easily verified, taking into account the action of $\widetilde{\Lambda}_1$ and $\widetilde{\Lambda}_\kappa$ on $\tilde{q}_1$ and $[\sigma_3,S]_-$, respectively. We do not account for the differences that could be compensated by explicit changes of the variables.)

In order to write down the conservation quantities (3.13) in the case (4.1), we denote [compare with (2.9)]

$$\left[X \overset{\wedge}{,} Y\right]:=\int_{-\infty}^{\infty}dx\langle X_{,\wedge}[\sigma_3,Y]_-\rangle,$$
$$\left[X \overset{\wedge}{,} Y\right]:=\int_{-\infty}^{\infty}dx\langle X_{,\wedge}[S,Y]_-\rangle, \qquad (4.16)$$
$$X,Y\in \mathrm{gl}(2,C).$$

In (4.16) we used the short notation $\left[u_1\overset{\wedge}{,}V_1\right]=\left[u_2\overset{\wedge}{,}V_2\right]$, which means that $[u_1,V_1]=[u_2,V_2]$ and $\left[u_1\overset{\wedge}{,}V_1\right]=\left[u\overset{\wedge}{,}V_2\right]$. Now we are able to give these conservation quantities, using (3.15), by means of (4.12) and (4.13) (see Appendix B),

$$\begin{aligned}
A'_{2n}&=-\frac{1}{n}\Big\{\int_{-\infty}^{\infty}dx\Big(\int_x^{\infty}dy+\int_x^{-\infty}dy\Big)\langle q'_1,\sigma_3(1-z')\Lambda'_1(\Lambda'_1-\omega^2)^n q'_1\rangle\\
&\quad+\frac{i}{4}[q'_1,\sigma_3(\Lambda'_1-\omega^2)^n q'_1]\Big\}\\
&=-\frac{1}{n}\Big\{\int_{-\infty}^{\infty}dx\Big(\int_x^{\infty}dy+\int_x^{-\infty}dy\Big)\langle\tilde{q}_1,\sigma_3\widetilde{\Lambda}_1(\widetilde{\Lambda}_1-\omega^2)^n\tilde{q}_1\rangle\\
&\quad+\frac{i}{4}[\tilde{q}_1,\sigma_3(1+\tilde{z})(\widetilde{\Lambda}_1-\omega^2)^n\tilde{q}_1]\Big\}\\
&=\frac{1}{4\beta_1 n\kappa_0^2}\Big\{\int_{-\infty}^{\infty}dx\Big(\int_x^{\infty}dy+\int_x^{-\infty}dy\Big)\Big\langle\delta y,S\Big(1-\frac{1}{\beta_1}z\Big)\widetilde{\Lambda}_\kappa(\widetilde{\Lambda}_\kappa-\omega^2)^n Sy\Big\rangle\\
&\quad+\frac{i}{4}[S_x,S(\widetilde{\Lambda}_\kappa-\omega^2)^n S_x]\Big\},\quad A'_{2n+1}=0,\quad n=\pm 1,\pm 2,...,\\
\delta A'_{2n}&=-(i/2)[\sigma_3\delta q'_1,(1-z')\Lambda'_1(\Lambda'_1-\omega^2)^{n-1}q_1]=-(i/2)[\sigma_3\delta\tilde{q}_1,(1+\beta_1\tilde{z})\widetilde{\Lambda}_1(\widetilde{\Lambda}_1-\omega^2)^{n-1}\tilde{q}_1]\\
&=(1/4\beta_1\kappa_0^2)[(\Lambda_\kappa-\kappa_0^2)\delta S,[1-1/\beta_1 z]\widetilde{\Lambda}_\kappa(\widetilde{\Lambda}_\kappa-\omega^2)^{n-1}S_x],\quad n=0,\pm 1\ldots.
\end{aligned}\qquad (4.17)$$

With the help of (4.13), (4.16), and (4.17) from (3.16) (see Appendix B), we derive that the Hamiltonian structures $\Omega^{(2n)}$, $H_{\mathscr{F}}^{(2n)}$, $\Omega^{(2n)}$, $\widetilde{H}_{\mathscr{F}}^{(2n)}$, $\Omega_g^{(2n)}$, and $H_{g\mathscr{F}}^{(2n)}$ are degenerated after the reduction (4.1). That is why (4.14) and (4.15) are generated by

$$\Omega^{(2n+1)}:=(i/4)\left[\sigma_3\delta q'_1 \overset{\wedge}{,} (1-z')\Lambda'_1(\Lambda'_1-\omega^2)^n\sigma_3\delta q'_1\right],$$
$$\widetilde{\Omega}^{(2n+1)}:=(i/4)\left[\sigma_3\delta\tilde{q}_1 \overset{\wedge}{,} (1+\beta\tilde{z})\widetilde{\Lambda}_1(\widetilde{\Lambda}_1-\omega^2)^n\{1+(\beta-1)\tilde{z}\}\sigma_3\delta\tilde{q}_1\right],$$
$$\Omega_g^{(2n+1)}:=\frac{1}{4\beta_1\kappa_0^2}\left[(\Lambda_\kappa-\kappa_0^2)\delta S \overset{\wedge}{,} [1-(1/\beta_1)z]\widetilde{\Lambda}_\kappa(\widetilde{\Lambda}_\kappa-\omega^2)^n(\Lambda_\kappa-\kappa_0^2)\delta S\right],\ \kappa_0^2\neq 0,$$
$$H'^{(2n+1)}_{\mathscr{F}}=\widetilde{H}^{(2n+1)}_{\mathscr{F}}=i\sum_{p=-[k/2]}^{[l/2]}f_{2p}A'_2(p+n+1),H^{(2n+1)}_{g\mathscr{F}}=i\sum_{p=-[k/2]}^{[l/2]}f_{2p}A'_{2(p+n-1)}\cdot$$

If $n\geqslant 0$ one can easily derive that

$$\Omega'^{(2n+1)}=\sum_{k=0}^{n}\binom{n}{k}(-\omega^2)^{n-k}\Omega^{(2k+1)}$$
$$=\sum_{k=0}^{n}\binom{n}{k}(-\omega^2)^{k}\Omega^{(2n-2k+1)},$$

where [see (2.3)]

$$\Omega^{(2n+1)}=\frac{i}{4}\left[\sigma_3\delta q_1 \overset{\wedge}{,} (1-z_1)\Lambda_1(\Lambda_1-\omega^2)^n\sigma_3\delta q_1\right],$$
$$\Lambda_1:=(1+z_1)\Lambda_0.$$

Coming to the end of the section we would like to note that in the case (4.1) the results (2.16) and (3.17), in general, are not compatible with the reduction (4.1). We have [see (2.15), (4.3), and (4.4)]

$$\langle q_0, q_1\rangle = 0, \quad \sigma_3 v_0(x,\lambda)\sigma_3 = v_0(x, -\lambda), \tag{4.18}$$

where $a_0^{\pm}(\lambda) = a_0^{\pm}(-\lambda)$ follows. That means that if λ_+ and λ_- are eigenvalues of (2.1), then so are $-\lambda_+$ and $-\lambda_-$, and they should also be included as poles in (2.12). In the case (4.1) with $\omega = 0$ we have [see (2.12), (2.15), and (4.3)]

$$\left(i\sigma_3 \frac{d}{dx} + \lambda q_1 + r\right)\underline{\psi}(x,\lambda) - 2\lambda^2 \underline{\psi}^a(x,\lambda) = 0. \tag{4.19}$$

The condition (4.18) leads to the expansion

$$\underline{\psi}(x,\lambda) = \sigma_0 + \sum_{n=1}^{\infty}\left\{\frac{1}{\lambda^{2n-1}}\underline{\psi}^a_{2n-1}(x) + \frac{1}{\lambda^{2n}}\underline{\psi}^d_{2n}(x)\right\}, \quad |\lambda| \gg 1,$$

of (2.12) at large $|\lambda|$, which we put in (4.19). Solving the recurrence we derive

$$\underline{\psi}^a_{2n-1} = \frac{1}{2}E^-(x)\{(1+N)D_x\}^{n-1}E^+(x)q_1,$$

$$\underline{\psi}^d_{2n} = -\frac{1}{2r}E^+(x)q_1 N D_x\{(1+N)D_x\}^{n-1}E^+(x)q_1,$$

$$E^{\pm}(x) := \exp \pm i\sigma_3\int_x^{\infty} dy\, r, \quad n = 1,2,...,$$

$$Nf(x) := \frac{i}{2}\sigma_3 E^+(x)q_1\int_x^{\infty} dy E^+(y)g_1 f(y).$$

Thus we have

$$\underline{\psi}(x,\lambda) = \sigma_0 + E^-(x)\sum_{n=0}^{\infty}\frac{1}{2\lambda^{2n+1}} \times\left\{1 - \frac{1}{\lambda r}(E^+(x))^2 q_1 N D_x\right\} \times\{(1+N)D_x\}^n E^+(x)q_1. \tag{4.20}$$

Because of (4.18), the representation (2.12) looks like

$$\underline{\psi}^{\pm}(x,\lambda) = e^{\pm} \pm 2\sum_{j=1}^{N}\frac{\zeta_j^{\mp}}{(\lambda_j^{\mp})^2 - \lambda^2} l_j^{\mp}(\lambda)\underline{\psi}_j^{\mp}(x),$$

where $e^{\pm}, \underline{\psi}_j^{\pm}(x)$, and $\zeta_j^{\pm}$ are defined by (2.14) and

$$l_j^+(\lambda) := \begin{pmatrix}\lambda_j^+, & 0\\ 0, & \lambda\end{pmatrix}, \quad l_j^-(\lambda) := \begin{pmatrix}\lambda, & 0\\ 0, & \lambda_j^-\end{pmatrix}.$$

In the simplest case $N = 1$ we have

$$\underline{\psi}(x,\lambda) = \sigma_0 - 2(\lambda_+^2 - \lambda_-^2)\sigma_3\mathscr{A}\left\{(\lambda_+^2 - \lambda_-^2)\sigma + \frac{2\zeta}{\lambda} l^2\right\}\sum_{n=0}^{\infty}\frac{1}{\lambda^{2n+1}} l^{2n}, \tag{4.21}$$

where ζ and σ are defined by (2.14) and

$$\mathscr{A} := \begin{pmatrix}\dfrac{1}{(\lambda_+^2 - \lambda_-^2)^2 - 4\zeta\lambda_+^2}, & 0\\ 0, & \dfrac{1}{(\lambda_+^2 - \lambda_-^2)^2 - 4\zeta\lambda_-^2}\end{pmatrix},$$

$$l := \begin{pmatrix}\lambda_+, & 0\\ 0, & \lambda_-\end{pmatrix}.$$

The results (4.20) and (4.21) give for the potential q_1

$$q_1(x,t) = -4(\lambda_+^2 - \lambda_-^2)^2\mathscr{A}\sigma_3\sigma. \tag{4.22}$$

The potential S (3.4) can be derived from (4.21),

$$S(x,t) = \sigma_3(1 + 2g^a g), \tag{4.23}$$

where

$$g := \underline{\psi}(x,\lambda_0)\exp - i\sigma_3\lambda_0^2 x,$$

$$\underline{\psi}(x,\lambda_0) = \sigma_0 + 2(\lambda_+^2 - \lambda_-^2)\mathscr{A}\sigma_3\{2\zeta l^2 + \lambda_0(\lambda_+^2 - \lambda_-^2)\sigma\}(l^2 - \lambda_0^2)^{-1},$$

and [compare with (2.15)]

$$g^a := \frac{1}{2}(g - \mathrm{Ad}_{\sigma_3} g) = 2\lambda_0(\lambda_+^2 - \lambda_-^2)\mathscr{A}\sigma_3\sigma(l^2 - \lambda_0^2)^{-1}\exp - i\sigma_3\lambda_0^2 x.$$

Thus we have both potentials q_1 of (4.2a) and S of (4.9) restored by the Jost solution (4.21). The soliton solutions (4.22) and (4.23) survive the additional reduction $q_1^\dagger = \epsilon q_1$, $\epsilon = \pm 1$, where "†" means the Hermitian conjugation; if $\lambda_+^* = \epsilon\lambda_-$ and $C_1^{*+} = -\epsilon C_1^-$ [see (2.2) and (2.14)].[7]

V. EXAMPLES

The examples studied in this section are of the case (4.1). One can obtain the different (in terms of coefficient functions) gauge-equivalent NLEEE from (4.14) and (4.15). As we mentioned in Sec. IV, we do not examine the differences that can be eliminated by explicit change of the variables.

Let us choose the dispersion law [see (4.5)]

$$\mathscr{F}(\lambda^2) = -4\lambda^4 + C_1\lambda^2 + C_2,$$

which gives us from (4.14) and (4.15),

$$i\sigma_3\tilde{q}_{1,t} + \tilde{q}_{1,xx} - i\beta(\langle\tilde{q}_1,\tilde{q}_1\rangle\sigma_3\tilde{q}_1)_x + i\nu_1\sigma_3\tilde{q}_{1,x} + \nu_2\tilde{q}_1 = 0, \tag{5.1}$$

$$S_t - (i/2)[S,S_{xx}]_- - (4\kappa_0^2 - C_1/2)S_x + (1/4\kappa_0^2)(S_x)^3 + (i/2)(4\kappa_0^4 - C_1\kappa_0^2 - C_2)[\sigma_3,S]_- = 0, \tag{5.2}$$

$$i[S,S_{xt}]_- + \frac{1}{2}[S,([S,S_{xx}]_-)_x]_- + \frac{i}{\kappa_0^2}\langle S_x,S_x\rangle SS_x + i\nu_3[S,S_{xx}]_- + \frac{(\beta_1 - 1)(2\beta_1 - 1)}{16\beta_1^2\kappa_0^4}(S_x)^5 + \frac{1}{\beta_1}(S_x)^3 + (\nu_4 + n)S_x = 0,$$

$$n_x = -2i\langle S,S_x S_t\rangle, \tag{5.3}$$

where

$$\nu_1 := -4\omega^2 - (\beta - 1)\langle \tilde{q}_1, \tilde{q}_1 \rangle + C_1/2,$$
$$\nu_3 := [(2\beta_1 - 1)/4\beta_1\kappa_0^2]\langle S_x, S_x\rangle - 4\kappa_0^2 + C_1/2,$$
$$\nu_2 := (\beta - 1)(2i\langle \tilde{q}_1, \sigma_3 \tilde{q}_1 x\rangle + [(2\beta - 1)/2]\langle \tilde{q}_1, \tilde{q}_1\rangle^2) - 2\omega^2\langle \tilde{q}_1, \tilde{q}_1\rangle - 4\omega^4 + C_1\omega^2 + C_2,$$
$$\nu_4 := \frac{i(\beta_1 - 1)}{\beta_1 \kappa_0^2}\langle S, S_x S_{xx}\rangle - 8\kappa_0^4 + 2C_1\kappa_0^2 + 2C_2.$$

The subclasses of NLEE (5.1), (5.2), and (5.3) conserve their gauge equivalence for any choice of β, $\beta_1 \in \mathbb{C}$, $\beta_1 \neq 0$, $\kappa_0^2 \neq 0$, and $\omega \in \Gamma$. What is more we can choose C_1 and C_2 (4.1) in a different way in each of (5.1), (5.2), and (5.3), preserving their equivalence. This is true because the change of variables $\tau = t$, $\xi = -(C_1/2)t + x$, $\tilde{q}_1 \to (\exp iC_1\omega^2\tau\sigma_3)\tilde{q}_1$ compensate C_1 and the change $\tilde{q}_1 \to (\exp iC_2 t\sigma_3)\tilde{q}_1$ compensate C_2 in these subclasses of NLEE. We use their gauge equivalence at a different choice of β, β_1, C_1, and C_2. The subclass (5.1) at $\omega = 0$, $\beta = 1$ is generated by the simplest, in some sense, Hamiltonian structure,

$$\tilde{\Omega}^{(-3)} = \frac{1}{2}\int_{-\infty}^{\infty} dx\left(\int_x^{\infty} dy + \int_x^{-\infty} dy\right)\langle \delta\tilde{q}_1(x) \overset{\wedge}{,} \delta\tilde{q}_1(y)\rangle,$$
$$\tilde{H}^{(-3)} = \frac{1}{2}\int_{-\infty}^{\infty} dx\{\langle \tilde{q}_1, \tilde{q}_1\rangle^2 + C_1\langle \tilde{q}_1, \tilde{q}_1\rangle + 2i\langle \tilde{q}_1, \sigma_3 \tilde{q}_{1,x}\rangle - \frac{iC_2}{2}\left(\int_x^{\infty} dy + \int_x^{-\infty} dy\right)\langle \tilde{q}_1(x), \sigma_3\tilde{q}_1(y)\rangle\}. \tag{5.4}$$

Its gauge-equivalent class, giving some modifications of DLLE (5.2), is generated by [for the sake of convenience we consider (5.2) to be obtained by (4.8b) when $\lambda_0 = 0$ and $\omega \neq 0$, i.e., $\kappa_0^2 = \omega^2$,]

$$\Omega_{\rm DLL} = \frac{i}{2}\int_{-\infty}^{\infty} dx\left(\langle \delta S \overset{\wedge}{,} S\delta S\rangle + \frac{i}{2\kappa_0^2}\langle \delta S \overset{\wedge}{,} \delta S_x\rangle\right),$$
$$H_{\rm DLL} = \int_{-\infty}^{\infty} dx\left\{\frac{i}{4\kappa_0^2}\langle S, S_x S_{xx}\rangle + \frac{1}{32\kappa_0^4}\langle S_x, S_x\rangle^2 - \frac{1}{2}\langle S_x, S_x\rangle + (4\kappa_0^4 - C_1\kappa_0^2 - C_2)\left(\langle \sigma_3, S\rangle + \frac{i}{4\kappa_0^2}\langle \sigma_3, SS_x\rangle\right)\right\} - 2i(4\kappa_0^2 - C_1/2)A_0^1.$$

Imposing

$$\tilde{q}_1 = \epsilon\tilde{q}_1^\dagger = \begin{pmatrix} 0, & \tilde{q} \\ \epsilon\tilde{q}^*, & 0\end{pmatrix}, \quad \epsilon = \pm 1, \tag{5.5}$$

and $C_1 = C_2 = 0$, we come upon the well-known DNSE,[7]

$$i\tilde{q}_t + \tilde{q}_{xx} - i\epsilon(|\tilde{q}|^2\tilde{q})_x = 0$$

for the $\tilde{q}$ variable and there (5.4) reduces to the given[7] Hamiltonian structure. In the case (5.5) from (4.8) ($\lambda_0 = 0$), we find that

$$(\tilde{g}_1^\dagger)^{-1} = g_1^d - \epsilon(\omega^*/\omega)g_1^a,$$

where for S (3.4), it follows that

$$S = \epsilon'\begin{pmatrix}\sqrt{1 + \epsilon|p|^2}, & p \\ -\epsilon p^*, & -\sqrt{1 + \epsilon|p|^2}\end{pmatrix},$$
$$\epsilon' = \pm 1, \quad \epsilon|p|^2 \geqslant -1$$

and (5.2) reduces to (compare with Ref. 10).

$$p_t + \epsilon'\left(\frac{\epsilon p \,{\rm Im}\, p^* p_x - ip_x}{\sqrt{1 + \epsilon|p|^2}}\right)_x - \frac{{\rm Im}^2 p^* p_x + \epsilon|p_x|^2}{4\kappa_0^2(1 + \epsilon|p|^2)}p_x + \frac{1}{2}(C_1 - 8\kappa_0^2)p_x + i(4\kappa_0^4 - C_1\kappa_0^2 - C_2)p = 0.$$

The simplest NLEE from (5.2) can be found choosing $C_1 = 8\kappa_0^2$,

$$C_2 = -4\kappa_0^4, \qquad S_t - (i/2)[S, S_{xx}]_- + \frac{1}{4\kappa_0^2}(S_x)^3 = 0. \tag{5.6}$$

It is easy to see that the choice $\kappa_0^2 = \alpha\lambda_0^2$, $C_1 = 0$, $C_2 = 4\alpha^2\lambda_0^4$ leads to the DLLE given in Ref. 11,

$$S_t - (i/2)[S, S_{xx}]_- + \frac{1}{4\alpha\kappa_0^2}(S_x)^3 - 4\alpha\kappa_0^2 S_x = 0.$$

Choosing C_1, C_2, ω, β, and β_1 in a different way, one can obtain from (5.1), (5.2), and (5.3) the whole class of NLEE equivalent to DNSE, (5.6), and DLLE, respectively. In this connection we would like to note two things. In Ref. 11 four diagrams showing the gauge equivalence of some NLEE[8–12] are given. As a consequence of the present paper the first two diagrams coincide. [It is not difficult to see that, for example

$${\rm GI}_1, \quad iQ_t + Q_{xx} + i\epsilon\alpha Q^2 Q_x^* + (\alpha^2/2)|Q|^4 Q = 0$$

from the second diagram is equivalent to

$${\rm GI}_2{:}\ iQ'_t + Q'_{xx} - 2\epsilon\omega^2\alpha|Q'|^2 Q' + (\alpha^2/2)|Q'|^4 Q' + i\epsilon\alpha Q'^2 Q'^*_x = 0$$

from the first diagram. One obtains GI_2 from GI_1 by the changes $Q \to (\exp 2i\omega^2 x)Q$ and

$$t = \tau, \quad \xi = 4\omega^2 t + x, \quad Q' = (\exp - 4i\omega^2\tau)Q. \tag{5.7}$$

From the other hand (5.1) at $\beta = \omega = C_1 = C_2 = 0$ and imposing

$$\tilde{q}_1 = \begin{pmatrix} 0, & Q \\ \epsilon\alpha Q^*, & 0\end{pmatrix}, \quad \alpha = \alpha^* \tag{5.8}$$

leads to GI_1. The choice $\beta = C_1 = C_2 = 0$. $\omega = $ fixed $\in \Gamma$, imposing (5.8) and (5.7) gives GI_2.]

The second remark concerns the linear problems

$$L_1 v_1 = (2D_x + \lambda\tilde{q}_1 - \lambda^2)v_1 = 0,$$
$$L_2 v_2 = \left(i\frac{d}{dx} + q_2 - \lambda^2\sigma_3\right)v_2 = 0,$$
$$\tilde{q}_1 := \begin{pmatrix} 0, & \tilde{q}^+ \\ \tilde{q}^-, & 0\end{pmatrix},$$
$$q_2 := \left(\exp - 2i\sigma_3\int_x^{\infty} dy\, r\right)\times\begin{pmatrix} 0, & \tilde{q}^+ \\ (i/2)\tilde{q}_x^- - \frac{1}{4}\tilde{q}^+(\tilde{q}^-)^2, & 0\end{pmatrix}$$

It is known (see, for example, Ref. 9 that L_1 and L_2 are connected by the Mikhailov transformation

$$\mathscr{M}\sigma_3 L_1 \mathscr{M}^{-1} = L_2,$$

$$\mathscr{M} := \left(\exp - i\sigma_3 \int_x^\infty dy\, r\right) \times \begin{pmatrix} \lambda^{-1/2}, & 0 \\ -\tilde{q}^-(4\lambda)^{-1/2}, & \lambda^{1/2} \end{pmatrix}.$$

Hence both problems are gauge equivalent. It is not true, however, to insist that the nonlinear Schrödinger equation is gauge equivalent to DNSE. (One can check that $\tilde{q}_1^\dagger = \epsilon \tilde{q}_1$ and $q_2^\dagger = \epsilon q_2$, $\epsilon = \pm 1$ are not compatible.) For that reason the other diagrams are to be separated.

ACKNOWLEDGMENTS

The author would like to express his gratitude to V. S. Gerdjikov for numerous discussions and constant support.

APPENDIX A

Let us denote

$$X^d := \tfrac{1}{2}(X + \mathrm{Ad}_{\sigma_3} X), \quad X^a := \tfrac{1}{2}(X - \mathrm{Ad}_{\sigma_3} X),$$

$$X \in \mathrm{gl}(2,C), \quad p_2 := p_0 + \lambda_0 p_1,$$

$$^p\mathbf{z}^\pm_{\alpha\beta} := ip_\alpha \int_x^{\pm\infty} dy \langle p_\beta, [\sigma_3, \cdot\,]_- \rangle,$$

$$\underline{\mathbf{z}}^\pm_{\alpha\beta} := i\,\mathrm{Ad}_g\, p_\alpha \int_x^{\pm\infty} dy \langle \mathrm{Ad}_g\, p_\beta, [S, \cdot\,]_- \rangle, \qquad \text{(A1)}$$

$$^p\mathbf{z}_{\alpha\beta} := \tfrac{1}{2}(^p\mathbf{z}^+_{\alpha\beta} + {}^p\mathbf{z}^-_{\alpha\beta}),$$

$$\underline{\mathbf{z}}_{\alpha\beta} := \tfrac{1}{2}(\underline{\mathbf{z}}^+_{\alpha\beta} + \underline{\mathbf{z}}^-_{\alpha\beta}), \quad \alpha,\beta = 0,1,2.$$

From (3.1) and (3.2) one can find that

$$\begin{aligned}(\delta g g^{-1})_x &= i\sigma_3 \delta p_2 - (i/2)\sigma_3 \langle p_1, [\sigma_3, \sigma_3 \delta p_1]_- \rangle \\ &\quad - ip_2 \langle (2\sigma_3 \delta g g^{-1})^d \rangle \\ &\quad - (i/2)\sigma_3 \langle p_2, [\sigma_3, (-2\sigma_3 \delta g g^{-1})^a]_- \rangle \\ &\quad - i(r_1 - \lambda_0^2)(-2\sigma_3 \delta g g^{-1})^a \end{aligned} \qquad \text{(A2)}$$

For the d part [see (A1)] of (A2) we derive

$$(\delta g g^{-1})^d = \frac{i}{2}\sigma_3 \int_x^{\pm\infty} dy (\langle p_1, [\sigma_3, \sigma_3 \delta p_1]_- \rangle + \langle p_2, [\sigma_3, (-2\sigma_3 \delta g g^{-1})^a]_- \rangle) + \sigma_3 \Delta_\pm, \qquad \text{(A3)}$$

where [see the remark following (3.4)]

$$\Delta_\pm := \lim_{x \to \pm\infty} \langle \sigma_3 \delta g g^{-1} \rangle, \quad \Delta_+ = 0,$$

$$\Delta_- = -\tfrac{1}{2}\delta \ln[a_1^+(\lambda_0)/a_1^-(\lambda_0)] = -\delta A''(\lambda_0). \qquad \text{(A4)}$$

For the a part [see (A1)] of (A2) by means of (A3), we have [see (3.8)]

$$(M_0 + {}^p\mathbf{z}_2^\pm - \lambda_0^2)(-2\sigma_3 \delta g g^{-1})^a + {}^p\mathbf{z}^\pm_{21} \sigma_3 \delta p_1 = \sigma_3 \delta p_2 - 2\Delta \pm p_2. \qquad \text{(A5)}$$

It is not difficult to verify that

$$\begin{aligned}\underline{\mathbf{z}}^\pm_{\alpha 1}\, \mathrm{Ad}_g\, \sigma_3 \delta p_1 &= i\,\mathrm{Ad}_g\, p_\alpha \int_x^{\pm\infty} dy\, \delta \langle p_1, p_1 \rangle \\ &= i\,\mathrm{Ad}_g\, p_\alpha \int_x^{\pm\infty} dy\, \delta \langle S', S' \rangle \\ &= i\,\mathrm{Ad}_g\, p_\alpha \int_x^{\pm\infty} dy \langle S', [S, S\delta S']_- \rangle \\ &= i\,\mathrm{Ad}_g\, p_\alpha \int_x^{\pm\infty} dy \langle S', [S, \tfrac{1}{2}[S, \delta S']_-]_- \rangle \\ &= \underline{\mathbf{z}}^\pm_{\alpha 1} S\delta S' = \underline{\mathbf{z}}^\pm_{\alpha 1} \tfrac{1}{2}[S, \delta S']_-, \end{aligned}$$

$$\delta S = \mathrm{Ad}_g [\sigma_3, \delta g g^{-1}]_- = \mathrm{Ad}_g (2\sigma_3 \delta g g^{-1})^a. \qquad \text{(A6)}$$

By means of (A3) and (A6) it is easy to calculate

$$\begin{aligned}&\mathrm{Ad}_g\, \sigma_3 \delta p_1 \\ &= \delta(\tfrac{1}{2}[S, S']_-) + [g^{-1}\delta g, SS']_- \\ &= \tfrac{1}{2}[\delta S, S']_- + \tfrac{1}{2}[S, \delta S']_- + \mathrm{Ad}_g [(\delta g g^{-1})^d, \sigma_3 p_1] \\ &\quad + \mathrm{Ad}_g [(\delta g g^{-1})^a, \sigma_3 p_1] \\ &= \tfrac{1}{2}[S, \delta S']_- + \mathrm{Ad}_g \{ {}^p\mathbf{z}_1^\pm \sigma_3 \delta p_1 \\ &\quad + {}^p\mathbf{z}^\pm_{12}(-2\sigma_3 \delta g g^{-1})^a + 2\Delta_\pm p_1 \} \\ &= (1 + \underline{\mathbf{z}}_1^\pm)\tfrac{1}{2}[S, \delta S']_- + \underline{\mathbf{z}}^\pm_{12}(-\delta S) + 2\Delta_\pm S'. \end{aligned} \qquad \text{(A7)}$$

The covariant derivative

$$\nabla_x := \frac{d}{dx} + [g^{-1}g_x, \cdot\,]_- = \frac{d}{dx} + \frac{1}{2}[SS_x, \cdot\,]_- + i(r_1 - \lambda_0^2)[S, \cdot\,]_-;$$

(A5) and (A6) allow one to obtain [see (3.7)] that

$$(\tilde{\Lambda}_0 + \underline{\mathbf{z}}_2^\pm)(-\delta S) + \underline{\mathbf{z}}^\pm_{21} \tfrac{1}{2}[S, \delta S']_- = \mathrm{Ad}_g\, \sigma_3 \delta p_2 - 2\Delta_\pm\, \mathrm{Ad}_g\, p_2.$$

This result (A4), (A7), (2.5), (3.7), and (3.8) give

$$\begin{aligned}\mathrm{Ad}_g \Sigma \begin{pmatrix} \delta p_1 \\ \delta p_2 \end{pmatrix} &= J\, \underline{\mathrm{Ad}}_g \Sigma \delta \mathbf{p} = \begin{pmatrix} \underline{\mathbf{z}}^\pm_{12}, & 1 + \underline{\mathbf{z}}_1^\pm \\ \tilde{\Lambda}_0 + \underline{\mathbf{z}}_2^\pm, & \underline{\mathbf{z}}^\pm_{21} \end{pmatrix} \mathbf{S}^\delta \\ &\quad + 2\Delta_\pm J\, \underline{\mathrm{Ad}}_g\, \mathbf{p} \\ &= \begin{pmatrix} \underline{\mathbf{z}}_{12}, & 1 + \underline{\mathbf{z}}_1 \\ \tilde{\Lambda}_0 + \underline{\mathbf{z}}_2, & \underline{\mathbf{z}}_{21} \end{pmatrix} \mathbf{S}^\delta \\ &\quad - \delta A''(\lambda_0) J\, \underline{\mathrm{Ad}}_g\, \mathbf{p}. \end{aligned} \qquad \text{(A8)}$$

With the help of (A1), (3.5), (3.7), and (3.9), one finds

$$\begin{pmatrix} \underline{\mathbf{z}}_{12}, & 1 + \underline{\mathbf{z}}_1 \\ \tilde{\Lambda}_0 + \underline{\mathbf{z}}_2, & \underline{\mathbf{z}}_{21} \end{pmatrix} = JB\tilde{M},$$

$$J\, \underline{\mathrm{Ad}}_g\, \mathbf{p} = \begin{pmatrix} S' \\ -(i/2)S_x \end{pmatrix}.$$

It is easy to verify that

$$\tilde{\Lambda}_0\left(-\frac{i}{2}[\sigma_3, S]_-\right) = \frac{1}{8} \frac{\langle S_x, [S, [\sigma_3, S_x]_-]_- \rangle}{\langle S_x, S_x \rangle} S_x = \tfrac{1}{2} \langle \sigma_3, S \rangle S_x. \qquad \text{(A9)}$$

In this way (A8), (A9), and (3.7) lead to (3.10).

APPENDIX B

With the help of (4.6), (4.8), (4.10), and (4.11), we derive

$$(\Lambda')^{-1}=\begin{pmatrix} -i\epsilon\omega\sigma_3(\Lambda_0+\omega^2 z')^{-1}z', & \sigma_3(\Lambda_0+\omega^2 z')^{-1}\sigma_3 \\ \omega^2 z'(\Lambda_0+\omega^2 z')^{-1}z'+1-z', & i\epsilon\omega z'(\Lambda_0+\omega^2 z')^{-1}\sigma_3 \end{pmatrix},$$

$$\widetilde{\Lambda}':=\operatorname*{Ad}_{\sigma_0\times g_0}\Lambda'=\begin{pmatrix} -i\epsilon\omega\tilde{z}\sigma_3, & 1+\tilde{z} \\ \Lambda_0-\beta_1 r+\omega^2\sigma_3\tilde{z}\sigma_3 & i\epsilon\omega\sigma_3\tilde{z}\end{pmatrix}, \tag{B1}$$

$$\widetilde{\widetilde{\Lambda}}':=\operatorname*{Ad}_{\sigma_0\times g}\Lambda'=\frac{1}{\kappa_0^2}\sigma_0\times B_1^{-1}(\lambda_0-i\omega S)$$

$$\times\begin{pmatrix} -\dfrac{i\epsilon\omega}{\beta_1}zS, & 1+\dfrac{1}{\beta_1}z \\ (\widetilde{\Lambda}_0-(\beta_1-1)r+\lambda_0^2+\dfrac{\omega^2}{\beta_1}SzS, & \dfrac{i\epsilon\omega}{\beta_1}Sz\end{pmatrix}\sigma_0\times B_1(\lambda_0+i\omega S).$$

Let us denote by X and $\widetilde{X}$ the elements of the space of 2×4 matrices,

$$X:=\begin{pmatrix} X_1 \\ i\epsilon\omega\sigma_3 X_1\end{pmatrix},\quad \widetilde{X}:=\begin{pmatrix} X_1 \\ i\epsilon\omega S X_1\end{pmatrix},$$

$$X_1(x)\in \mathrm{gl}(2,c),\ \lim_{x\to\pm\infty} X_1(x)=0.$$

From (4.6) and (B1) by induction we have [see (4.7), (4.13), and (4.14)]

$$(\Lambda')^{2n}X=\begin{pmatrix} (\Lambda_1'-\omega^2)^n X_1 \\ i\epsilon\omega\sigma_3(\Lambda_1'-\omega^2)^n X_1\end{pmatrix},\quad (\Lambda')^{2n+1}X=\begin{pmatrix} i\epsilon\omega\sigma_3(\Lambda_1'-\omega^2)^n X_1 \\ \Lambda_0(\Lambda_1'-\omega^2)^n X_1\end{pmatrix},$$

$$(\widetilde{\Lambda}')^{2n}X=\begin{pmatrix} (1+\tilde{z})(\widetilde{\Lambda}_1-\omega^2)^n(1-\tilde{z})X_1 \\ i\epsilon\omega\sigma_3(1+\tilde{z})(\widetilde{\Lambda}_1-\omega^2)^n(1-\tilde{z})X_1\end{pmatrix},\quad (\widetilde{\Lambda}')^{2n+1}X=\begin{pmatrix} i\epsilon\omega\sigma_3(1+\tilde{z})(\widetilde{\Lambda}_1-\omega^2)^n(1-\tilde{z})X_1 \\ (\Lambda_0-\beta_1 r)(1+\tilde{z})(\widetilde{\Lambda}_1-\omega^2)^n(1-\tilde{z})X_1\end{pmatrix},$$

$$(\widetilde{\widetilde{\Lambda}}')^{2n}X=\frac{1}{\kappa_0^2}\sigma_0\times B_1^{-1}(\lambda_0-i\omega S)\begin{pmatrix} (\widetilde{\Lambda}_\kappa-\omega^2)^n B_1(\lambda_0+i\omega S)X_1 \\ i\epsilon\omega S(\widetilde{\Lambda}_\kappa-\omega^2)^n B_1(\lambda_0+i\omega S)X_1\end{pmatrix}, \tag{B2}$$

$$(\widetilde{\widetilde{\Lambda}}')^{2n+1}X=\frac{1}{\kappa_0^2}\sigma_0\times B_1^{-1}(\lambda_0)-i\omega S)\begin{pmatrix} i\epsilon\omega S(\widetilde{\Lambda}_\kappa-\omega^2)^n B_1(\lambda_0+i\omega S)X_1 \\ \{\widetilde{\Lambda}_0-(\beta_1-1)r+\lambda_0^2\}(\widetilde{\Lambda}_\kappa-\omega^2)^n B_1(\lambda_0+i\omega S)X_1\end{pmatrix},$$

$$n=0,\pm 1,\ldots .$$

It is easy to see that [see (B2)]

$$\{(\Lambda')^2+\omega^2\}^{-1}X=\sum_{K=0}^{\infty}(-1)^K\frac{(\Lambda')^{2K}}{\omega^{2(K+1)}}X=\begin{pmatrix} (\Lambda_1')^{-1}X_1 \\ i\epsilon\omega\sigma_3(\Lambda_1')^{-1}X_1\end{pmatrix}.$$

From this result and (B1), by induction, we derive

$$\{(\Lambda')^2+\omega^2\}^n X=\begin{pmatrix} (\Lambda_1')^n X_1 \\ i\epsilon\omega\sigma_3(\Lambda_1')^n X_1\end{pmatrix},\quad n=0,\pm 1,\ldots .$$

Because of (4.5), from this result and (2.5), follows (4.7). From (4.7), (4.8), (4.10), and the fact that

$$\mathrm{Ad}_{g_0}\{(\Lambda_1')^n q_1'\}=(1+\tilde{z})(\widetilde{\Lambda}_1)^n(1-\tilde{z})\tilde{q}_1,$$

one obtains (4.13). In the same way (4.11), (4.12), and

$$\mathrm{Ad}_g\{(\Lambda_1')^n q_1'\}=-(i/2\kappa_0^2)B_1^{-1}(\lambda_0-i\omega S)(\widetilde{\Lambda}_\kappa)^n S_x,$$

lead to (4.14). By means of (B2) we derive [see (2.9), (3.13), and the remark following (4.15)]

$$\langle\{(\Lambda^1)^{2n}X\}^R \overset{\wedge}{,} \Sigma_1 Y\rangle=\langle X^R \overset{\wedge}{,} \Sigma_1(\Lambda')^{2n}Y\rangle=\langle\{(\widetilde{\Lambda}')^{2n}X\}^R \overset{\wedge}{,} \Sigma_1 Y\rangle$$

$$=\langle X^R \overset{\wedge}{,} \Sigma_1(\widetilde{\Lambda}')^{2n}Y\rangle=\langle\{(\widetilde{\widetilde{\Lambda}}')^{2n}\widetilde{X}\}^R \overset{\wedge}{,} \widetilde{\Sigma}_1\widetilde{Y}\rangle=\langle\widetilde{X}^R \overset{\wedge}{,} \widetilde{\Sigma}_1(\widetilde{\widetilde{\Lambda}}')^{2n}\widetilde{Y}\rangle=0. \tag{B3}$$

By means of (4.15) we have from (2.9) and (3.13)

$$\left[\!\left[\begin{pmatrix}X_1\\X_2\end{pmatrix}_{,\wedge}\begin{pmatrix}Y_1\\Y_2\end{pmatrix}\right]\!\right] = [X_{1,\wedge}\, Y_2] + [X_{2,\wedge}\, Y_1],\quad \left[\!\left[\begin{pmatrix}X_1\\X_2\end{pmatrix}_{,\wedge}\begin{pmatrix}Y_1\\Y_2\end{pmatrix}\right]\!\right]_{\sim} = [X_{j,\wedge}\, Y_2]_{\sim} + [\chi_{,\wedge}\, Y_1]_{\sim}.$$

Using (B2) and (B3), we derive

$$\begin{aligned}
&[\![(\Lambda')^{2n}X_{,\wedge}\, Y]\!] = [\![X_{,\wedge}\,(\Lambda')^{2n}Y]\!] = [\![(\widetilde{\Lambda}')^{2n}X_{,\wedge}\, Y]\!] = [\![X_{,(\wedge)}\,(\widetilde{\Lambda}')^{2n}Y]\!]\\
&\qquad = [\![(\widetilde{\widetilde{\Lambda}}')^{2n}\widetilde{X}_{,\wedge}\, \widetilde{Y}]\!] = [\![\widetilde{X}_{,\wedge}\,(\widetilde{\widetilde{\Lambda}}')^{2n}\widetilde{Y}]\!] = 0,\\
&[\![(\Lambda')^{2n+1}X_{,\wedge}\, Y]\!] = [\![X_{,\wedge}\,(\Lambda')^{2n+1}Y]\!]\\
&\qquad = [\,(\Lambda_0+\omega^2)(\Lambda_1'-\omega^2)^nX_{1,\wedge}\, Y_1] = [X_{1,\wedge}\,(\Lambda_0+\omega^2)(\Lambda_1'-\omega^2)^nY_1],\\
&[\![(\widetilde{\Lambda}')^{2n+1}X_{,\wedge}\, Y]\!] = [\![X_{,\wedge}\,(\widetilde{\Lambda}')^{2n+1}Y]\!] = [\,\widetilde{\Lambda}_1(\widetilde{\Lambda}_1-\omega^2)^n(1-\mathbf{z})X_{1,\wedge}\, Y_1]\\
&\qquad = [X_{1,\wedge}\,\widetilde{\Lambda}_1(\widetilde{\Lambda}_1-\omega^2)^n(1-\tilde{z})Y_1],\\
&[\![(\widetilde{\widetilde{\Lambda}}')^{2n+1}\widetilde{X}_{,\wedge}\, \widetilde{Y}]\!] = [\![\widetilde{X}_{,\wedge}\,(\widetilde{\Lambda}')^{2n+1}Y]\!]\\
&\qquad = \frac{1}{\beta_1\kappa_0^2}\left[\left(1-\frac{1}{\beta_1}z\right)\widetilde{\Lambda}_\kappa(\widetilde{\Lambda}_\kappa-\omega^2)^nB_1(\lambda_0+i\omega S)\,X_{1,\wedge}\,B_1(\lambda_0+i\omega S)\,Y_1\right]_{\sim}\\
&\qquad = \frac{1}{\beta_1\kappa_0^2}\left[B_1(\lambda_0+i\omega S)X_{1,\wedge}\left(1-\frac{1}{\beta_1}z\right)\widetilde{\Lambda}_\kappa(\widetilde{\Lambda}_\kappa-\omega^2)^nB_1(\lambda_0+i\omega S)\,Y_1\right]_{\sim}.
\end{aligned}$$

It is not difficult to verify that [see (A9)]

$$\widetilde{M}\mathrm{S} = -\frac{i}{2\kappa_0^2}\sigma_0\times B_1^{-1}(\lambda_0-i\omega\mathrm{S})\begin{pmatrix}S_x\\ i\epsilon\omega SS_x\end{pmatrix},\quad \widetilde{M}\Lambda_g\mathrm{S} = \widetilde{\widetilde{\Lambda}}'\widetilde{M}\mathrm{S}, \tag{B5}$$

that way from (3.14), (4.12), (B2), (B4), and (B5) one can obtain (4.16) and (4.17).

[1]A. V. Mikhailov and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **74**, 1953 (1978); V. E. Zakharov and L. A. Takhtadjan, Teor. Mat. Fiz. **38**, 26 (1979).
[2]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[3]D. J. Kaup, J. Math. Appl. **54** (3), 849 (1976); D. J. Kaup and A. C. Newell, Adv. Math. **31**, 67 (1979); V. S. Gerdjikov and E. Kh. Khristov, Bulg. J. Phys. 7, 28, 119 (1980).
[4]V. S. Gerdjikov and A. B. Yanovski, Phys. Lett. A **103**, 232 (1984).
[5]V. S. Gerdjikov and A. B. Yanovski, Phys. Lett. A **110**, 53 (1985).
[6]V. S. Gerdjikov, M. I. Ivanov, and Y. S. Vaklev, Inverse Problems **1**, 413 (1986).
[7]D. J. Kaup and A. C. Newell, J. Math. Phys. **19**, 798 (1978).
[8]V. S. Gerdjikov and M. I. Ivanov, Bulg. J. Phys. **10**, 13, 130 (1983).
[9]V. S. Gerdjikov, M. I. Ivanov, and P. P. Kulish, Teor. Mat. Fiz. **44**, 342 (1980).
[10]Y. S. Vaklev and M. I. Ivanov, Teor. Mat. Fiz. **77** (1988).
[11]A. Kundu, J. Math. Phys. **25**, 3433 (1984).
[12]H. H. Chen, Y. C. Lee, and C. S. Liu, Phys. Scr. **20**, 490 (1979); M. Wadati and K. Sogo, J. Phys. Soc. Jpn. **52**, 394 (1983).
[13]V. E. Zakharov and A. B. Shabat, Funct. Anal. Appl. **8**, 43 (1974); **13**, 13 (1979).
[14]M. I. Ivanov, *Hamiltonian structures of the modified nonlinear Schrödinger equations*, Ph.D. thesis, JINR, Dubna 1982.

Construction of the eigenstates for the quantum nonlinear Schrödinger model with the most general supermatrices

Yu-Kui Zhou
Center of Theoretical Physics, CCAST (World Laboratory) and Institute of Modern Physics, Xibei University, Xian, People's Republic of China
Guo-Hun Yun
Department of Physics, Neimenggu University, Hohhot, People's Republic of China

(Received 6 September 1988; accepted for publication 2 February 1989)

A quantum nonlinear Schrödinger model with the most general supermatrices, having a structure $(n-i, i)\times(m-j,j)$, is studied by using the quantum inverse scattering method. The scattering states and bound states of the model are constructed.

I. INTRODUCTION

A one-dimensional many-body system, which contains both the fermions and bosons, with the two-body delta potentials, has been extensively studied.[1-10] As a field theory, this system is the quantum nonlinear Schrödinger model (QNSM) and the Hamiltonian is

$$H=\operatorname{str}\int dx\Big(\frac{\partial q^+}{\partial x}\frac{\partial q}{\partial x}-c{:}q^+qq^+q{:}\Big),\tag{1.1}$$

where the colons : : imply normal product and $c>0$ is the coupling constant. Note that $q(x)=[q(x)_{ba}]$ and $q^+(x)=[q^+(x)_{ab}]$ with $q^+(x)_{ab}=q(x)_{ba}{}^+(-1)^{p(\alpha)}$ are the supermatrices with structures $q\sim(n-i,i)\times(m-j,j)$ and $q^+\sim(m-j,j)\times(n-i,i)$ (Ref. 6). They satisfy the commutation relations

$$\begin{aligned}&q(x)_{\alpha\beta}q^+(y)_{ba}-(-1)^{[p(a)+p(\beta)][p(\alpha)+p(b)]}\\&\quad\times q^+(y)_{ab}q(x)_{\alpha\beta}=(-1)^{p(\alpha)}\delta_{ab}\delta_{\alpha\beta}\delta(x-y),\\&q(x)_{\alpha\beta}q(y)_{ba}-(-1)^{[p(a)+p(\beta)][p(\alpha)+p(b)]}\\&\quad\times q(y)_{ba}q(x)_{\alpha\beta}=0.\end{aligned}\tag{1.2}$$

For $q\sim(n-i,i)\times(m-j,j)$, the supertrace makes the kinetic energy term in H positive definite.

The model (1.1) is the most general form for QNSM. For $0<j<m$ it describes the many-body system with the mixture of a two-body attractive delta potential and repulsive delta potential. In Refs. 7–9 the model (1.1) for the case of $q\sim(1,0)\times(1,0)$, $(0,2)\times(1,0)$, and $(n-k,k)\times(m,0)$, which describes the system with a two-body attractive delta potential, has been studied and the eigenstates have been constructed. In this paper we find the eigenstates for the Hamiltonian (1.1) and the infinite number of the conserved quantities of the model with an arbitrary positive integer $n-i$, $m-j$, i, and j.

This model is integrable.[5] We can find the auxiliary linear equations of the model and they read

$$\begin{aligned}&\frac{\partial}{\partial x}T(x,y|\lambda)={:}L(x|\lambda)T(x,y|\lambda){:},\\&T(y,y|\lambda)=E_{aa}+E_{\alpha\alpha},\end{aligned}\tag{1.3}$$

$$\begin{aligned}&\frac{\partial}{\partial t}T(x,y|\lambda)\\&\quad={:}M(x|\lambda)T(x,y|\lambda){:}-{:}T(x,y|\lambda)M(y|\lambda){:},\end{aligned}\tag{1.4}$$

where L and M are the Lax pair of the model and have the form

$$\begin{aligned}&L(x|\lambda)=i\tfrac{1}{2}\lambda J+S(x),\\&S(x)=i\sqrt{c}q(x)_{b\alpha}E_{b\alpha}+i\sqrt{c}q^+(x)_{ab}E_{ab},\end{aligned}\tag{1.5}$$

$$\begin{aligned}M(x|\lambda)=&-i\tfrac{1}{2}\lambda^2J+ic{:}q(x)_{b\alpha}q^+(x)_{\alpha a}{:}E_{ba}\\&-icq^+(x)_{\alpha b}q(x)_{b\beta}E_{\alpha\beta}-\sqrt{c}\Big(\frac{\partial}{\partial x}q(x)_{a\alpha}\\&+i\lambda q(x)_{a\alpha}\Big)E_{a\alpha}+\sqrt{c}\Big(\frac{\partial}{\partial x}q^+(x)_{\alpha a}\\&-i\lambda q^+(x)_{\alpha a}\Big)E_{\alpha a},\end{aligned}\tag{1.6}$$

$$J=E_{aa}-E_{\alpha\alpha}.$$

Here λ is the spectral parameter, E_{ij} is a $(n+m)\times(n+m)$ matrix with $(E_{ij})_{kl}=\delta_{ik}\delta_{jl}$, the double latin letters a, b imply summations over 1 to n, and the double greek letters α, β mean summations over $n+1$ to $n+m$. From $q\sim(n-i,i)\times(m-j,j)$ we know that the Lax pair (L,M) and $T(x,y|\lambda)$ have the supermatrix structure $(n+m-i-j, i+j)\times(n+m-i-j,i+j)$. From (1.3) we have the following auxiliary equation:

$$\begin{aligned}&\frac{\partial}{\partial y}T(x,y|\lambda)=-{:}T(x,y|\lambda)L(y|\lambda){:},\\&T(x,x|\lambda)=E_{aa}+E_{\alpha\alpha}.\end{aligned}\tag{1.7}$$

Taking (1.3) and (1.4) as the basis of the inverse transform we can find the Yang–Baxter relation (YBR) of the $T(x,y|\lambda)$ and also find some useful permutation relations for the construction of the eigenstates of the system.

II. NEUMANN SERIES, YBR, AND CONSERVED LAW

Firstly, we define the monodromy matrix $T(\lambda)$ and the Jost function $T^{(+)}(x|\lambda)$,

$$T(\lambda)=\lim_{L\to\infty}V(-L|\lambda)T(L,-L|\lambda)V(-L|\lambda),\tag{2.1}$$

$$T^{(+)}(x|\lambda)=\lim_{L\to\infty}V(-L+x|\lambda)T(L,x|\lambda),\tag{2.2}$$

where

$$V(x|\lambda)=\exp(i\tfrac{1}{2}\lambda xJ).$$

 0022-2488/89/081756-04$02.50

Using the auxiliary linear equations (1.3) and (1.7), we have the Neumann series expansions for $T(\lambda)$ and $T^{(+)}(x|\lambda)$,

$$T(\lambda) = 1 + \sum_{i=1}^{\infty} \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{y_1} dy_2 \cdots \int_{-\infty}^{y_{i-1}} dy_i \times :V(-y_1|\lambda)S(y_1)V(y_1-y_2|\lambda)\cdots \times S(y_i)V(y_i|\lambda):, \tag{2.3}$$

$$T^{(+)}(x|\lambda) = 1 + \sum_{i=1}^{\infty} \int_{x}^{\infty} dy_1 \int_{y_1}^{\infty} dy_2 \cdots \int_{y_{i-1}}^{\infty} dy_i \times :V(x-y_i|\lambda)S(y_i)\cdots V(y_2-y_1|\lambda) \times S(y_1)V(y_1-x|\lambda):. \tag{2.4}$$

We write $T(\lambda)$ and $T^{(+)}(x|\lambda)$ as

$$T(\lambda) = \begin{bmatrix} A(\lambda)_{ab}, B(\lambda)_{a\beta} \\ C(\lambda)_{ab}, D(\lambda)_{\alpha\beta} \end{bmatrix},$$

$$T^{(+)}(x|\lambda) = \begin{bmatrix} A(x|\lambda)_{ab}, B(x|\lambda)_{a\beta} \\ C(x|\lambda)_{ab}, D(x|\lambda)_{\alpha\beta} \end{bmatrix},$$

where $a, b = 1, 2, ..., n$ and $\alpha, \beta = n+1, n+2, ..., n+m$. After an inspection on the Neumann series (2.3) and (2.4), we can know that $C(x|\lambda)$ and $D(x|\lambda)$ are analytic in the region $\mathrm{Im}\,\lambda \geqslant 0$ and $C(\lambda)$ is defined only for $\lambda = $ real.

Using (1.3), we can show that the following YBR is valid;

$$R(\lambda-\mu)T(x,y|\lambda) \otimes T(x,y|\mu) = T(x,y|\mu) \otimes T(x,y|\lambda)R(\lambda-\mu), \tag{2.5}$$

$$R(\lambda)_{ji}^{kl} = b(\lambda)\delta_{ik}\delta_{jl} + (-1)^{p(i)p(j)} \times a(\lambda)\delta_{il}\delta_{jk},$$
$$b(\lambda) = 1 - a(\lambda) = ic/(\lambda + ic), \tag{2.6}$$

where the tensor product is defined by

$$(A \otimes B)_{ij,kl} = (-1)^{p(j)[p(i)+p(k)]} A_{ik}B_{jl}.$$

With the help of (2.1) and (2.2) and using (2.5), we can find the following permutation relations that we need for the construction of the eigenstates of the model:

$$D(x|\lambda)_{\beta\delta}C(x|\mu)_{ac} = (-1)^{p(\delta)[p(\alpha)+p(c)]+p(\beta')p(c)} \times C(x|\mu)_{\alpha'c}D(x|\lambda)_{\beta'\delta}\frac{R(\mu-\lambda)_{\beta\alpha}^{\beta'\alpha'}}{a(\mu-\lambda)} + (-1)^{p(\delta)[p(\alpha)+p(c)]+p(\alpha)p(c)} \times C(x|\lambda)_{\beta c}D(x|\mu)_{\alpha\delta}\frac{b(\lambda-\mu)}{a(\lambda-\mu)}, \tag{2.7}$$

$$R(ic)_{\alpha\beta}^{\alpha'\beta'}C(x|\lambda+ic)_{\beta'b}C(x|\lambda)_{\alpha'c}(-1)^{p(\alpha')p(b)} = (-1)^{p(b)p(c)+p(c)p(\alpha')}R(ic)_{\alpha\beta}^{\alpha'\beta'} \times C(x|\lambda+ic)_{\beta'c}C(x|\lambda)_{\alpha'b}, \tag{2.8}$$

$$R(ic)_{\alpha\beta}^{\alpha'\beta'}C(x|\lambda+ic)_{\alpha'c}D(x|\lambda)_{\beta'\delta}(-1)^{p(c)[p(\delta)+p(\beta')]} = (-1)^{p(\delta)p(\beta')}R(ic)_{\alpha\beta}^{\alpha'\beta'}D(x|\lambda+ic)_{\alpha'\delta}C(x|\lambda)_{\beta'c}, \tag{2.9}$$

$$D(\lambda)_{\beta\delta}C(\mu)_{ac} = (-1)^{p(\delta)[p(\alpha)+p(c)]+p(\beta')p(c)} \times C(\mu)_{\alpha'c}D(\lambda)_{\beta'\delta}\frac{R(\mu-\lambda)_{\beta\alpha}^{\beta'\alpha'}}{a(\mu-\lambda)}, \quad (\lambda \neq \mu), \tag{2.10}$$

$$[\mathrm{str}\,D(\lambda),\ \mathrm{str}\,D(\mu)] = 0, \tag{2.11}$$

where the double indices α' and β' mean summations over $n+1$ to $n+m$.

From Eq. (2.11) we know that $\mathrm{str}\,D(\lambda) = (-1)^{p(\alpha)} D(\lambda)_{\alpha\alpha}$ with different spectral parameters can be diagonalized simultaneously. This also enables us to consider $\mathrm{str}\,D(\lambda)$ as a generating function of the set of the infinite conserved quantities of the model. These conserved quantities can be obtained by expanding $\mathrm{str}\,D(\lambda)$ as $i\lambda \to \infty$. Notably, three of them are the particle number

$$N = \mathrm{str}\int dx\, q^+ q,$$

the momentum

$$P = -i\,\mathrm{str}\int dx\, q^+ \frac{\partial q}{\partial x},$$

and the Hamiltonian (1.1).

III. SCATTERING STATES

Firstly, let us consider an $n \times n$ matrix $T_{f_0}(\lambda)_{\alpha'|\alpha}$. Its matrix elements are defined by the following equation for $i = 0$:

$$T_{f_i}^{\delta\beta}(\lambda)_{\alpha'|\alpha} = R(\lambda_{f_i}^{(i)} - \lambda)_{\alpha'_{f_i}\delta}^{\alpha_{f_i}\delta_{f_i-1}} \cdots R(\lambda_2^{(i)} - \lambda)_{\alpha'_2\delta_2}^{\alpha_2\delta_1} \times R(\lambda_1^{(i)} - \lambda)_{\alpha'_1\delta_1}^{\alpha_1\beta} u^{(i)}(\beta),$$
$$n+1 \leqslant \delta,\ \beta,\ \alpha \leqslant n+m; \quad 0 \leqslant i \leqslant m-2, \tag{3.1}$$

where the double indices mean summations, and

$$u^{(0)}(\beta) = \exp\left\{i\pi\left[p(\beta)\sum_{s=1}^{f_0}(p(c_s)+p(\alpha_s)) + p(\beta) + p(c_{f_0})p(\delta) + \sum_{s=1}^{f_0-1} p(c_s)p(\delta_s)\right]\right\},$$

$$u^{(i)}(\beta) = \exp\left[i\pi p(\beta) \times \left(\sum_{s=1}^{i}(f_{s-1}+f_s)p(n+s) + \sum_{s=1}^{f_0} p(\alpha_s) + 1\right)\right]. \tag{3.2}$$

The eigenvalue problems of $\mathrm{tr}\ T_{f_0}(\lambda)$ have been solved.[5] The eigenstates are

$$F_{\alpha_{f_0}\cdots\alpha_1} = [T_{f_0}^{n+1,\beta_1}(\lambda_1^{(1)})T_{f_0}^{n+1,\beta_2}(\lambda_2^{(1)})\cdots T_{f_0}^{n+1,\beta_{f_1}}(\lambda_{f_1}^{(1)})]_{\alpha|n+1} \times [T_{f_1}^{n+2,\delta_1}(\lambda_1^{(2)})T_{f_1}^{n+2,\delta_2}(\lambda_2^{(2)})\cdots T_{f_1}^{n+2,\delta_{f_2}}(\lambda_{f_2}^{(2)})]_{\beta|n+2}\cdots \times [T_{f_{m-2}}^{n+m-1,n+m}(\lambda_1^{(m-1)})\cdots T_{f_{m-2}}^{n+m-1,n+m}(\lambda_{f_{m-1}}^{(m-1)})]_{\rho|n+m-1}, \tag{3.3}$$

and the eigenvalues are $t_1(\lambda)$ given by the recurrence relations

$$t_d(\lambda) = U(d)\prod_{s=1}^{f_d}\frac{b(\lambda-\lambda_s^{(d)}) + (-1)^{p(n+d)}a(\lambda-\lambda_s^{(d)})}{a(\lambda-\lambda_s^{(d)})}$$
$$\times\prod_{s=1}^{f_{d-1}}[b(\lambda_s^{(d-1)}-\lambda) + (-1)^{p(n+d)}a(\lambda_s^{(d-1)}-\lambda)] + \prod_{s=1}^{f_d}\frac{1}{a(\lambda_s^{(d)}-\lambda)}\prod_{s=1}^{f_{d-1}}a(\lambda_s^{(d-1)}-\lambda)t_{d+1}(\lambda),$$
$$1\leqslant d\leqslant m-1, \tag{3.4}$$
$$t_m(\lambda) = u^{(m-1)}(m+n),$$
$$U(d) = \exp\Big[i\pi p(n+d)$$
$$\times\Big(\sum_{s=1}^{f_0}p(\alpha_s) + \sum_{s=1}^{d-1}(f_{s-1}+f_s)p(n+s) + \sum_{s=d+1}^{m-1}(f_{s-1}-f_s)p(n+s) + f_{m-1}p(n+m)+1\Big)\Big]. \tag{3.5}$$

These $\lambda_u^{(r)}$ $(1\leqslant u\leqslant f_r,\ 1\leqslant r\leqslant m-1)$ satisfy

$$\frac{U(d)}{U(d+1)}\prod_{s=1}^{f_{d-1}}\frac{b(\lambda_s^{(d-1)}-\lambda_u^{(d)}) + (-1)^{p(n+d)}a(\lambda_s^{(d-1)}-\lambda_u^{(d)})}{a(\lambda_s^{(d-1)}-\lambda_u^{(d)})}$$
$$=\prod_{s=1}^{f_d}\frac{a(\lambda_u^{(d)}-\lambda_s^{(d)})}{a(\lambda_s^{(d)}-\lambda_u^{(d)})}\frac{b(\lambda_s^{(d)}-\lambda_u^{(d)}) + (-1)^{p(n+d+1)}a(\lambda_s^{(d)}-\lambda_u^{(d)})}{b(\lambda_u^{(d)}-\lambda_s^{(d)}) + (-1)^{p(n+d)}a(\lambda_u^{(d)}-\lambda_s^{(d)})}$$
$$\times\prod_{s=1}^{f_{d+1}}\frac{b(\lambda_u^{(d)}-\lambda_s^{(d+1)}) + (-1)^{p(n+d+1)}a(\lambda_u^{(d)}-\lambda_s^{(d+1)})}{a(\lambda_u^{(d)}-\lambda_s^{(d+1)})}. \tag{3.6}$$

We define a pseudovacuum $|0\rangle$ by $q|0\rangle = 0$, and then we have

$$D(\lambda)_{\alpha\beta}|0\rangle = \delta_{\alpha\beta}|0\rangle,$$
$$C(\lambda)_{ab}|0\rangle \neq 0.$$

Using (2.10), we can show that the scattering states

$$|N\rangle = C(\lambda_1^{(0)})_{\alpha_1c_1}C(\lambda_2^{(0)})_{\alpha_2c_2}\cdots$$
$$\times C(\lambda_{f_0}^{(0)})_{\alpha_{f_0}c_{f_0}}|0\rangle F_{\alpha_{f_0}\cdots\alpha_2\alpha_1} \tag{3.7}$$

are the eigenstates of str $D(\lambda)$, and the eigenvalues are

$$t^s(\lambda) = \prod_{j=1}^{f_0}\frac{1}{a(\lambda_j^{(0)}-\lambda)}t_1(\lambda),$$

where the repeated indices mean summations. The $t_i(\lambda)$ and F's are given by (3.4) and (3.3), respectively.

Since $R_{\alpha\beta}^{\delta\eta}$ satisfies $\alpha+\beta=\delta+\eta$, we can show that $F_{\alpha_{f_0}\cdots\alpha_1}$'s have the conservation condition as follows: the set $(\alpha_{f_0}\cdots\alpha_1)$ contains $N_i = f_{i-1}-f_i$ α's equal to $n+i(1\leqslant i\leqslant m-1)$, and $N_m = f_{m-1}\alpha$'s equal to $n+m$. These f_i $(0\leqslant i\leqslant m-1)$ satisfy $f_0\geqslant f_1\geqslant\cdots f_{m-1}$.

IV. BOUND STATES

We define the operators $(0\leqslant s\leqslant N)$

$$X(x|\lambda,N,s) = C(x|\lambda+i\tfrac{1}{2}c(N-1))_{\alpha_1b_1}\cdots$$
$$\times C(x|\lambda+i\tfrac{1}{2}c(N-2s+1))_{\alpha_sb_s}$$
$$\times D(x|\lambda+i\tfrac{1}{2}c(N-2s-1))_{\alpha_{s+1}\beta_{s+1}}\cdots$$
$$\times D(x|\lambda-i\tfrac{1}{2}c(N-1))_{\alpha_N\beta_N}f_{\alpha_N\cdots\alpha_1}, \tag{4.1}$$

where the double indices mean summations, and

$$f_{\alpha_N\cdots\alpha_1} = F_{\alpha_N\cdots\alpha_2\alpha_1}\big|_{N_i=0} \quad\text{for } p(n+i)=1 \quad (1\leqslant i\leqslant m), \tag{4.2}$$

i.e., f's are given by taking $N_i=0$ for $p(n+i) = 1$ $(1\leqslant i\leqslant m)$ in (3.3). Thus we have

$$U(d) = (-1)^{p(n+d)}$$
$$\text{as } N_i=0 \quad\text{for } p(n+1)=1 \quad (1\leqslant i\leqslant m). \tag{4.3}$$

The X's are well-defined in the region Im $\lambda\geqslant\frac{1}{2}c(N-1)$. Using (1.2), (1.7), (2.8), (2.9), the YBR

$$R(\lambda-\mu)_{\alpha_1\alpha_2}^{\delta_1\delta_2}R(\lambda)_{\delta_2\alpha_3}^{\beta_2\delta_3}R(\mu)_{\delta_1\delta_3}^{\beta_1\beta_3}$$
$$=R(\mu)_{\alpha_1\alpha_3}^{\delta_1\delta_3}R(\lambda)_{\alpha_2\delta_3}^{\delta_2\beta_3}R(\lambda-\mu)_{\delta_1\delta_2}^{\beta_1\beta_2},$$

and the boundary condition

$$X\to\delta_{s,0}\prod_{i=1}^{N}\delta_{\alpha_i\beta_i}f_{\alpha_N\cdots\alpha_1},$$

we can show that X can be analytically continued into the region Im $\lambda\geqslant0$ and

$$X\to0 \quad\text{for } s\neq N, \quad\text{as } x\to-\infty. \tag{4.4}$$

Let p be a real spectral, and we define

$$B_N^+(p) = \lim_{x\to-\infty}X(x|p,N,s=N)\exp(iNpx). \tag{4.5}$$

Using the permutation relation (2.7), (4.4), (2.1), and (2.2), we can show the eigenequation

$$\text{str } D(\lambda)B_N^+(p)|0\rangle = t^b(\lambda)B_N^+(p)|0\rangle, \tag{4.6}$$

with the eigenvalues

$$t^b(\lambda) = \prod_{j=1}^{N}\frac{1}{a(\lambda_j^{(0)}-\lambda)}t_1(\lambda),$$

where $t_1(\lambda)$ is given by taking $N_i=0$ for $p(n+i) = 1$ $(1\leqslant i\leqslant m)$ in (3.4) and

$$\lambda_j^{(0)} = p - i\tfrac{1}{2}(N-2j+1)c \quad (1\leqslant j\leqslant N). \tag{4.7}$$

From (2.8) we have

$$B_N^+(p) = 0, \quad \text{if } p(b_i) = p(b_i) = 1 \quad \text{and } b_i = b_j. \qquad (4.8)$$

It means we can not construct the bound state with the same indices, i.e., $b_i = b_j$, for the fermions. This also leads that a maximum of i fermions can be held in a bound state as $N > i$.

V. CONCLUSION

We have constructed the eigenstates, both the scattering states and bound states, for the infinite conserved quantities of the system. The momentum and energy of the N particles eigenstate are, respectively,

$$\sum_{j=1}^{N} \lambda_j^{(0)} \quad \text{and} \quad \sum_{j=1}^{N} (\lambda_j^{(0)})^2, \quad \text{for the scattering state}$$

and

$$Np \quad \text{and} \quad Np^2 - \frac{c^2}{12} N(N^2 - 1), \quad \text{for the bound state.}$$

These states generally contain both the fermions and bosons.

For the cases of $q \sim (1,0) \times (1,0)$, $(0,2) \times (1,0)$, and $(n-k,k) \times (m,0)$, our results coincide with those in Refs. 7–9.

ACKNOWLEDGMENTS

This work was supported in part by the science fund of the Chinese Academy of Science and the Chinese Young Teacher.

[1] H. B. Thacker, Rev. Mod. Phys. **53**, 253 (1981).
[2] L. D. Faddeev, Sov. Rev. Math. Phys. C **1**, 107 (1981).
[3] C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[4] H. T. Fan, F. C. Pu, and B. H. Zhao, Nucl. Phys. B **299**, 52 (1988).
[5] Y. K. Zhou, to be published in Nucl. Phys. B.
[6] Y. K. Zhou, J. Phys. A **21**, 2391 (1988).
[7] M. Göckler, Z. Phys. C **7**, 263 (1981); **11**, 125 (1981).
[8] F. C. Pu and B. H. Zhao, Nucl. Phys. B **275** (FS17), 77 (1986).
[9] Y. K. Zhou, Z. Phys. **39**, 215 (1988).
[10] F. C. Pu and B. H. Zhao, Phys. Rev. D **30**, 2253 (1984).

The asymptotic behavior of nonlinear waves near a cosmological Cauchy horizon

Vincent Moncrief
Department of Mathematics and Department of Physics, Yale University, P. O. Box 6666, New Haven, Connecticut 06511

(Received 14 February 1989; accepted for publication 29 March 1989)

The asymptotic behavior of solutions of the nonlinear wave equation (with cubic nonlinearity) is studied on background vacuum space-times containing compact Cauchy horizons at the boundaries of their maximal Cauchy developments. The analytic form of the general solution of the linearized equations and that of the nth-order perturbation equations for arbitrarily large n are derived. It is shown how to sum the leading-order terms of the full perturbation series to derive what is presumably the asymptotic form of the general solution of the original nonlinear wave equation near the Cauchy horizon of the background space-time. It is also shown how the indicated asymptotic behavior leads to a natural classification of solutions in terms of "Lagrangian submanifolds" of an associated phase space.

I. INTRODUCTION

In a recent paper[1] we studied the gravitational perturbations, to arbitrarily high order, of a family of vacuum, cosmological space-times that contained compact Cauchy horizons at the boundaries of their maximal Cauchy developments. The aim was to see in detail how perturbations generically destroy such horizons and produce curvature singular boundaries in their place. The background space-times considered in Ref. 1 all had one spacelike Killing field in their globally hyperbolic regions and, for simplicity, the perturbations studied were all required to preserve the isometry groups of the backgrounds.

As a first step towards the study of completely general, nonsymmetric perturbations of these same space-times we here apply the higher-order perturbation method to the model problem of solving the nonlinear wave equation on such backgrounds. We show how one can determine the analytic form of the general solution of the nth-order perturbation equations for arbitrary n and how one can sum the leading-order terms of the full perturbation series to determine what is presumably the asymptotic form of the general solution of the nonlinear wave equation near the Cauchy horizon of the background space-time. We also show how one can classify the singularities of the solutions of the wave equation in a natural way, using the asymptotic form just mentioned, into "Lagrangian submanifolds" of an associated phase space for the problem. At the level of the perturbation equations (to arbitrary order) our results are essentially rigorous, but our determination of the asymptotic behavior of solutions and their associated classification is, at present, only heuristic.

The nonlinear wave equation, with a cubic nonlinearity of the appropriate sign (chosen to ensure positivity of the "energy"), is known to admit global solutions for arbitrarily large data in suitably chosen Sobolev spaces. This is true not only in Minkowski space but also in quite general four-dimensional, globally hyperbolic space-times with sufficiently smooth metrics and appropriate boundary conditions (e.g., asymptotic flatness or compact Cauchy surfaces).[2,3] This result follows from using Sobolev inequalities and applying straightforward higher-order energy estimates to bound the norms of the solutions. Thus the solutions of the wave equation can only blow up at the boundaries of the maximal Cauchy developments of such space-times. In our problem this boundary is always taken to be of the mildest possible type—a smooth (in fact analytic) compact Cauchy horizon across which the space-time can be analytically extended to a casuality violating (hence, non-globally-hyperbolic) region. Such space-times have, if Einstein's equations are imposed, (at least) one spacelike Killing field in their globally hyperbolic regions[4,5] but provide an infinite-dimensional family of vacuum solutions of Einstein's equations.[6,7]

The simplest such space-times have the structures of circle bundles—the fibers of the bundles being defined by the orbits of the Killing fields mentioned above. We briefly review the metrical properties of such vacuum space-times in Sec. II A for the simplest topological category—the product circle bundles. In Sec. II B we formulate the nonlinear wave equation on such a background and show how to determine an infinite-dimensional (but incomplete) family of its analytic solutions by means of a slight extension of the Cauchy–Kowalewski theorem. Each of these solutions, by construction, shares the symmetry of the background. In Sec. II C, using any one of the aforementioned solutions as a background, we show how to derive the general solution of the associated linearized equations and in Sec. II D we extend this result to the perturbation equations of arbitrarily high order. In Sec. III we identify the leading-order terms in the full perturbation series and show how their formal summation leads to a simple picture of the asymptotic behavior of solutions of the nonlinear wave equation near the Cauchy horizon boundary of space-time. This last result, though nonrigorous, suggests a natural classification of the solutions in terms of Lagrangian submanifolds of an associated phase space for the problem.

 0022-2488/89/081760-09$02.50

II. ANALYSIS OF THE NONLINEAR WAVE EQUATION AND ITS PERTURBATIONS

A. Background space-times

As in Ref. 1 we consider Lorentzian metrics defined on manifolds of the form ${}^{(4)}V = K\times\mathbb{R}\times S^1$, where K is a compact, orientable two-manifold. We view these as (trivial) circle bundles over the base manifolds $K\times\mathbb{R}$ and impose upon the metrics to be considered the isometry group of invariance under translations along the circular fibers. For simplicity, we shall only treat trivial (i.e., product) bundles here. However, the same methods could be applied to nontrivial S^1 bundles such as $S^3\times\mathbb{R}\to S^2\times\mathbb{R}$ (cf. Ref. 7).

Let $\{x^a, a=1,2\}$ represent local coordinates on K, x^3 (defined mod 2π) represent an angle coordinate on the circle, and $x^0 = t\in\mathbb{R}$ represent the "time." We consider analytic Lorentzian metrics on ${}^{(4)}V$ expressible in the form[1,6,7]

$$\begin{aligned} ds^2 &= {}^{(4)}g_{\mu\nu}\,dx^\mu\,dx^\nu \\ &= e^{-2\varphi}[-N^2\,dt^2 + g_{ab}\,dx^a\,dx^b] \\ &\quad + t^2e^{2\varphi}(dx^3+\beta_a\,dx^a)^2, \end{aligned} \tag{2.1}$$

where $\partial/\partial x^3$ is a Killing field.

By analogy with the well-known Kaluza–Klein–Jordan reduction program we may view φ, $\beta_a\,dx^a$, and $-N^2\,dt^2 + g_{ab}\,dx^a\,dx^b$ as a scalar field, one-form, and Lorentzian metric induced on the base manifold $K\times\mathbb{R}$ by the space-time metric on $K\times\mathbb{R}\times S^1$. For simplicity (and without any essential loss in generality), we have imposed the coordinate condition of zero shift field. This corresponds to dropping the time component of the one-form field and the shift field of the $(2+1)$-dimensional Lorentzian metric induced on $K\times\mathbb{R}$.

The line element (2.1) degenerates at $t=0$. However, if we reexpress it through the change of coordinates given by either

$$t' = t^2,\quad x^{3'} = x^3 - \ln t,\quad x^{a'} = x^a \tag{2.2}$$

or

$$t' = t^2,\quad x^{3'} = x^3 + \ln t,\quad x^{a'} = x^a, \tag{2.3}$$

then we can show that the transformed metric is analytic and Lorentzian on a neighborhood $\mathcal{N} = K\times S^1\times(-\lambda,\lambda)$ of the hypersurface $t'=0$ provided (i) $\varphi(t',x^{a'})$, $N(t',x^{a'})$, $\beta_a(t',x^{b'})dx^{a'}$, and $g_{ab}(t',x^{c'})dx^{a'}\,dx^{b'}$ are analytic on $\mathcal{N}$; (ii) $N>0$ and g_{ab} is positive definite on $\mathcal{N}$; and (iii) $(N^2 - e^{4\varphi})/4t'$ is analytic on $\mathcal{N}$. By examining the transformed metric in more detail one can also show that (iv) the hypersurface $t'=0$ is a null hypersurface with $\partial/\partial x^{3'}$ tangent to its null generators; and (v) the Killing field $\partial/\partial x^{3'}$ is spacelike in the region $t'>0$ but timelike in the region $t'<0$, where its orbits are closed timelike curves.

Space-times satisfying conditions (i)–(iii) above are globally hyperbolic in the regions $t'>0$ (which were covered by the original coordinates with either $t>0$ or $t<0$), have Cauchy horizons diffeomorphic to $K\times S^1$ at $t'=0$, and are acausal in the regions $t'<0$. The two inequivalent coordinate transformations (2.2) and (2.3) lead to two inequivalent analytic extensions of the original metric (2.1) through its Cauchy horizon at $t=0$.

If, as in Refs. 1, 6, and 7, we impose Einstein's equations upon metrics of the form (2.1), then we may prove the existence of infinite-dimensional families of solutions having all the properties (i)–(v) above provided we impose a suitable coordinate condition to fix the lapse function N. The basic step in the proof is an application of the generalized Cauchy–Kowalewski theorem sketched in Ref. 6 and proved in detail in Ref. 7. The main result is that every choice of analytic initial data $\{\mathring{\varphi}, \mathring{\beta}_a, \mathring{g}_{ab}\}(0,x^c)$ specified over K (with $\mathring{\varphi}$ a function, $\mathring{\beta}_a\,dx^a$ a one-form, and $\mathring{g}_{ab}\,dx^a\,dx^b$ a Riemannian metric) determines a unique, analytic solution of the vacuum Einstein equations having all the properties (i)–(v) above provided the lapse function is chosen to satisfy conditions (i)–(iii) above and the condition

$$\text{(vi)}\quad (N/\sqrt{{}^{(2)}g})_{,t} = 0, \tag{2.4}$$

where ${}^{(2)}g$ is the determinant of g_{ab}. These restrictions lead to the requirement that

$$N/\sqrt{{}^{(2)}g} = e^{2\mathring{\varphi}}/\sqrt{{}^{(2)}\mathring{g}}, \tag{2.5}$$

which fixes N completely.

These rigid coordinate conditions [i.e., zero shift together with (2.4)] are not strictly necessary but were chosen to simplify the form of Einstein's equations and to facilitate the application of the generalized Cauchy–Kowalewski theorem in Refs. 6 and 7.

Many of the solutions determined by data $\{\mathring{\varphi}, \mathring{\beta}_a, \mathring{g}_{ab}\}$ prescribed on K are isometric to one another. For any such solution, however, one can, without disturbing the coordinate conditions imposed above, find a diffeomorphism of ${}^{(4)}V$ that takes ${}^{(4)}g$ to a canonical gauge in which (a) $\mathring{g}_{ab}$ is a constant curvature metric on K depending only on the choice of zero (if $K\approx S^2$), two (if $K\approx T^2$), or $6g-6$ (if K has genus $g\geqslant 2$) real parameters; (b) $\mathring{\beta}_a$ has zero divergence with respect to $\mathring{g}_{ab}$; and (c) there is a residual gauge subgroup action of dimension 6 (if $K\approx S^2$) or dimension 2 (if $K\approx T^2$) generated by the conformal Killing fields of $(K,\mathring{g}_{ab})$ that acts on the data $\{\mathring{\varphi}, \mathring{\beta}_a, \mathring{g}_{ab}\}$. Thus $\mathring{\varphi}$ and the divergence-free part of $\mathring{\beta}_a$ together with the "Teichmüller parameters" for $\mathring{g}_{ab}$ (modulo the action of a finite-dimensional Lie group in the case $K\approx S^2$ or T^2) represent the truly independent data that parametrize the nonisometric solutions of Einstein's equations on ${}^{(4)}V$ which admit compact Cauchy horizons of the type described above.

In fact, as was shown by Isenberg and the author,[4,5] the existence of a compact Cauchy horizon with closed null generators together with the requirements of analyticity and satisfaction of Einstein's equations imply the existence of the Killing field $\partial/\partial x^3$, which we have assumed.

B. The nonlinear wave equation and a class of its analytic solutions

Motivated by the discussion of the previous section, we choose an analytic metric on ${}^{(4)}V$ of the form (2.1), where $\{\varphi, \beta_a, g_{ab}\}(t,x^c)$ are analytic and even in t (hence analytic in $t'=t^2$) and where N is fixed by

$$\frac{N}{\sqrt{{}^{(2)}g}} = \frac{e^{2\mathring{\varphi}}}{\sqrt{{}^{(2)}\mathring{g}}} = \left.\frac{e^{2\varphi}}{\sqrt{{}^{(2)}g}}\right|_{t=0} \tag{2.6}$$

and thus satisfies the coordinate condition $(N/\sqrt{^{(2)}g})_{,t}=0$ and the regularity condition $N^2-e^{4\varphi}=O(t^2)$ discussed in the previous section.

The nonlinear wave equation

$$\Box_{^{(4)}g}\psi=\lambda\psi^3 \tag{2.7}$$

(where λ is a positive constant) takes the explicit form

$$-\left(\psi_{,tt}+\frac{1}{t}\psi_{,t}\right)+{}^{(2)}\widetilde{\Delta}\psi+\left(\frac{1}{t^2}+B\right)\psi_{,33}-\frac{A}{3}(\psi)^3$$
$$-NV^a\psi_{,a3}-(N/\sqrt{^{(2)}g})(\sqrt{^{(2)}g}V^a\psi_{,3})_{,a}=0, \tag{2.8}$$

where

$$A=3\lambda N^2e^{-2\varphi},$$
$$1/t^2+B=N^2e^{-4\varphi}/t^2+N^2\beta_a\beta_bg^{ab}, \tag{2.9}$$
$${}^{(2)}\widetilde{\Delta}\psi=(N/\sqrt{^{(2)}g})(N\sqrt{^{(2)}g}g^{ab}\psi_{,b})_{,a},$$
$$V^a=Ng^{ab}\beta_b,$$

and where g^{ab} is the inverse of the two-metric g_{ab}. The functions A and B as well as the vector field V^a and the coefficients of the linear operator ${}^{(2)}\widetilde{\Delta}$ are all analytic and even in t by virtue of the aforementioned assumptions on ${}^{(4)}g$.

A large family of analytic solutions of Eq. (2.8) may be established by first imposing the invariance condition, $\partial\psi/\partial x^3=0$, and then appealing to the generalized Cauchy–Kowalewski theorem of Ref. (7). Setting $\partial\psi/\partial x^3=0$ everywhere we see that Eq. (2.8) reduces to

$$-(\psi_{,tt}+(1/t)\psi_{,t})+{}^{(2)}\widetilde{\Delta}\psi-(A/3)(\psi)^3=0. \tag{2.10}$$

The generalized Cauchy–Kowalewski (CK) theorem (developed first for nonlinear wave equations by Fusaro[8] and extended to systems of equations in Ref. 7) applies directly to this equation and shows that for any analytic function $\mathring{\psi}(x^a)$ defined on K there exists a solution $\psi(t,x^a)$ of Eq. (2.10) that is analytic and even in t and that has $\psi(0,x^a)=\mathring{\psi}(x^a)$. The evenness in t (hence analyticity in $t'=t^2$) follows from a straightforward inductive argument that shows that all odd t derivatives of ψ vanish at $t=0$ for any analytic solution ψ. The fact every analytic solution has $(\partial\psi/\partial t)|_{t=0}=0$ shows that this set of solutions has only half the free data expected for the general solution of Eq. (2.10). The reason for this is that the remaining solutions must "blow up" in some way as $t\to 0$ and thus fail to satisfy the analyticity demand made in applying the generalized CK theorem.

Unfortunately, therefore, the extended CK theorem does not seem capable of determining the remaining solutions of Eq. (2.10) directly. It is, however, adequate for attacking this equation perturbatively, not only to first order, but in fact to all orders in a perturbation expansion. This follows from a straightforward application of the methods introduced in Ref. 1 for the much more intricate problem of higher-order perturbations of Einstein's equations. Here, however, we want to show that these same higher-order perturbation methods can be applied to Eq. (2.8) as well and yield the general solution of its perturbations to all orders. In this way, we go beyond the analysis of Ref. 1 by showing, at least for the nonlinear wave equation, that the higher-order perturbations technique (based on the generalized CK theorem) is not limited to fields satisfying the invariance condition (e.g., $\partial\psi/\partial x^3=0$) assumed previously.

C. General solution of the linearized equation

We begin by linearizing Eq. (2.8) about an analytic solution $\psi(t,x^a)$ of the type described in the previous section. Since both the space-time metric and the background solution are invariant relative to the Killing field $\partial/\partial x^3$, it is clear that the coefficients in the linearized equation will be independent of the fiber coordinate x^3 and thus that we can conveniently Fourier analyze the perturbation in the x^3 variable.

The linearized equation for a first-order perturbation $\psi^{(1)}$ of the background solution ψ is

$$-(\psi^{(1)}_{,tt}+(1/t)\psi^{(1)}_{,t})+{}^{(2)}\widetilde{\Delta}\psi^{(1)}$$
$$+(1/t^2+B)\psi^{(1)}_{,33}-A(\psi)^2\psi^{(1)}$$
$$-NV^a\psi^{(1)}_{,a3}-(N/\sqrt{^{(2)}g})(\sqrt{^{(2)}g}V^a\psi^{(1)}_{,3})_{,a}=0. \tag{2.11}$$

An arbitrary real solution of Eq. (2.11) can be expanded as

$$\psi^{(1)}=\sum_{m=0}^{\infty}h^{(1)}_m(t,x^a)e^{imx^3}+\text{c.c.}, \tag{2.12}$$

where c.c. signifies "complex conjugate," and where the Fourier coefficient $h^{(1)}_m$ (in general complex) satisfies

$$h^{(1)}_{m,tt}+(1/t)h^{(1)}_{m,t}+(m^2/t^2)h^{(1)}_m+m^2Bh^{(1)}_m-{}^{(2)}\widetilde{\Delta}h^{(1)}_m$$
$$+im(NV^ah^{(1)}_{m,a}+(N/\sqrt{^{(2)}g})(\sqrt{^{(2)}g}V^ah^{(1)}_m)_{,a})$$
$$+A(\psi)^2h^{(1)}_m=0. \tag{2.13}$$

First consider the special case $m=0$ for which Eq. (2.13) reduces to

$$h^{(1)}_{0,tt}+(1/t)h^{(1)}_{0,t}-{}^{(2)}\widetilde{\Delta}h^{(1)}_0+A(\psi)^2h^{(1)}_0=0. \tag{2.14}$$

The latter can be solved by the same methods introduced in Ref. 1. One begins by seeking solutions of the form

$$h^{(1)}_0=(\ln t)a^{(1)}_1+a^{(1)}_0 \tag{2.15}$$

for certain real analytic functions $\{a^{(1)}_1, a^{(1)}_0\}(t,x^a)$. Substituting expression (2.15) into Eq. (2.14) and requiring that the coefficient of each of the two separately occurring powers of (ln t) vanish independently leads to two equations for $a^{(1)}_1$ and $a^{(1)}_0$. The generalized CK theorem is applicable to the first of these equations and shows that an analytic solution $a^{(1)}_1(t,x^a)$ is uniquely determined by arbitrarily specified analytic initial data $\mathring{a}^{(1)}_1(x^a)=a^{(1)}_1(0,x^a)$ prescribed on K. A straightforward inductive argument shows that $a^{(1)}_1$ is even in t (hence analytic in $t'=t^2$). From these properties of $a^{(1)}_1$ it now follows that the second equation is also amenable to the generalized CK theorem and that the latter determines a unique analytic solution $a^{(1)}_0(t,x^a)$ from arbitrarily specified analytic initial data $\mathring{a}^{(1)}_0(x^a)=a^{(1)}_0(0,x^a)$ prescribed on K.

That the two initial data functions $\mathring{a}^{(1)}_1(x^a)$ and $\mathring{a}^{(1)}_0(x^a)$ are genuinely independent may be verified by evaluating the symplectic product of any pair of solutions of the above type. This calculation (which we shall include below as part of a more general result) shows that $\mathring{a}^{(1)}_1$ and $\mathring{a}^{(1)}_0$ are essentially canonically conjugate variables in a Hamiltonian formulation of Eq. (2.14).

Now, for $m\neq 0$, we seek solutions of Eq. (2.13) of the form

$$h_m^{(1)} = e^{im(\ln t)}\lambda_m^{(1)+} + e^{-im(\ln t)}\lambda_m^{(1)-}, \tag{2.16}$$

for some (in general, complex) analytic functions $\{\lambda_m^{(1)\pm}(t,x^a)\}$. Substituting expression (2.16) into Eq. (2.13) and requiring that the coefficients of the two independent factors, $e^{\pm im(\ln t)}$, vanish separately lead to the following equations for $\lambda_m^{(1)\pm}$:

$$\begin{aligned}\lambda_{m,tt}^{(1)\pm} &+ (1/t)(1\pm 2im)\lambda_{m,t}^{(1)\pm} + m^2 B\lambda_m^{(1)\pm} - {}^{(2)}\widetilde{\Delta}\lambda_m^{(1)\pm}\\ &+ im(NV^a\lambda_{m,a}^{(1)\pm} + (N/\sqrt{{}^{(2)}g})(\sqrt{{}^{(2)}g}V^a\lambda_m^{(1)\pm})_{,a})\\ &+ A(\psi)^2\lambda_m^{(1)\pm} = 0.\end{aligned} \tag{2.17}$$

These equations are of the type covered by Fusaro's version of the extended CK theorem.[8] A straightforward application of this theorem shows that analytic solutions $\{\lambda_m^{(1)\pm}(t,x^a)\}$ are uniquely determined by arbitrary analytic data, $\mathring{\lambda}_m^{(1)\pm}(x^a) = \lambda_m^{(1)\pm}(0,x^a)$ prescribed on K. Once again one proves inductively that all odd t derivatives of $\lambda_m^{(1)\pm}$ vanish at $t=0$ and thus that the solutions $\lambda_m^{(1)\pm}(t,x^a)$ are analytic and even in t.

We thus arrive at a formal solution of Eq. (2.11) expressible as

$$\begin{aligned}\psi^{(1)} = \Big\{&(\ln t)a_1^{(1)} + a_0^{(1)} + \sum_{m=1}^{\infty}[e^{im(x^3+\ln t)}\lambda_m^{(1)+}\\ &+ e^{im(x^3-\ln t)}\lambda_m^{(1)-}] + \text{c.c.}\Big\},\end{aligned} \tag{2.18}$$

where $\{a_1^{(1)},a_0^{(1)},\lambda_m^{(1)\pm}\}(t,x^a)$ are all analytic and even in t and are all determined by their (analytic) initial values prescribed on K at $t=0$ (with $\mathring{a}_1^{(1)}$ and $\mathring{a}_0^{(1)}$ real and $\mathring{\lambda}_m^{(1)\pm}$, in general, complex).

To establish the generality of this solution, we shall evaluate the symplectic two-form ω naturally associated with the linearized equation on an arbitrary pair of solutions of the type (2.18). Writing a Lagrangian for Eq. (2.11) and defining the momentum $\pi_\psi^{(1)}$, conjugate to $\psi^{(1)}$, in the usual way, one finds that

$$\pi_\psi^{(1)} = (t/N)\sqrt{{}^{(2)}g}\,\psi_{,t}^{(1)}. \tag{2.19}$$

If $h = (\psi^{(1)}, \pi_\psi^{(1)})$ and $h' = (\psi^{(1)\prime}, \pi_\psi^{(1)\prime})$ represent an arbitrary pair of solutions then the symplectic two-form ω, evaluated on h and h', takes the form

$$\omega(h,h') = \int_\Sigma (\pi_\psi^{(1)}\psi^{(1)\prime} - \pi_\psi^{(1)\prime}\psi^{(1)})d^3x, \tag{2.20}$$

where Σ is a ($t=\text{const}$) Cauchy hypersurface in the globally hyperbolic region of the space-time. The quantity $\omega(h,h')$ is, in fact, hypersurface invariant (i.e., independent of the choice of Σ) and thus has a well defined limit as $t\to 0$. One can evaluate this limit explicitly for any pair of solutions of the form (2.18). Defining real analytic functions $\{\alpha_m^\pm,\beta_m^\pm\}(t,x^a)$ via

$$\lambda_m^{(1)+} = \alpha_m^+ + i\beta_m^+,\quad \lambda_m^{(1)-} = \alpha_m^- + i\beta_m^-, \tag{2.21}$$

so that

$$\begin{aligned}\psi^{(1)} = 2\Big\{&(\ln t)a_1^{(1)} + a_0^{(1)} + \sum_{m=1}^{\infty}[\cos(m(x^3+\ln t))\alpha_m^+\\ &+ \cos(m(x^3-\ln t))\alpha_m^- - \sin(m(x^3+\ln t))\\ &\times\beta_m^+ - \sin(m(x^3-\ln t))\beta_m^-]\Big\},\end{aligned} \tag{2.22}$$

with a similar expression for $\pi_\psi^{(1)}$, one finds that

$$\begin{aligned}\lim_{t\to 0}\omega(h,h') = \Big\{\Big[\int_{\Sigma_h} 4\frac{\sqrt{{}^{(2)}g}}{N}\Big[&a_0^{(1)\prime}a_1^{(1)} - a_0^{(1)}a_1^{(1)\prime}\\ &+ \sum_{m=1}^{\infty} m(\alpha_m^+\beta_m^{+\prime} - \alpha_m^{+\prime}\beta_m^+\\ &+ \alpha_m^{-\prime}\beta_m^- - \alpha_m^-\beta_m^{-\prime})\Big]\Big]\Big\}\Big|_{t=0}.\end{aligned} \tag{2.23}$$

It follows that the initial values of the Fourier coefficients fall naturally into canonically conjugate pairs,

$$\{(a_0^{(1)},a_1^{(1)}),(-\alpha_m^+,\beta_m^+),(\alpha_m^-,\beta_m^-)\}\big|_{t=0},$$

and thus occur independently in the general expression (2.22), which, therefore, has the appropriate generality.

Of course, convergence of the infinite series occurring in Eqs. (2.22) and (2.23) requires an additional restriction upon the Fourier coefficients, beyond satisfaction of the field equations. To sidestep such convergence questions here and in the following sections, we shall, for simplicity, consider only those solutions with finite Fourier series expansions (i.e., those for which $\lambda_m^{(1)\pm}$ vanish identically for sufficiently large m).

D. General solution of the higher-order perturbation equations

Given a background $\psi(t,x^a)$ of the type discussed previously and a solution $\psi^{(1)}(t,x^a,x^3)$ of the linearized equation one can proceed to search for solutions $\psi^{(2)}$ of the second-order perturbation equation. The equation for $\psi^{(2)}$ may be obtained from that for $\psi^{(1)}$ by replacing $\psi^{(1)}$ by $\psi^{(2)}$ in Eq. (2.11) and adding the "source" term, $2A\psi(\psi^{(1)})^2$, to the right-hand side of that equation.

Using the form of the general solution for $\psi^{(1)}$ derived in the previous section we see that this source term takes the form

$$\begin{aligned}2A\psi(\psi^{(1)})^2 = (\ln t)^2 r &+ (\ln t)r + r\\ &+ \sum_{m=1}^{\infty}\{e^{im(x^3+\ln t)}((\ln t)r_m^+ + r_m^+)\\ &+ e^{im(x^3-\ln t)}((\ln t)r_m^- + r_m^-) + \text{c.c.}\}\\ &+ \sum_{k,l=-\infty}^{\infty}\{e^{ikx^3+il(\ln t)}r_{k,l} + \text{c.c.}\},\end{aligned} \tag{2.24}$$

where the symbols r, $r_m^\pm$, and r_{kl} stand generically for certain functions of (t,x^a) that are analytic and even in t. Since this notion of analyticity in (t,x^a) and evenness in t will reoccur repeatedly in the following we shall use the shorthand expression "regular" to refer to functions having these properties.

Furthermore, since we have confined our attention to first-order solutions $\psi^{(1)}$ which have only a finite number of terms in their Fourier expansions, the series expansions in

Eq. (2.24) are actually finite as well (i.e., the regular coefficients $r_m^{\pm}$ and $r_{k,l}$ vanish for sufficiently large k, l, and m).

The source thus consists of a (finite) series of terms of the form

$$S_{k,l,m} = (\ln t)^k e^{ilx^3 + im(\ln t)} r_{k,l,m}, \tag{2.25}$$

for certain integers k, l, and m (with $k \geqslant 0$) and certain regular coefficients $r_{k,l,m}(t,x^a)$. In fact, as we shall see, the corresponding source terms at higher order will all consist of (finite) sums of terms of the same type (provided we always restrict the homogeneous solution at each order to be one of the finite Fourier series type as we did for $\psi^{(1)}$). Therefore, the main step in solving not only the second-order perturbation equation but also the higher-order perturbation equations will be to solve the inhomogeneous wave equation:

$$\begin{aligned} &-(\psi^{k,l,m}_{,tt} + (1/t)\psi^{k,l,m}_{,t}) + {}^{(2)}\widetilde{\Delta}\psi^{k,l,m} \\ &\quad + ((1/t^2) + B)\psi^{k,l,m}_{,33} - A(\psi)^2\psi^{k,l,m} \\ &\quad - NV^a\psi^{k,l,m}_{,a3} - (N/\sqrt{{}^{(2)}g})(\sqrt{{}^{(2)}g}V^a\psi^{k,l,m}_{,3})_{,a} \\ &\quad = (\ln t)^k e^{ilx^3 + im(\ln t)} r_{k,l,m}. \end{aligned} \tag{2.26}$$

Since the homogeneous form of this equation is, of course, identical to that we solved for the first-order perturbation $\psi^{(1)}$, it suffices to show how to find a particular solution of Eq. (2.26).

We begin by seeking a solution of the form

$$\psi^{k,l,m} = e^{ilx^3 + im(\ln t)}\gamma^{k,l,m}(t,x^a), \tag{2.27}$$

which leads to the following equation for $\gamma^{k,l,m}$:

$$\begin{aligned} &\gamma^{k,l,m}_{,tt} + [(1+2im)/t]\gamma^{k,l,m}_{,t} \\ &\quad + [(l^2 - m^2)/t^2]\gamma^{k,l,m} - {}^{(2)}\widetilde{\Delta}\gamma^{k,l,m} \\ &\quad + (A(\psi)^2 + l^2B)\gamma^{k,l,m} + il\,[NV^a\gamma^{k,l,m}_{,a} \\ &\quad + (N/\sqrt{{}^{(2)}g})(\sqrt{{}^{(2)}g}V^a\gamma^{k,l,m})_{,a}] = -(\ln t)^k r_{k,l,m}. \end{aligned} \tag{2.28}$$

To solve Eq. (2.28) we follow the method of Ref. 1 and seek a solution of the form

$$\gamma^{k,l,m} = \sum_{p=0}^{k} (\ln t)^p \eta_p^{k,l,m}(t,x^a), \tag{2.29}$$

for some analytic functions $\{\eta_p^{k,l,m}(t,x^a)\}$. To simplify the notation in the following, we shall suppress the fixed indices and simply write $\eta_p(t,x^a)$ for $\eta_p^{k,l,m}(t,x^a)$.

We substitute expression (2.29) into Eq. (2.28) and demand that the coefficient of each independently occurring power of $\ln t$ vanish separately. This leads to a set of equations, one for each of the η_p. For $p = k$ we always get

$$\begin{aligned} &\eta_{k,tt} + [(1+2im)/t]\eta_{k,t} + ((l^2 - m^2)/t^2)\eta_k \\ &\quad - {}^{(2)}\widetilde{\Delta}\eta_k + (A(\psi)^2 + l^2B)\eta_k \\ &\quad + il\,[NV^a\eta_{k,a} + (N/\sqrt{{}^{(2)}g})(\sqrt{{}^{(2)}g}V^a\eta_k)_{,a}] \\ &\quad + r_{k,l,m} = 0. \end{aligned} \tag{2.30}$$

If $k \geqslant 1$, we also get

$$\begin{aligned} &\eta_{k-1,tt} + [(1+2im)/t]\eta_{k-1,t} + ((l^2 - m^2)/t^2)\eta_{k-1} \\ &\quad + (2imk/t^2)\eta_k + (2k/t)\eta_{k,t} - {}^{(2)}\widetilde{\Delta}\eta_{k-1} \\ &\quad + (A(\psi)^2 + l^2B)\eta_{k-1} + il\,[NV^a\eta_{k-1,a} \\ &\quad + (N/\sqrt{{}^{(2)}g})(\sqrt{{}^{(2)}g}V^a\eta_{k-1})_{,a}] = 0, \end{aligned} \tag{2.31}$$

and, if $k \geqslant 2$, we get, for each $p = k-2,\ldots,0$,

$$\begin{aligned} &\eta_{p,tt} + [(1+2im)/t]\eta_{p,t} + ((l^2 - m^2)/t^2)\eta_p \\ &\quad + [2im(p+1)/t^2]\eta_{p+1} + [2(p+1)/t]\eta_{p+1,t} \\ &\quad + [(p+2)(p+1)/t^2]\eta_{p+2} \\ &\quad - {}^{(2)}\widetilde{\Delta}\eta_p + (A(\psi)^2 + l^2B)\eta_p \\ &\quad + il\,[NV^a\eta_{p,a} + (N/\sqrt{{}^{(2)}g})(\sqrt{{}^{(2)}g}V^a\eta_p)_{,a}] = 0. \end{aligned} \tag{2.32}$$

We first consider the case for which $l^2 = m^2$. In this case each of Eqs. (2.30)–(2.32) reduces to an inhomogeneous generalization of the homogeneous equation (2.17). In particular, the inhomogeneity $r_{k,l,m}$ in Eq. (2.30) is regular by assumption so that Fusaro's version of the extended CK theorem applies to this equation and assures the existence of an analytic solution $\eta_k(t,x^a)$ uniquely determined by arbitrarily prescribed, analytic initial data $\mathring{\eta}_k(x^a) = \eta_k(0,x^a)$. A straightforward inductive argument shows as before that all odd t derivatives of η_k vanish at $t = 0$ and thus that η_k is "regular" (in the sense defined above).

If $k \geqslant 1$ we proceed to Eq. (2.31) (still assuming, for the moment, that $l^2 = m^2$). If $m = 0$ then the inhomogeneous term in Eq. (2.31) is automatically regular [since $(2k/t)\eta_{k,t}$ has that property for regular η_k] so that we may solve, *á la* Fusaro, for a regular η_{k-1} which is uniquely determined by arbitrarily prescribed analytic, initial data $\mathring{\eta}_{k-1}(x^a) = \eta_{k-1}(0,x^a)$. If $m \neq 0$ then we render the inhomogeneous term regular by retroactively demanding that the (heretofore arbitrary) initial data for η_k vanish [i.e., by demanding that $\mathring{\eta}_k(x^a) = 0$.] We then get, as before, a regular solution η_{k-1} of Eq. (2.31) uniquely determined by arbitrarily prescribed analytic initial data $\mathring{\eta}_{k-1}(x^a) = \eta_{k-1}(0,x^a)$.

Now, if $k \geqslant 2$ (still assuming that $l^2 = m^2$), we proceed to Eqs. (2.32). If $m = 0$, we can solve these in the reverse sequence $p = k-2,\ldots,0$ by (retroactively) imposing the conditions

$$\eta_k\,|_{t=0} = 0 \quad \text{(to solve for } \eta_{k-2}\text{)},$$

$$\vdots$$

$$\eta_2|_{t=0} = 0 \quad \text{(to solve for } \eta_0\text{)}.$$

In each case the imposed condition renders the corresponding inhomogeneous term regular and allows one to obtain a regular solution $\eta_p(t,x^a)$ through application of the extended CK theorem. If $m \neq 0$, we have already set $\eta_k\,|_{t=0} = 0$ but, to render the source term in the η_{k-2} equation regular, we must also (retroactively) impose the condition $\eta_{k-1}\,|_{t=0} = 0$ as well. This permits us to apply the extended CK theorem and obtain a regular solution η_{k-2}. Proceeding in this way we find that regular solutions for each of the η_p with $p = k-2,\ldots,0$ may be obtained using the extended CK theorem provided that we restrict the (otherwise arbitrary) initial data by the conditions (for $m \neq 0$)

$$\eta_k\,|_{t=0} = \eta_{k-1}\,|_{t=0} = \cdots = \eta_1|_{t=0} = 0.$$

To summarize the results for $l^2 = m^2$ we have found that Eqs. (2.30)–(2.32) can always be solved for regular

$\eta_p(t,x^a)$ provided we impose the following restrictions upon the (otherwise undetermined) initial data:

$$\eta_k\,|_{t=0} = \eta_{k-1}\,|_{t=0} = \cdots = \eta_2|_{t=0} = 0,$$
$$\text{if } m = 0 \text{ (and } l^2 = m^2), \qquad (2.33)$$

and

$$\eta_k\,|_{t=0} = \eta_{k-1}\,|_{t-0} \doteq \cdots = \eta_1|_{t=0} = 0,$$
$$\text{if } m \neq 0 \text{ (and } l^2 = m^2). \qquad (2.34)$$

Thus, only $\eta_1|_{t=0}$ and $\eta_0|_{t=0}$ remain arbitrary for the case $m = 0$ whereas only $\eta_0|_{t=0}$ remains arbitrary for the cases $m \neq 0$. The same pattern holds for $k = 1$ and $k = 0$, i.e., if $k = 1$ and $m = 0$ then $\eta_1|_{t=0}$ and $\eta_0|_{t=0}$ remain arbitrary, if $k = 1$ and $m \neq 0$ then only $\eta_0|_{t=0}$ remains arbitrary and if $k = 0$ then $\eta_0|_{t=0}$ is always arbitrary.

We now return to Eqs. (2.30)–(2.32) and consider the remaining cases—those for which $l^2 \neq m^2$. We begin by attempting to solve Eq. (2.30) for an analtyic solution η_k. Strictly speaking, Fusaro's version of the extended CK theorem does not apply to this equation because of the presence of the additional singular term, $((l^2 - m^2)/t^2)\eta_k$. Nevertheless, Fusaro's argument (which involves comparison of a formal series solution with the convergent series solution of an associated nonsingular problem) works equally well, with only obvious modifications, for equations of the type (2.30). Indeed, the inclusion of the term $((l^2 - m^2)/t^2)\eta_k$ modifies the singularity structure of the wave operator so that it has the more general form of an operator with "regular singular point" in the t variable—a form familiar in the study of second-order ordinary differential equations. This is a natural generalization of the "Euler–Poisson–Darboux" type singularity considered by Fusaro and it is not surprising that the same methods of analysis apply equally well to it.

This slight further extension of the Cauchy–Kowalewski theorem (which we shall continue to refer to as the extended CK theorem) shows that the formal power series solution of Eq. (2.30) does indeed converge to yield an analytic solution $\eta_k(t,x^a)$ on some neighborhood of the surface $t = 0$. In this case, however, the two singular terms in the equation force both $\eta_k|_{t=0}$ and $(\partial\eta_k/\partial t)\,|_{t=0}$ to vanish identically and a straightforward inductive argument shows, as before, that all odd t derivatives of η_k vanish at $t = 0$. Thus Eq. (2.30) always admits an analytic solution that, moreover, is regular (i.e., analytic and even in t) and that automatically satisfies $\eta_k\,|_{t=0} = (\partial\eta_k/\partial t)\,|_{t=0} = 0$. Thus (when $l^2 \neq m^2$) there is no free data in the analytic solution of Eq. (2.30), which, therefore, would vanish identically in the absence of the inhomogeneous term $r_{k,l,m}$.

If $k \geqslant 1$ we proceed to Eq. (2.31), which, by virtue of the aforementioned properties of η_k, has exactly the same form as Eq. (2.30) [with $(2imk/t^2)\eta_k + (2k/t)\eta_{k,t}$ playing the role of the regular inhomogeneity]. Thus the extended CK theorem applies to this equation and yields a unique regular solution that automatically satisfies

$$\eta_{k-1}\,|_{t=0} = \left.\frac{\partial\eta_{k-1}}{\partial t}\right|_{t=0} = 0.$$

If $k \geqslant 2$ we proceed to Eqs. (2.32) and treat these in the reverse sequence $p = k - 2,\ldots,0$. It is clear that each of these equations will be identical in form to Eq. (2.30) with a regular inhomogeneity provided by the terms

$$\frac{2im(p+1)}{t^2}\eta_{p+1} + \frac{2(p+1)}{t}\eta_{p+1,t} + \frac{(p+2)(p+1)}{t^2}\eta_{p+2}.$$

The regularity of these (apparently singular) terms follows, for each p, from the fact that η_{p+1} and η_{p+2} have already been shown, by a preceding step in the argument, to be regular and to satisfy

$$\eta_{p+1}\,|_{t=0} = \left.\frac{\partial\eta_{p+1}}{\partial t}\right|_{t=0} = \eta_{p+2}\,|_{t=0} = \left.\frac{\partial\eta_{p+2}}{\partial t}\right|_{t=0} = 0.$$

Thus for each p one proceeds as before to obtain a unique regular solution $\eta_p(t,x^a)$ that, moreover, automatically satisfies

$$\eta_p\,|_{t=0} = \left.\frac{\partial\eta_p}{\partial t}\right|_{t=0} = 0.$$

To summarize the results for $l^2 \neq m^2$ we have found that Eqs. (2.30)–(2.32) can always be solved for regular $\eta_p(t,x^a)$ but, in contrast to the previous case for which $l^2 = m^2$, the regular solution is unique and each η_p satisfies, for each $p = k, k-1,\ldots,0$,

$$\eta_p\,|_{t=0} = \left.\frac{\partial\eta_p}{\partial t}\right|_{t=0} = 0 \quad (l^2 \neq m^2). \qquad (2.35)$$

We have thus found that Eq. (2.26) can always be solved by a particular solution of the form

$$\psi^{k,l,m} = e^{ilx^3 + im(\ln t)} \sum_{p=0}^{k} (\ln t)^p \eta_p^{k,l,m}(t,x^a), \qquad (2.36)$$

where each of the $\{\eta_p^{k,l,m}(t,x^a)\}$ is analytic and even in t. Furthermore, since the linear operator on the left-hand side of Eq. (2.26) is purely real, it follows that $(\psi^{k,l,m} + \text{c.c.})$ satisfies the corresponding real equation obtained from (2.26) by replacing the source term $S_{k,l,m}$ on the right-hand side by the real source $(S_{k,l,m} + \text{c.c.})$. Thus the general solution of the second-order perturbation equation can always be written as a finite series (under the simplifying assumption introduced previously) of terms of the form $(\psi^{k,l,m} + \text{c.c.})$ superimposed with the general solution of the homogeneous equation [c.f. Eq. (2.18)].

Under the same simplifying assumption (of restricting the Fourier series in the homogeneous solution to a finite series) one thus finds that an arbitrary solution of the second-order perturbation equation has the form of a finite sum of terms of the form

$$\chi^{p,l,m} = e^{ilx^3 + im(\ln t)}(\ln t)^p \eta^{p,l,m}(t,x^a) + \text{c.c.},$$

where $\eta^{p,l,m}(t,x^a)$ is analytic and even in t. From this it follows that the source term in the third-order perturbation equation [given by $2A(\psi^{(1)})^3 + 6A\psi^{(1)}\psi^{(2)}$] takes the same form as that of the source encountered in the second-order equation—a finite sum of terms of the type $S_{k,l,m}$ [c.f. Eq. (2.25)]. Consequently the same methods used to solve the second-order equation apply equally well to the third-order equation whose general solution thus consists of a series of terms of the same type as $\chi^{p,l,m}$.

This pattern clearly continues inductively to arbitrarily high order in perturbation theory. The source at each order arises from finite sums of products of functions of the type $\chi^{p,l,m}$ and therefore always takes the form of a series of terms of the type $S_{k,l,m}$. Thus the pattern always continues to the succeeding order. The artifical restriction that each homogeneous solution only contain a finite set of Fourier modes could surely be replaced by a weaker condition permitting infinite Fourier series solutions but we shall not try to deal with the associated convergence questions here.

III. SUMMATION OF LEADING-ORDER TERMS IN THE PERTURBATION SERIES AND CLASSIFICATION OF THE ASYMPTOTIC SOLUTIONS

A closer inspection of the arguments of the previous section reveals a remarkable simplicity in the asymptotic form of the general solution of the nth-order perturbation equations and thus strongly suggests (without, however, rigorously proving) a corresponding simplicity in the asymptotic form of the general solution of the nonlinear wave equation. This asymptotic form in turn suggests a natural classification of the solutions in terms of "Lagrangian submanifolds" of an associated phase space for the wave equation.

To see how this arises, let us return to the analysis of Eqs. (2.30)–(2.32) and first consider the case for which $l^2 \neq m^2$. Since the source term $r_{k,l,m}$ in Eq. (2.30) is regular, it has an expansion of the form

$$r_{k,l,m} = \rho_{k,l,m}(x^a)t^{2q-2} + O(t^{2q}), \tag{3.1}$$

for some integer $q \geqslant 1$ and some analytic function $\rho_{k,l,m}(x^a)$. The recurrence relation for Eq. (2.30) then leads to an expression for η_k of the form

$$\eta_k = C_{k,l,m}(x^a)t^{2q} + O(t^{2q+2}), \tag{3.2}$$

where

$$C_{k,l,m}(x^a) = -\rho_{k,l,m}(x^a)/[2q(2q+2im) + l^2 - m^2]. \tag{3.3}$$

Now, if $k \geqslant 1$, then η_k provides a source term in Eq. (2.31) for η_{k-1} of the form

$$\begin{aligned}(2imk/t^2)\eta_k + (2k/t)\eta_{k,t} \\ = 2k(2q+im)C_{k,l,m}t^{2q-2} + O(t^{2q}),\end{aligned} \tag{3.4}$$

and thus leads to a solution for η_{k-1} of the form

$$\begin{aligned}\eta_{k-1} &= \frac{-2k(2q+im)C_{k,l,m}}{[2q(2q+2im)+l^2-m^2]}t^{2q} + O(t^{2q+2}) \\ &= \frac{2k(2q+im)\rho_{k,l,m}}{[2q(2q+2im)+l^2-m^2]^2}t^{2q} + O(t^{2q+2}).\end{aligned} \tag{3.5}$$

If $k \geqslant 2$, then one continues with the solution of Eqs. (2.32) for each $p = k-2,...,0$. However, it is clear that, for each p, the functions η_{p+1} and η_{p+2} will provide a source term of precisely the same type, i.e.,

$$\begin{aligned}&\frac{2im(p+1)}{t^2}\eta_{p+1} + \frac{2(p+1)}{t}\eta_{p+1,t} \\ &+ \frac{(p+2)(p+1)}{t^2}\eta_{p+2} \\ &= \delta^p_{k,l,m}(x^a)t^{2q-2} + O(t^{2q})\end{aligned} \tag{3.6}$$

where $\delta^p_{k,l,m}(x^a)$ is a uniquely determined, constant multiple of the original source function $\rho_{k,l,m}(x^a)$. Thus, for each $p = k-2,...,0$, one gets a solution of the form

$$\eta_p = \frac{-\delta^p_{k,l,m}(x^a)}{[2q(2q+2im)+l^2-m^2]}t^{2q} + O(t^{2q+2}), \tag{3.7}$$

for fixed $q \geqslant 1$, where each $\delta^p_{k,l,m}(x^a)$ is a given constant multiple of $\rho_{k,l,m}(x^a)$.

The situation for $l^2 = m^2$ is similar except that now, if $m \neq 0$, then one is free to choose $\eta_0|_{t=0}$ arbitrarily, and, if $m = 0$, then one is free to choose $\eta_1|_{t=0}$ and $\eta_0|_{t=0}$ arbitrarily. This latter freedom merely reflects the freedom to add an arbitrary solution of the linearized equation to any particular solution of the (inhomogeneous) nth-order perturbation equation. The simplest choice for such a particular solution clearly results from imposing the additional restrictions (for $l^2 = m^2$)

$$\begin{aligned}&\eta_1|_{t=0} = \eta_0|_{t=0} = 0, \quad \text{if } m = 0, \\ &\eta_0|_{t=0} = 0, \quad \text{if } m \neq 0.\end{aligned} \tag{3.8}$$

This choice renders the particular solution unique and expressible in the same form as the solution obtained above for $l^2 \neq m^2$. In other words, if $r_{k,l,m}$ is written as in Eq. (3.1) then each of the η_p, for $p = k, k-1,...,0$, has an expression of the form

$$\eta_p = C^p_{k,l,m}(x^a)t^{2q} + O(t^{2q+2}), \tag{3.9}$$

where $C^p_{k,l,m}(x^a)$ is a uniquely determined, constant multiple of $\rho_{k,l,m}(x^a)$.

Furthermore, at each order of perturbation theory, the collection of leading-order coefficients in the source terms [i.e., the collection of functions $\rho_{k,l,m}(x^a)$ defined at that order] are determined by purely algebraic operations (multiplication, addition, and multiplication by regular "background" functions) from the leading-order coefficients in the lower-order perturbations. These latter consist of the leading-order terms from the particular solutions described above together with the leading-order (and arbitrarily specifiable) terms from the general solution of the associated homogeneous equation [given in Eq. (2.18)].

The key point is that all of these leading-order terms are determined by purely algebraic operations from the free data introduced at each order in the form of the arbitrarily specified analytic functions

$$\{a_1^{(1)}, a_0^{(1)}, \lambda_m^{(1)+}, \lambda_m^{(1)-}\}|_{t=0} \quad [\text{cf. Eq. (2.18)}],$$

which determine an arbitrary solution of the homogeneous perturbation equation. In fact, it is not hard to show that these leading-order terms are, to all orders in perturbation theory, precisely the same as those we would obtain by an analogous perturbative solution of the "truncated nonlinear wave equation,"

$$-\left(\psi_{,tt}+\frac{1}{t}\psi_{,t}\right)+\frac{1}{t^2}\psi_{,33}=\frac{A}{3}(\psi)^3, \tag{3.10}$$

provided we choose the free data at each order in the same way for the two problems.

The truncated equation can also be written in the form

$$-\psi_{,\tau\tau}+\psi_{,33}=e^{-2\tau}(A/3)(\psi)^3, \tag{3.11}$$

where $\tau=-\ln t$, and in which A is analytic in the variables x^a and $t^2=e^{-2\tau}$. This is a simple $(1+1)$-dimensional nonlinear wave equation defined on each of the "sheets" $\{x^a=\text{const}\}$ with the additional feature that the coefficient of the nonlinearity decays exponentially as $\tau\to\infty$ (i.e., $t\to 0^+$). Thus, at least as judged by the leading-order terms in their perturbation expansions, the solutions of the original nonlinear wave equation and the truncated nonlinear wave equation have the same asymptotic behavior as $\tau\to\infty$ (i.e., as one approaches the Cauchy horizon of the background space-time).

To proceed further with the study of this asymptotic behavior, it is convenient to impose a restriction upon the free data that one can introduce at each order of perturbation theory in the form of an arbitrary solution of the linearized equation. This "embarassment of riches" of free data corresponds, of course, to the fact that the perturbation series allows one, in principle, to express an arbitrary (analytic) curve of exact solutions of the nonlinear wave equation that passes through the background solution. Insofar as we are only interested in arbitrary solutions, as opposed to curves of solutions, we can, without any real loss of generality, simply turn off all the free data at every order of perturbation theory beyond the first. In other words, at first order we take the general solution given by Eq. (2.18) but at each successive order we take the unique particular solution of the corresponding inhomogeneous equation defined by the prescription outlined above and refrain from adding any nonvanishing solution of the homogeneous equation to these particular solutions.

This choice has the virtue of uniquely parametrizing the full perturbative solution in terms of the free data introduced at first order. Recalling the aforementioned results on the form of the particular solutions of the higher-order perturbation equations, one now finds that the full perturbation series

$$\psi(x^\mu;\epsilon)=\overset{\circ}{\psi}(t,x^a)+\sum_{n=1}^{\infty}\frac{\epsilon^n}{n!}\psi^{(n)}(x^\mu) \tag{3.12}$$

takes the form (upon evaluation at $\epsilon=1$)

$$\psi(x^\mu)=\overset{\circ}{\psi}(t,x^a)+\psi^{(1)}(x^\mu)+t^2\Bigg\{\sum_{l,m=-\infty}^{\infty}\sum_{p=0}^{\infty}e^{ilx^3+im(\ln t)}\times\mu_p^{l,m}(t,x^a)(\ln t)^p+\text{c.c.}\Bigg\}. \tag{3.13}$$

Here, each of the $\{\mu_p^{l,m}(t,x^a)\}$ is, at least formally, expressible as a series in positive powers of t^2 with coefficients uniquely determined by the data chosen for the arbitrary solution of the linearized equation, $\psi^{(1)}(x^\mu)$. A variation of this solution is therefore generated by varying the free data in $\psi^{(1)}$ and computing the uniquely determined variations in the $\mu_p^{l,m}(t,x^a)$. Thus such a variation has the form

$$\delta\psi=\delta\psi^{(1)}(x^\mu)+t^2\Bigg\{\sum_{l,m=-\infty}^{\infty}\sum_{p=0}^{\infty}e^{ilx^3+im(\ln t)}\times\delta\mu_p^{l,m}(t,x^a)(\ln t)^p+\text{c.c.}\Bigg\}. \tag{3.14}$$

We want to compute the symplectic product of an arbitrary pair of perturbations of the perturbative solution $\psi(x^\mu)$ given in Eq. (3.13). The momentum conjugate to ψ is

$$\pi_\psi=(t/N)\sqrt{^{(2)}g}\,\psi_{,t} \tag{3.15}$$

and the symplectic product of any pair of perturbations $h=(\delta\psi,\delta\pi_\psi)$ and $h'=(\delta\psi',\delta\pi'_\psi)$ of a solution (ψ,π_ψ) is given by

$$\Omega(h,h')=\int_\Sigma(\delta\pi_\psi\delta\psi'-\delta\pi'_\psi\delta\psi)d^3x. \tag{3.16}$$

One can evaluate this, at least formally, in the limit as $t\to 0$ by assuming the validity of a term by term computation of the limits occurring in the infinite series expressions. The main point is that, because of the extra factors of t^2 multiplying the higher-order terms in $\delta\psi$ and $\delta\psi'$ [cf. Eq. (3.14)], none of the terms involving the $\{\delta\mu_p^{l,m}\}$ or the $\{\delta\mu_p^{l,m'}\}$ make a contribution to the limit. If we parametrize the linearized solution $\psi^{(1)}$ in Eq. (3.13) as in Eq. (2.22) and designate the variation of its parameters defining $\delta\psi^{(1)}$ and $\delta\psi^{(1)'}$ by $\{\delta a_1^{(1)},\delta a_0^{(1)},\delta\alpha_m^\pm,\delta\beta_m^\pm\}$ and $\{\delta a_1^{(1)'},\delta a_0^{(1)'},\delta\alpha_m^{\pm\prime},\delta\beta_m^{\pm\prime}\}$ then we get

$$\lim_{t\to 0}\Omega(h,h')=\Bigg\{\int_{\Sigma_h}4\frac{\sqrt{^{(2)}g}}{N}\Big[\delta a_0^{(1)'}\delta a_1^{(1)}-\delta a_0^{(1)}\delta a_1^{(1)'}+\sum_{m=1}^{\infty}m(\delta\alpha_m^+\delta\beta_m^{+\prime}-\delta\alpha_m^{+\prime}\delta\beta_m^+ +\delta\alpha_m^{-\prime}\delta\beta_m^- -\delta\alpha_m^-\delta\beta_m^{-\prime})\Big]\Bigg\}\Bigg|_{t=0}, \tag{3.17}$$

which closely resembles Eq. (2.23). Now, however, h and h' represent perturbations of the solution $\psi(x^\mu)$, not the original background solution $\psi(t,x^a)$.

The data determining an arbitrary formal solution $\psi(x^\mu)$ [cf. Eq. (3.13)] are, of course, just the "initial values" (fixed at $t=0$) of the functions $\{a_1^{(1)}, a_0^{(1)}, \alpha_m^\pm, \beta_m^\pm\}$ that determine the first-order perturbation $\psi^{(1)}$ and, according to our prescription, all the higher-order perturbations as well. Thus we may regard the initial values of these functions as parametrizing an "asymptotic phase space" for the nonlinear wave equation. These initial data fall naturally into canonically conjugate pairs as one sees from Eq. (3.17). Suppose, for any particular solution $\psi(x^\mu)$, we consider the corresponding affine subspace of the asymptotic phase space defined by holding

$$(a_1^{(1)}|_{t=0},(\alpha_m^+-\alpha_m^-)|_{t=0},(\beta_m^+-\beta_m^-)|_{t=0})$$

fixed while allowing the complementary data

$$(a_0^{(1)}|_{t=0},(\alpha_m^++\alpha_m^-)|_{t=0},(\beta_m^++\beta_m^-)|_{t=0})$$

to vary arbitrarily. Any pair of perturbations h and h' of $\psi(x^\mu)$ that lie in this subspace will therefore have data satisfying

$$\delta\alpha_1^{(1)}\big|_{t=0} = \delta\alpha_1^{(1)\prime}\big|_{t=0} = 0,$$
$$(\delta\alpha_m^+ - \delta\alpha_m^-)\big|_{t=0} = 0 = (\delta\alpha_m^{+\prime} - \delta\alpha_m^{-\prime})\big|_{t=0}, \tag{3.18}$$
$$(\delta\beta_m^+ - \delta\beta_m^-)\big|_{t=0} = 0 = (\delta\beta_m^{+\prime} - \delta\beta_m^{-\prime})\big|_{t=0},$$

and thus yield $\lim_{t\to 0} \Omega(h,h') = 0$ when substituted into Eq. (3.17). Thus every such subspace is "isotopic" relative to the symplectic form Ω and has, roughly speaking, half the dimension of the full asymptotic phase space. A complementary isotopic subspace can readily be defined through $\psi(x^\mu)$ by exchanging the roles of those data held fixed and those allowed to vary freely.

Each such isotopic subspace is, therefore, at least at this formal level, a "Lagrangian submanifold" of the asymptotic phase space. The relative asymptotic behavior of any pair of solutions ψ and $\widetilde{\psi}$ lying in such a Lagrangian subspace may be found by subtracting the two expressions of the form (3.13), dropping all the terms that tend to zero as $t\to 0$, and imposing the conditions (3.18) upon the variations of the various parameters from their background values. The result is

$$\psi - \widetilde{\psi} \sim f(x^a) + g(x^3 - \ln t, x^a) + g(x^3 + \ln t, x^a). \tag{3.19}$$

The divergent term in $\ln t$ has canceled and the "right and left moving waves" in each "sheet," $x^a = \text{const}$, have the same shape by virtue of the conditions imposed from Eq. (3.18).

ACKNOWLEDGMENTS

I am grateful to the Centre for Mathematical Analysis at the Australian National University for their hospitality while some of the work was being carried out.

This research was supported in part by National Science Foundation Grant No. PHY-8503072 to Yale University.

[1]V. Moncrief, "Symmetry preserving deformations of generalized Taub–NUT spacetimes to all orders in perturbation theory," to be published in J. Math. Phys.

[2]For the global existence theory in Minkowski space-time see M. Reed, "Abstract non-linear wave equations," *Springer Lecture Notes in Mathematics*, Vol. 507 (Springer, Berlin, 1976).

[3]A recent study of global existence results for nonlinear wave equations in more general Lorentzian manifolds has been made by P. Chrusciel and J. Baez (unpublished).

[4]V. Moncrief and J. Isenberg, Commun. Math. Phys. **89**, 387 (1983); see also J. Isenberg and V. Moncrief, J. Math. Phys. **26**, 1024 (1985).

[5]A paper extending the results of Ref. 4 to the case of compact analytic, vacuum Cauchy horizons with nonclosed null generators is in preparation by V. Moncrief and J. Isenberg.

[6]V. Moncrief, Ann. Phys. (NY) **141**, 83 (1982).

[7]V. Moncrief, J. Geom. Phys. **1**, 107 (1984).

[8]B. A. Fusaro, SIAM Rev. **10**, 417 (1968). For a more recent treatment of such singular initial value problems that uses modern functional analytic methods, see M. S. Baouendi and C. Goulaouic, J. Differ. Eqs. **22**, 268 (1976); Comm. Pure Appl. Math. **26**, 455 (1973).

World sheet diffeomorphisms and the canonical string

K. V. Kuchař and C. G. Torre[a)]
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 8 February 1989; accepted for publication 5 April 1989)

The degree to which the phase space of a closed bosonic string carries a representation of the two-dimensional diffeomorphism group Diff(M) is investigated. In particular, homomorphic mappings from the associated Lie algebra diff(M) into the Poisson algebra of functions on three natural phase spaces associated with the string are constructed. Two of these spaces are extended phase spaces based on the conformal and harmonic gauges, respectively. The third space is essentially the original phase space; the homomorphism in this case relies on the validity of the light-cone gauge. Homomorphisms from diff(M) into the extended phase spaces of the Batalin–Fradkin–Vilkovisky formalism are also constructed. While the methods used to represent diff(M) cannot be extended to represent all of Diff(M), the phase spaces *do* carry a representation of the subgroup of conformal isometries. It is argued that this subgroup is sufficiently large to serve as the *dynamical* group for the string. The implications of this work for a true Dirac quantization of the string via operator representations of diff(M) are discussed.

I. INTRODUCTION

It is well known that the presence of an infinite-dimensional invariance group for a dynamical system manifests itself through the appearance of constraints in the Hamiltonian formalism.[1] For gauge theory, the constraint functions are linear in the canonical momenta and serve as comoments for the action of the gauge group on the phase space, that is, the Poisson algebra of the constraint functions is a homomorphic image of the Lie algebra of the gauge group. When the invariance group is the diffeomorphism group Diff(M) for the underlying space-time arena M, one finds a Hamiltonian constraint quadratic in momenta and, for field theories, a momentum constraint linear in momenta. In this case, the relationship between the constraint functions and the invariance group is more obscure. Indeed, if one computes the Poisson brackets between the constraint functions, one does not find the Lie algebra diff(M) of the diffeomorphism group, but rather the open algebra of hypersurface deformations (see, e.g., the first reference in Ref. 1). The hypersurfaces being deformed are the Cauchy surfaces on which the canonical variables are defined. Therefore, insofar as the role of the diffeomorphism group is concerned, the canonical treatment of the classical and quantum dynamics of generally covariant systems is quite different from other "manifestly covariant" formulations. It is usually maintained that the loss of covariance is an unavoidable by-product of any canonical formalism. However, as emphasized by Kuchař and Isham,[2] space-time covariance is not so much lost as it appears in a different guise. This is well illustrated by the canonical formulation of ordinary Minkowski space field theory: There, manifest Poincaré invariance is lost in the usual sense; nevertheless, the phase space *does* carry a representation of the Poincaré group via the energy- and angular-momentum tensors. This is, of course, how one knows the canonical formulation is Poincaré covariant.

In Ref. 2 it was shown how a similar game can be played with Diff(M). The key observation made in Ref. 2 is that the needed link between intrinsically space-time quantities, i.e., diffeomorphisms and canonically defined quantities, i.e., constraints, is given by the *embeddings* that implant the Cauchy surfaces into the space-time. By explicitly incorporating the embeddings into the phase space, one links the action of space-time diffeomorphisms with the corresponding hypersurface deformations. More specifically, given $V \in \mathrm{diff}(M)$, one finds the corresponding phase space function, the "diffeomorphism Hamiltonian" $H(\vec{V})$, by composing a pair of Lie algebra antihomomorphisms. The first antihomomorphism is the familiar map from diff(M) into the commutator algebra of (complete) vector fields on M. The vector field $\vec{V}$ corresponding to V is then restricted to a given embedding and one uses the resulting functional of the embeddings as a smearing field for an appropriate combination of the embedding momenta and the Hamiltonian and momentum constraint functions. This map from the commutator algebra of vector fields into the Poisson bracket algebra of phase space functions is another antihomomorphism. The net effect is a homomorphism from diff(M) into the phase space Poisson bracket algebra $V \rightarrow \vec{V} \rightarrow H(\vec{V})$. In Ref. 2 this technique was utilized to represent diff(M) on the phase space of a parametrized scalar field and on an extended phase space for general relatively obtained by using Gaussian coordinate conditions.

In this paper we will investigate the above ideas in the context of the Hamiltonian formulation of a closed bosonic string. Our reason for studying this relatively simple system comes in two parts. First, as is well known, the string provides the most promising (or at least the most popular) framework for unifying all fundamental forces, including

[a)] Present address: Space Research Institute, Florida Institute of Technology, Melbourne, FL 32901.

 0022-2488/89/081769-25$02.50

gravity, in a quantum mechanical setting. One of the central features of string theory is the important role that world sheet symmetries of the first-quantized formalism play in defining the properties of the second-quantized field theoretic formalism. Now, while lip service is usually paid to the world sheet diffeomorphism group, the constraint algebra for the string is actually the Lie algebra conf(M,g) of the group Conf(M,g) of conformal isometries of the metric g on M, where M is the manifold which is embedded as a world sheet in Minkowski space. It is this symmetry group that has played the main role in the canonically quantized theory. Thus to date, the role (if any) of the full two-dimensional diffeomorphism group [of which Conf(M,g) is a subgroup] has remained obscure in the canonical approach to string quantum mechanics and string field theory. From the point of view of the Polyakov formalism,[3,4] the role of Diff(M) is more pronounced. For example, the origin of the critical dimension can be seen as a consequence of one's inability to define a functional integration measure which is simultaneously diffeomorphism and Weyl invariant. A direct translation of the results of the Polyakov formalism into the vernacular of the canonically quantized string is not available. By making the canonical role of Diff(M) explicit, one should, to some degree, be able to unify the canonical and covariant (Polyakov) quantization schemes.

Our second purpose for studying the string de-emphasizes its role as providing a "theory of everything" and instead stresses the role of the string as a simple paradigm for studying the quantization of more conventional theories of gravity. The string, like gravity, is a generally covariant system with Hamiltonian and momentum constraints. Unlike gravity, the string is sufficiently simple such that one can study the canonical quantization process in some detail. By studying the quantum mechanical status of Diff(M) in this simple setting, we hope to shed light on corresponding issues in quantum gravity.[2] For example, one can investigate the effect of operator-ordering difficulties and/or anomalies on quantum mechanical covariance within the framework of canonical quantization. A preliminary examination of this particular question is given in Ref. 5.

This paper is meant to serve primarily as performing the necessary classical groundwork for studying the above quantum mechanical issues. Our goal will be to study several ways in which one can naturally incorporate diff(M) into the Hamiltonian formulation of the closed string. With proper attention to boundary conditions, the results obtained can be easily translated into the open string framework.

We organize the paper as follows. Section II is mainly devoted to preliminaries and is divided into two parts. Section II A summarizes the Hamiltonian formulation of parametrized harmonic maps from a cylindrical space-time into a (possibly curved) target manifold. The way we represent diff(M) and conf(M,g) here is to serve as the model for all the work to follow. Section II B contains a sketch of the relevant aspects of the canonical formulation of the closed bosonic string. In Sec. III we utilize the techniques of Ref. 2 to represent diff(M) on extended phase spaces associated with the conformal and harmonic gauges. In Sec. IV we show that one can go quite a long way toward extracting the embeddings and their conjugate momenta directly from the usual string phase space of Sec. II. In contrast to Sec. III (or Ref. 2), no gauge fixing is needed here, although for technical reasons, it is convenient to extend slightly the conventional string phase space. With an eye on quantization, in Sec. V and we construct BRST (Becchi–Rouet–Stora–Tyutin) invariant extensions of the diff(M) representatives obtained in Secs. III and IV. In Sec. VI we comment on the distinguished role of the group Conf(M,g) of conformal isometries as both symmetry and *bona fide* dynamical groups. We discuss the results obtained and the corresponding implications for their quantum mechanical counterparts in Sec. VII. We have provided three appendices. Appendix A summarizes some rudimentary features of the two-dimensional conformal group. Needed results from the BRST formalism are given in Appendix B. Finally, we summarize our notation and conventions in Appendix C.

II. PRELIMINARIES

A. Parametrized scalar fields on the cylinder

We consider a set of massless scalar fields φ^A $(A = 1,...,d)$ propagating on a globally hyperbolic cylindrical space-time (M,g), where $M = R \times S^1$ and g_{ab} is a fixed, externally prescribed metric of Lorentzian signature. The action functional describing the fields is taken to be

$$S[\varphi] = -\frac{1}{2}\int_M \sqrt{-g}\, g^{ab} G_{AB} \nabla_a \varphi^A \nabla_b \varphi^B. \qquad (2.1)$$

In (2.1) we have introduced a "target space" metric G_{AB} which is a nondegenerate ultralocal function of the scalar fields. The stationary points of $S[\varphi]$ correspond to harmonic maps $\varphi^A{:}M \to M^d$, where M^d is a d-dimensional target space

$$\frac{\delta S}{\delta \varphi^A} = \sqrt{-g}\, g^{ab}\big(\nabla_a (G_{AB} \nabla_b \varphi^B) - \tfrac{1}{2} G_{BD,A} \nabla_a \varphi^B \nabla_b \varphi^D\big) = 0 \qquad (2.2)$$

subject to the boundary conditions that the fields are fixed at some initial and final times.

The action can be cast into Hamiltonian form by introducing a foliation

$$X{:}R \times S^1 \to M. \qquad (2.3)$$

The foliation is a one-parameter (denoted τ) family of embeddings of a circle into M. Since the space-time is globally hyberbolic, the map (2.3) will be taken to be a diffeomorphism. In terms of local coordinates X^α and $\sigma \in [0,2\pi]$ on M and S^1, respectively, the foliation is described by specifying the coordinate location of the embedded circles:

$$X^\alpha = X^\alpha(\sigma,\tau). \qquad (2.4)$$

The embedded circles are to serve as Cauchy surfaces for M and hence are spacelike with respect to g_{ab}. This means that the metric γ induced in each leaf X_τ of the foliation

$$\gamma = X_\tau^* g$$

or

$$\gamma(\sigma) := \gamma_{11}(\sigma) = g_{\alpha\beta}(X_\tau(\sigma))X^\alpha{}_{,1}X^\beta{}_{,1} \tag{2.5}$$

is positive definite ($X^\alpha{}_{,1} = \partial X^\alpha/\partial\sigma$). Alternatively, at each embedding there exists a unique timelike normal covector n_α defined via

$$X^\alpha_{,1} n_\alpha = 0, \tag{2.6}$$

$$g^{\alpha\beta} n_\alpha n_\beta = -1, \tag{2.7}$$

where $n^\alpha = g^{\alpha\beta} n_\beta$ is future directed. Through (2.6) and (2.7) n_α is defined as a fixed functional of the embeddings.

Expression (2.4) can be viewed as providing a coordinate transformation between X^α and $\sigma^\alpha = (\tau,\sigma)$. The one-parameter set of embedded circles then arises as the level surfaces $\tau = \text{const}$. Neighboring embeddings, labeled by τ and $\tau + \delta\tau$, are connected by the *deformation vector*

$$N^\alpha := \frac{\partial X^\alpha}{\partial \tau} =: \dot{X}^\alpha. \tag{2.8}$$

We decompose N^α into its components parallel and perpendicular to the embedding:

$$N^\alpha = N^\perp n^\alpha + N^1 X^\alpha_{,1}, \tag{2.9}$$

where

$$N^\perp = -n_\alpha N^\alpha, \tag{2.10}$$

$$N^1 = X^1_\alpha N^\alpha, \tag{2.11a}$$

$$X^1_\alpha := g_{\alpha\beta}\gamma^{11} X^\beta_{,1}. \tag{2.11b}$$

In standard terminology $N^\perp$ is known as the lapse function and N^1 is known as the shift vector. Any tensor field on M may be restricted to a given embedding and likewise decomposed into tangential and normal pieces. In particular, $\nabla_a \varphi^A$ and g_{ab} may be decomposed and pulled back to $R \times S^1$, yielding the hypersurface form of the action:

$$S[\varphi] = \int_{R\times S^1} \frac{1}{2} N^\perp \sqrt{\gamma} G_{AB}(\varphi)[(N^\perp)^{-2}(-\dot{\varphi}^A + \varphi^A{}_{,1}N^1)(-\dot{\varphi}^B + \varphi^B{}_{,1}N^1) - \gamma^{11}\varphi^A{}_{,1}\varphi^B{}_{,1}]. \tag{2.12}$$

The momenta π_A conjugate to φ^A are easily obtained, from which we can pass to the Hamiltonian form of the action:

$$S[\varphi,\pi] = \int_{R\times S^1} (\pi_A \dot{\varphi}^A - N^\perp h_\perp - N^1 h_1), \tag{2.13}$$

where

$$h_\perp = \gamma^{-1/2} h = \gamma^{-1/2}(\tfrac{1}{2} G^{AB}\pi_A\pi_B + \tfrac{1}{2} G_{AB}\varphi^A{}_{,1}\varphi^B{}_{,1}), \tag{2.14}$$

$$h_1 = \pi_A \varphi^A{}_{,1}. \tag{2.15}$$

The functional $S[\varphi,\pi]$ is to be varied with respect to φ^A and π_A with $N^\perp$, N^1, and γ treated as fixed external functions. We will gain some simplicity in what follows if we rescale the lapse function to be a scalar density of weight minus 1 (equivalent to a vector in this one-dimensional case):

$$N := \gamma^{-1/2} N^\perp. \tag{2.16}$$

The action then takes the form

$$S[\varphi,\pi] = \int_{R\times S^1} (\pi_A \dot{\varphi}^A - Nh - N^1 h_1), \tag{2.17}$$

where h is a scalar density of weight 2. The energy and momentum densities (h,h_1) are independent of the metric induced on the embedding. All foliation and metric dependence is concentrated in N and N^1.

A significant feature of the lapse density and shift vector is that they are "Weyl invariant," i.e., invariant under conformal rescalings of the metric g_{ab}. This can be seen from (2.5), (2.7), (2.11), and (2.16). If we perform a conformal rescaling

$$\bar{g}_{ab} = e^{2\omega(X)} g_{ab},$$

then

$$\bar{n}_\alpha = e^\omega n_\alpha, \quad \bar{\gamma}_{11} = e^{2\omega}\gamma_{11}.$$

Hence

$$\bar{N}^1 = \bar{\gamma}^{11}\bar{g}_{\alpha\beta}X^\beta{}_{,1}N^\alpha = N^1,$$

$$\bar{N} = \bar{\gamma}^{-1/2}\bar{N}^\perp = [e^{2\omega}\gamma]^{-1/2}[e^\omega N^\perp] = N.$$

This, of course, is just a reflection of the fact that only the Weyl invariant combination $\sqrt{-g}g^{ab}$ enters the action (2.1).

The Hamiltonian

$$H = \int_{S^1} (Nh + N^1 h_1) \tag{2.18}$$

generates the dynamical flow in the phase space Γ_0, which is the cotangent bundle (with the usual L^2 dual) over the space of smooth maps from the circle into M^d: $\Gamma_0 = T^*C^\infty(S^1,M^d)$. The variables (φ^A,π_B) constitute a canonical chart on Γ_0, with their only nonvanishing Poisson brackets being

$$[\varphi^A(\sigma),\pi_B(\sigma')] = \delta^A_B\delta(\sigma,\sigma'). \tag{2.19}$$

The parametrization process, to be described shortly, allows us to represent the Lie algebra diff(M) on an extended version of Γ_0. However, first we will indicate why Γ_0 alone cannot carry a natural representation of diff(M). To do this it is somewhat more efficient to look at the space of solutions $\mathscr{S}$ of (2.2) rather than Γ_0 directly. Since the Cauchy problem for (2.2) is well posed, we know that every point in Γ_0 uniquely determines a point in $\mathscr{S}$.

Given a solution φ^A to (2.2), the natural action of Diff(M) is by pullback (this is the *right* action). Thus if $\phi \in \text{Diff}(M)$, the induced action on $\mathscr{S}$ is given by

$$\bar{\varphi}^A := \phi^*\varphi^A = \varphi^A \circ \phi. \tag{2.20}$$

Near the identity diffeomorphism, $\bar{\varphi}^A$ and φ^A differ by the Lie derivative along the vector field V^a which generates ϕ:

$$\delta\varphi^A := \bar{\varphi}^A - \varphi^A = \mathscr{L}_V \varphi^A. \tag{2.21}$$

A straightforward computation shows that $\bar{\varphi}^A \in \mathscr{S}$ if and only if

$$\nabla_a(C^{ab}(\vec{V})G_{AB}\nabla_b\varphi^B) - \tfrac{1}{2}C^{ab}(\vec{V})G_{BD,A}\nabla_a\varphi^B\nabla_b\varphi^D = 0,$$

where

$$C^{ab}(\vec{V}) := \nabla^a V^b + \nabla^b V^a - g^{ab}\nabla_c V^c. \tag{2.22}$$

For an arbitrary vector field, (2.22) does not vanish, so that there is no (natural) action of Diff(M) on $\mathscr{S}$ and hence no action on Γ_0. Note that the vanishing of C^{ab} is the requirement that V^a be a conformal Killing vector. Hence we *can* represent on $\mathscr{S}$ the subalgebra conf(M,g) of conformal isometries. This, of course, is just a reflection of the fact that Conf(M,g) is a symmetry group for this Weyl invariant field theory. Moreover, any solution of the conformal Killing equation is completely determined by its restriction to a Cauchy surface (see Appendix A). We therefore expect to be able to realize conf(M,g) on Γ_0. This can be done as follows.

Let v^a denote a conformal Killing vector, thus satisfying

$$C_{ab}(\vec{v}) = 0; \tag{2.23}$$

in the coordinate chart X^α its components are the local functions v^α on M:

$$v^\alpha = v^a(dX^\alpha)_a.$$

Using the embedding $X^\alpha(\sigma)$ we can pull these two functions back to the circle:

$$v^\alpha(\sigma) := v^\alpha(X(\sigma)). \tag{2.24}$$

It is straightforward to verify that the map from conf(M,g) into $C^\infty(\Gamma_0,R)$ given by

$$v \to \vec{v} \to h(\vec{v}),$$

with

$$h(\vec{v}) := \int_{S^1} v^\alpha(\sigma) h_\alpha \tag{2.25}$$

and

$$h_\alpha = -\gamma^{-1/2} n_\alpha h + X^1_{\ \alpha} h_1 \tag{2.26}$$

is an antihomomorphism, i.e., for two conformal Killing vectors v^a and w^a,

$$[h(\vec{v}),h(\vec{w})] = h([\vec{v},\vec{w}]). \tag{2.27}$$

The reason we obtain an *anti*homomorphism stems from the following facts: The Lie bracket is minus the conventional commutator of vector fields[2] and the action of Conf(M,g) on $\mathscr{S}$ is the *right* action. The antihomomorphism depends vitally on the way in which v^a is locked into the metric and the embedding through (2.23) and (2.24), respectively. Relation (2.27) fails for arbitrary vector fields. That $h(\vec{v})$ does in fact represent a symmetry is verified by computing its dynamical evolution via the Hamiltonian (2.18), from which it is seen that $h(\vec{v})$ is a constant of motion. This result also depends crucially on the fact that $v^\alpha(\sigma)$ is the restriction to an embedding of a conformal Killing vector.

The inability to represent more than the subalgebra conf(M,g) of diff(M) is easily traced to the fact that we are not letting Diff(M) act on the space-time metric. In the Hamiltonian framework, g_{ab} is a fixed functional of the embeddings, which are themselves dynamically inert. The parametrization procedure allows Diff(M) to act on the metric by adjoining the embeddings (and their conjugate momenta) to the phase space and then utilizing the natural left action of Diff(M) on the embeddings.

Thus let $\phi \in$Diff(M) and X be an embedding, i.e.,

$$\phi: M \to M, \quad X: S^1 \to M.$$

Then we have

$$\bar{X}: S^1 \to M$$

via

$$\bar{X} = \phi \circ X. \tag{2.28}$$

Of course, this action of Diff(M) is not guaranteed to preserve the spacelike character of the embedding. Given an embedding, there will always be some diffeomorphism which yields a nonspacelike embedding. Since we aim to work with a phase space representation of the Cauchy problem, we are obliged to use initial data placed on spacelike surfaces only. Hence we certainly cannot expect to represent the full diffeomorphism group on (the extension of) Γ_0. By the same token, there will always exist an open neighborhood of the identity in Diff(M) which *will* map spacelike circles into spacelike circles. We can therefore aspire to represent the Lie algebra diff(M).

The parametrization process is facilitated by writing the phase space action (2.17) in the form

$$S[\varphi,\pi] = \int_{R\times S^1} (\pi_A \dot{\varphi}^A - \dot{X}^\alpha h_\alpha), \tag{2.29}$$

where h_α is given in (2.26). We now adjoin $X^\alpha(\sigma)$ and its conjugate momenta $P_\alpha(\sigma)$ to the phase space, so that we work in an extended phase space $\Gamma_0' = T^*[C^\infty(S^1,M^d) \times Emb_g(S^1,M)]$, where the subscript g indicates that the embeddings are to be spacelike with respect to g_{ab}. The only nonvanishing Poisson brackets between $X^\alpha(\sigma)$ and $P_\beta(\sigma)$ are

$$[X^\alpha(\sigma),P_\beta(\sigma')] = \delta^\alpha_\beta \delta(\sigma,\sigma').$$

Since the embedding "velocity" $\dot{X}^\alpha$ enters (2.29) linearly, the definition of P_α is a constraint:

$$\mathbf{h}_\alpha := P_\alpha + h_\alpha \approx 0. \tag{2.30}$$

Next, we adjoin to the action the constraints, smeared with the Lagrange multipliers N^α, to yield the phase space action on Γ_0':

$$S[\varphi,\pi,X,P,\vec{N}] = \int_{R\times S^1} (\pi_A \dot{\varphi}^A + P_\alpha \dot{X}^\alpha - N^\alpha \mathbf{h}_\alpha). \tag{2.31}$$

The action is now to be varied freely with respect to all of its arguments. The equations of motion associated with (2.31) for (φ^A,π_B) are the usual ones; they are equivalent to (2.2): In Hamiltonian form they are

$$\dot{\varphi}^A = [\varphi^A,\mathbf{h}(\vec{N})], \quad \dot{\pi}_A = [\pi_A,\mathbf{h}(\vec{N})],$$

where

$$\mathbf{h}(\vec{N}) = \int_{S^1} N^\alpha \mathbf{h}_\alpha. \tag{2.32}$$

We also have

$$\dot{X}^\alpha = [X^\alpha,\mathbf{h}(\vec{N})] = N^\alpha,$$

which justifies our use of the symbol N^α for the multipliers. Finally, the Hamilton equations for P_α read as

$$\dot{P}_\alpha = [P_\alpha,\mathbf{h}(\vec{N})] = -\frac{\delta}{\delta X^\alpha}\mathbf{h}(\vec{N}).$$

The X dependence of $\mathbf{h}(\vec{N})$ comes from the factors $\gamma^{-1/2}n_\alpha$ and $X^1_{\ \alpha}$ in (2.26). From the constraints (2.30) we see that the embedding momenta P_α are combinations of the energy and momentum densities h and h_1. The equations of motion for P_α just amount to the corresponding conservation laws.

Alternatively, these equations guarantee that the constraints (2.30) are preserved in time; this implies that they are *first class* in the terminology of Dirac. In fact, it is straightforward to verify that the Poisson algebra of the functions $\mathbf{h}(\vec{N})$ is *Abelian*:

$$[\mathbf{h}(\vec{N}),\mathbf{h}(\vec{M})] = 0 \tag{2.33}$$

even off the constraint surface (2.30) (by definition, the first class nature of the constraints guarantees (2.33) on the constraint surface).

To recover the constraint functions that generate the non-Abelian algebra of hypersurface deformations, we project the functions $\mathbf{h}_\alpha$ perpendicular and parallel to a given embedding:

$$\mathbf{h} := \gamma^{1/2} n^\alpha \mathbf{h}_\alpha, \tag{2.34}$$

$$\mathbf{h}_1 := X^\alpha_{,1} \mathbf{h}_\alpha. \tag{2.35}$$

When smeared with a lapse density, $\mathbf{h}$ generates the dynamical evolution associated with a normal deformation of the embeddings. Likewise, $\mathbf{h}_1$ generates the action of Diff(S^1) on the embeddings and dynamical variables. It will be useful in Sec. IV to have the projected functions written in terms of embeddings registered in a conformal coordinate system. Thus let $T\in(-\infty,\infty)$ and $S\in[0,2\pi]$ be coordinates on M such that the metric takes the form

$$g_{TT} = -g_{SS}, \quad g_{ST} = 0.$$

If we denote $X^\alpha(\sigma) = (T(\sigma),S(\sigma))$ and $P_\alpha = (P_T,P_S)$ we then have

$$\mathbf{h} = P_T S_{,1} + P_S T_{,1} + \tfrac{1}{2}(G^{AB}\pi_A\pi_B + G_{AB}\varphi^A{}_{,1}\varphi^B{}_{,1}), \tag{2.36}$$

$$\mathbf{h}_1 = P_T T_{,1} + P_S S_{,1} + \pi_A \varphi^A{}_{,1}. \tag{2.37}$$

Relation (2.33) allows us to implement diff(M) on Γ_0' in an analogous fashion to our previous representation of conf(M,g). Given any (complete) vector field $\vec{U}$ on M, we again form two scalars via the coordinate one-form basis. We then pull these functions back along the embedding and use the resulting functions on S^1 to smear $\mathbf{h}_\alpha$:

$$\mathbf{h}(\vec{U}) = \int_{S^1} U^\alpha(X(\sigma))\mathbf{h}_\alpha. \tag{2.38}$$

Using (2.33) it follows immediately that the map from diff(M) into $C^\infty(\Gamma_0',R)$ given by

$$U \to \vec{U} \to \mathbf{h}(\vec{U})$$

is a homomorphism. [Here we obtain a homomorphism due to the fact that the action of Diff(M) on the embeddings is the *left* action.]

$$[\mathbf{h}(\vec{U}),\mathbf{h}(\vec{V})] = \mathbf{h}(-[\vec{U},\vec{V}]). \tag{2.39}$$

Physically, we can think of the action of diff(M) on Γ_0' as follows. Decompose $\mathbf{h}(\vec{V})$ as

$$\mathbf{h}(\vec{V}) = \int_{S^1} V^\alpha(\sigma)P_\alpha + \int_{S^1} V^\alpha(\sigma)h_\alpha := P(\vec{V}) + h(\vec{V}).$$

The function $P(\vec{V})$ acts on $T^*Emb_g(S^1,M)$ with the effect of displacing the embedded circle along the deformation vector $\vec{V}$ by the infinitesimal version of (2.28). The function $h(\vec{V})$ then evolves the canonical data in the correct dynamical fashion associated with the Hamiltonian equations. Proceeding along the integral curves of $\vec{V}$, one builds a solution to the Hamiltonian equations with lapse and shift defined by the projections of $\vec{V}$.

B. The closed string

The action functional for the massless relativistic string can be obtained from (2.1) by taking the target space for the harmonic maps to be d-dimensional Minkowski space. Conventionally, one uses an inertial coordinate system x^μ, $\mu = 0,\ldots,d-1$ on M^d in the role of φ^A to yield

$$S[g,x] = -\frac{1}{2}\int_M \sqrt{-g}\, g^{ab}\eta_{\mu\nu}\nabla_a x^\mu \nabla_b x^\nu. \tag{2.40}$$

Aside from the change in notation, the primary difference between the variational principles associated with (2.1) and (2.40) is that in the former we extremize the action with the metric treated as given, i.e., fixed. In the variational principle for the string we vary both x^μ and g_{ab}. As a result, when we extremize (2.40) we obtain two sets of equations:

$$\frac{\delta S}{\delta x^\mu} = \sqrt{-g}\, g^{ab}\eta_{\mu\nu}\nabla_a\nabla_b x^\nu = 0 \tag{2.41}$$

and

$$\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g_{ab}} = \eta_{\mu\nu}\left(\nabla^a x^\mu \nabla^b x^\nu - \frac{1}{2} g^{ab} g^{cd} \nabla_c x^\mu \nabla_d x^\nu\right) = 0. \tag{2.42}$$

As before, (2.41) is the equation for a harmonic map $x: M \to M^d$. However, the metric g_{ab} is now coupled to the induced metric on the world sheet through (2.42). Specifically, (2.42) tells us that the Weyl invariant parts of the two metrics are equal:

$$g_{ab} = e^{2\omega}(x^*\eta)_{ab}.$$

Here ω is a function on M which cannot be determined, owing to the Weyl invariance of the theory. In contrast to the previous situation, the system of equations (2.41) and (2.42) is degenerate in the sense that the equations obey an infinite set of ("Bianchi") identities

$$\int_M \left(\frac{\delta S}{\delta x^\mu}\mathscr{L}_V x^\mu + \frac{\delta S}{\delta g_{ab}}\mathscr{L}_V g_{ab}\right) = 0, \tag{2.43}$$

where V^a is any vector field on M with compact support. The identities (2.43) are a direct consequence of the invariance of the action with respect to variations of (x^μ, g_{ab}) induced by diffeomorphic mappings of M onto itself. As a direct consequence of the diffeomorphism invariance of the action we have the following immediate result: If (x,g) is a solution to the equations of motion, then for $\phi\in$Diff(M), (ϕ^*x,ϕ^*g) is also a solution. The identities (2.43), while complicating the Cauchy problem for (2.41) and (2.42), thus serve to indicate that the space of solutions to the equations of motion admits an action of Diff(M). From our earlier discussion we also know that Conf(M,g) is a *symmetry* group for the system: Given a solution (x,g) and $\psi\in$Conf(M,g), both (ψ^*x,g) and (x,ψ^*g) will be solutions [of course, we can also let ψ act on the string and metric variables simultaneously, but this is just a special case of the Diff(M) action].

The Hamiltonian formulation for the string is obtained in exactly the same fashion as for the harmonic maps of Sec. II A. Again, the important new feature that enters is that the metric is now freely variable. The phase space Γ can be taken to be the cotangent bundle over the space of smooth embeddings of a circle into M^d which are spacelike with respect to the target space metric $\eta_{\mu\nu}$:

$$\Gamma = T^*[Emb_\eta(S^1,M^d)].$$

A canonical coordinate patch is given by the pair (x^μ,p_ν):

$$[x^\mu(\sigma),p_\nu(\sigma')] = \delta^\mu_\nu\delta(\sigma,\sigma').$$

The action functional on $R\times S^1$ takes the form

$$S[x,p,\vec{N}] = \int_{R\times S^1}(p_\mu\dot{x}^\mu - N\mathscr{H} - N^1\mathscr{H}_1), \qquad (2.44)$$

where

$$\mathscr{H} = \tfrac{1}{2}(\eta^{\mu\nu}p_\mu p_\nu + \eta_{\mu\nu}x^\mu{}_{,1}x^\nu{}_{,1}) \qquad (2.45)$$

and

$$\mathscr{H}_1 = p_\mu x^\mu{}_{,1}. \qquad (2.46)$$

From the Hamiltonian point of view, the distinction between (2.13) and (2.44) is that in the former case the lapse and shift are fixed, externally prescribed functions on S^1, while in the present case they are independently variable: Thus they serve as Lagrange multipliers enforcing the constants

$$\mathscr{H}\approx 0, \qquad (2.47)$$

$$\mathscr{H}_1\approx 0. \qquad (2.48)$$

Note that the constraints (2.47) and (2.48) arise in an entirely different manner than those for the parametrized scalar fields: There, the constraints appear after adjoining a nondynamical set of variables (the embeddings and their conjugate momenta) to the phase space. Here, no adjoining is necessary and, as one often says, the theory is "already parametrized."

The Hamiltonian

$$H(\vec{N}) := \int_{S^1}(N\mathscr{H} + N^1\mathscr{H}_1) \qquad (2.49)$$

generates the dynamical flow, which depends on a choice for the lapse and shift:

$$\dot{x}^\mu = [x^\mu,H(\vec{N})] = Np^\mu + N^1x^\mu{}_{,1}, \qquad (2.50)$$

$$\dot{p}_\mu = [p_\mu,H(\vec{N})] = \partial_1(Nx_{\mu,1} + N^1p_\mu). \qquad (2.51)$$

From (2.50) we see that the *super-Hamiltonian*

$$H(N) = \int_{S^1}N\mathscr{H} \qquad (2.52)$$

generates the dynamical evolution associated with a normal deformation of the embeddings. The *supermomentum*

$$H(N^1) = \int_{S^1}N^1\mathscr{H}_1 \qquad (2.53)$$

generates the motion in Γ associated with a tangential deformation of the circles. This, of course, is just an action of Diff(S^1) on the phase space which is generated by the vector N^1.

The constraints (2.47) and (2.48) are complete in the sense that if they are satisfied on some initial embedding, i.e., at some initial τ, then they are satisfied everywhere along the dynamical flow (2.50) and (2.51). This is guaranteed by the "Dirac algebra" which the constraint functions satisfy:

$$[H(N),H(M)] = H(L^1), \qquad (2.54)$$

$$[H(N),H(M^1)] = H(K), \qquad (2.55)$$

$$[H(N^1),H(M^1)] = H(J^1), \qquad (2.56)$$

where

$$L^1 = N\partial_1M - M\partial_1N,\quad K = N\partial_1M^1 - M^1\partial_1N,$$
$$J^1 = N^1\partial_1M^1 - M^1\partial_1N^1.$$

In contrast to general relativity or parametrized theories on backgrounds of dimension greater than 2, the Dirac algebra in two dimensions is a Lie algebra. This algebra is in fact isomorphic to diff$(S^1)\oplus$diff(S^1) [which in turn is isomorphic to conf(M,g); see Appendix A], as can be seen by rearranging the constraints as follows:

$$_{(\pm)}\mathscr{H} := \tfrac{1}{2}(\mathscr{H}\pm\mathscr{H}_1)\approx 0. \qquad (2.57)$$

When smeared with vector fields on S^1, the constraint algebra becomes

$$[_{(\pm)}H(N^1),_{(\pm)}H(M^1)] = \pm_{(\pm)}H(\mathscr{L}_{N^1}M^1), \qquad (2.58a)$$

$$[_{(+)}H(N^1),_{(-)}H(M^1)] = 0, \qquad (2.58b)$$

which is diff$(S^1)\oplus$diff(S^1). Note that the equivalence between vectors and scalar densities of weight minus 1, which occurs only in one dimension, is crucial for the existence of the algebraic structure just described.

We are now in a position to elaborate on the work to follow. We are aspiring to find a homomorphic mapping from diff(M) into the Poisson algebra of functions $C^\infty(\Gamma,R)$. Obviously, the action of Diff(M) on Γ should lead to the appropriate dynamical evolution as we move from point to point in M. One therefore expects that the natural place to look for such a realization of diff(M) is the constraint functions (2.52) and (2.53). Indeed, as we have seen, these functions do form a Lie algebra. There are problems, however. First, the algebra is the wrong algebra [diff$(S^1)\oplus$diff(S^1) rather than diff(M)]. More important, however, is the fact that there is no (obvious) way to link phase space quantities, defined on S^1, to space-time quantities, in particular vector fields on M. The homomorphism obtained via the parametrization procedure of Sec. II A relied heavily on the fact that the metric on M was known in advance. This allowed for the canonical–covariant connection to be obtained since the metric (used to define projections) was a fixed, known functional of the embeddings. In the present situation, the metric is effectively a part of the configuration space (albeit a nondynamical part) to be varied independent of the embedding. The metric and foliation are now decoupled and in fact, the embeddings no longer directly feature in the formalism. Consequently, a direct mimicry of the procedure used in Sec. II A is impossible. Based on these considerations, two courses of action suggest themselves. One can try to reestablish the coupling between the metric and embedding by a form of "gauge fixing." This line of attack, originally devised for geometrodynamics by Isham and Kuchař,[2] is the subject of Sec. III. Alternatively, one can try to take advantage of the notion that the string is in some sense "already parametrized." In Sec. IV we investigate the degree to which one can extract the embedding variables (and their conjugate momenta) directly from the string phase space Γ. Success in either of these approaches leads to the desired representation of diff(M) through the techniques of Sec. II A.

Finally, we point out (with some irony intended) that the same difficulties arise in trying to homomorphically map conf(M,g) into the Poisson algebra of functions on Γ. Granted, the phase space does carry the Lie algebra diff$(S^1)\oplus$diff(S^1), which is isomorphic to conf(M,g), as seen in (2.58). However, the lack of any connection between phase space functions and space-time (conformal Killing) vectors obstructs any attempt at finding a homomorphic mapping. Nevertheless, the fact that the two-dimensional algebra of hypersurface deformations is isomorphic to conf(M,g) does suggest a fundamental *dynamical* role for the conformal group. We will elucidate this point in Sec. VI.

III. REPRESENTATION OF diff(*M*) ON EXTENDED PHASE SPACES

Here we shall follow the strategy of Ref. 2. The basic idea is to recouple the metric on M to the embeddings through gauge-fixing conditions. For aesthetic reasons, we will use an "auxiliary structure" to fix the gauge, so that we always deal with geometrically well-defined objects. We will, however, occasionally pause to make contact with the possibly more familiar coordinate-dependent expressions.

The auxiliary structure that we will need is simply a fixed foliation

$$\widetilde{Y}:R\times S^1\to M$$

(which will be a diffeomorphism) and its inverse, denoted Y:

$$Y:=\widetilde{Y}^{-1},\quad Y:M\to R\times S^1.$$

If we fix once and for all local coordinates $\sigma^{(\alpha)}$ on $R\times S^1$,

$$\sigma^{(\alpha)}:R\times S^1\to R^2,$$

we can compose them with the map Y to obtain local coordinates $Y^{(\alpha)}$ on M:

$$Y^{(\alpha)}=\sigma^{(\alpha)}\circ Y,\quad Y^{(\alpha)}:M\to R^2.$$

While a choice of $\widetilde{Y}$ is necessarily *ad hoc* (we have an infinity of possibilities), we are to think of having made a definite choice. The gauge-fixing procedure will amount to fixing two components of g_{ab} with respect to the given auxiliary structure. This structure will then serve to provide the needed link between the metric and the embeddings which we introduce to construct the Hamiltonian formulation. At any point along the way, we can put a given expression into a more conventional form by choosing coordinates X^α on M, which are adapted to the foliation $\widetilde{Y}$ via

$$X^\alpha=Y^{(\alpha)}.$$

A. Conformal gauge

The use of the conformal gauge is ubiquitous in string theory. Among its virtues we have the simplicity of the resulting dynamical equations and the fact that one can always find conformal coordinates which cover all of M except for a set of measure zero.

In terms of the auxiliary structure Y, the conformal gauge conditions take the form

$$g^{ab}\delta_{(\alpha)(\beta)}\nabla_a Y^{(\alpha)}\nabla_b Y^{(\beta)}=0, \tag{3.1}$$

$$g^{ab}e_{(\alpha)(\beta)}\nabla_a Y^{(\alpha)}\nabla_b Y^{(\beta)}=0, \tag{3.2}$$

where

$$e_{(\alpha)(\beta)}=\begin{pmatrix}1&1\\0&0\end{pmatrix}.$$

If we denote by $\widetilde{\gamma}$ the metric induced on the leaves of the foliation,

$$\widetilde{\gamma}=\widetilde{Y}^*g,$$

then the metric can be expressed as a functional of $\widetilde{\gamma}$ and Y via

$$g_{ab}=\widetilde{\gamma}\eta_{(\alpha)(\beta)}\nabla_a Y^{(\alpha)}\nabla_b Y^{(\beta)}. \tag{3.3}$$

As summarized in Appendix A, every metric on M is conformal to a flat metric. For an arbitrary auxiliary structure, (3.3) just serves to capture that fact, i.e., (3.3) is simply an invariant parametrization of the space of metrics. By *fixing* the auxiliary structure we are fixing the flat metric to which g_{ab} is conformally related. The only remaining "degree of freedom" is the conformal factor $\widetilde{\gamma}$. Weyl invariance for the classical string will guarantee that this remaining metric degree of freedom does not play a role in the theory.

If we choose coordinates adapted to the foliation as discussed above, the gauge-fixing conditions (3.1) and (3.2) take the familiar form

$$g^{00}=-g^{11},\quad g^{01}=0.$$

Alternatively, (3.3) reads as

$$g_{\alpha\beta}=\widetilde{\gamma}\eta_{(\alpha)(\beta)}$$

in the adapted coordinates $X^\alpha=Y^{(\alpha)}$.

We can now substitute the metric (3.3) into the string action (2.40). The conformal degree of freedom $\widetilde{\gamma}$ drops out of the action. This leaves an action that is a functional of the fixed auxiliary structure and the string variables. Only the latter quantities are to be varied; Eqs. (2.47) and (2.48) are now suspended and we have arrived back at the formalism for the harmonic maps of Sec. II A.

Following the Hamiltonian treatment of the parametrized harmonic maps, we now work in an extended phase space Γ' for the string, which is given by

$$\Gamma'=T^*[Emb_\eta(S^1,M^d)\times Emb_g(S^1,M)].$$

The corresponding action functional becomes

$$S[x,p,X,P,\vec{N}]=\int_{R\times S^1}(p_\mu\dot{x}^\mu+P_\alpha\dot{X}^\alpha-N^\alpha\mathbf{H}'_\alpha), \tag{3.4}$$

where

$$\mathbf{H}'_\alpha:=P_\alpha+\mathscr{H}_\alpha\approx 0 \tag{3.5}$$

and

$$\mathscr{H}_\alpha = -\gamma^{-1/2} n_\alpha \mathscr{H} + X^1_\alpha \mathscr{H}_1, \tag{3.6}$$

with $\mathscr{H}$ and $\mathscr{H}_1$ given by (2.45) and (2.46). In (3.6) we have defined $\gamma = X^* g$, with g_{ab} given by (3.3). Note that the induced metric (γ or $\tilde{\gamma}$) dutifully drops out of the functions (3.6), i.e.,

$$\mathscr{H}_\alpha = \mathscr{H}_\alpha [x,p,X].$$

This can be verified using the definition of n_α [Eqs. (2.6) and (2.7)] and X^1_α [Eq. (2.11b)] in conjunction with the gauge-fixed metric (3.3). We therefore need not include γ (or $\tilde{\gamma}$) in the phase space. Alternatively, the equation of motion for γ,

$$\frac{\delta S}{\delta \gamma} = 0,$$

is identically satisfied.

The constraints (3.5) obey an Abelian Poisson algebra; consequently, we can immediately construct the desired Lie algebra homomorphism. For V, $W \in \mathrm{diff}(M)$, the map

$$V \to \vec{V} \to \mathbf{H}'(\vec{V}),$$

where

$$\mathbf{H}'(\vec{V}) = \int_{S^1} V^\alpha(X(\sigma)) \mathbf{H}'_\alpha, \tag{3.7}$$

yields

$$[\mathbf{H}'(\vec{V}), \mathbf{H}'(\vec{W})] = \mathbf{H}'(-[\vec{V},\vec{W}]). \tag{3.8}$$

The symmetry algebra conf(M,g) can also be represented: To represent it as a subalgebra of diff(M) we use

$$v \to \vec{v} \to \mathbf{H}'(\vec{v}),$$

where $\vec{v}$ is a conformal Killing vector for the metric (3.3). Note that such vectors are in fact independent of $\tilde{\gamma}$. This representation emphasizes the role of conf(M,g) as a *dynamical* subalgebra of diff(M). As discussed earlier, Conf(M,g) also acts as a *symmetry* group on the space of solutions to the string dynamical equations. In Hamiltonian language, this representation is implemented via the following two mappings:

$$v \to \vec{v} \to P(\vec{v}), \quad P(\vec{v}) = \int_{S^1} v^\alpha P_\alpha, \tag{3.9}$$

and

$$v \to \vec{v} \to \mathscr{H}(\vec{v}), \quad \mathscr{H}(\vec{v}) = \int_{S^1} v^\alpha \mathscr{H}_\alpha. \tag{3.10}$$

The map (3.9) is a homomorphism from conf(M,g) into the Poisson algebra of functions on Γ':

$$[P(\vec{v}), P(\vec{w})] = P(-[\vec{v},\vec{w}]),$$

while (3.10) is an antihomomorphism

$$[\mathscr{H}(\vec{v}), \mathscr{H}(\vec{w})] = \mathscr{H}([\vec{v},\vec{w}]).$$

As in Sec. II A, the functions $\mathscr{H}(\vec{w})$ are indeed symmetry generators since they are constants of motion. Thus for any vector $\vec{V}$ and any conformal Killing vector $\vec{w}$,

$$[\mathscr{H}(\vec{w}), \mathbf{H}'(\vec{V})] = 0.$$

The functions $P(\vec{w})$ are generators of a *conditional symmetry*, i.e., they are conserved modulo the constraints (3.5):

$$[P(\vec{w}), \mathbf{H}'(\vec{V})] = \mathbf{H}'(-[\vec{w},\vec{V}]) \approx 0.$$

In exactly the same manner as for the parametrized fields, the functions $\mathbf{H}'(\vec{V})$ act on Γ' by displacing the embedding along the deformation vector $\vec{V}$ which generates the diffeomorphism, while simultaneously evolving the string variables along the flow lines of $\vec{V}$. It is important to realize that while we use a conformal coordinate system (in the guise of the auxiliary structure) to describe the kinematics of the dynamical evolution, we are *not* evolving with some fixed (conformal) foliation. In other words, we still evolve from one arbitrary Cauchy surface to another. The choice of deformation vector chooses the slicing. Having chosen a deformation vector we can, if we like, use the gauge-fixed metric to extract the corresponding lapse and shift. The string evolution is then governed by the following:

$$\dot{x}^\mu = [x^\mu, \mathbf{H}'(\vec{N})] = Np^\mu + N^1 x^\mu{}_{,1}, \tag{3.11}$$

$$\dot{p}^\mu = [p^\mu, \mathbf{H}'(\vec{N})] = \partial_1 (N x_{\mu,1} + N^1 p_\mu). \tag{3.12}$$

Despite the fact that (N, N^1) obtained from $\vec{V}$ are functionals of the embeddings, Eqs. (3.11) and (3.12) are identical to the usual string equations of motion (2.50), (2.51); the evolution within the extended phase space is still the correct one. This occurs because the induced metric γ does not play a role in the formalism. Moreover, even if the metric on M were defined to be the metric induced by virtue of the string embeddings x^μ, thus rendering γ a functional of the string variables, Weyl invariance would again guarantee that the equations of motion would be correct. (This should be contrasted with the case of geometrodynamics,[2] where the metric *is* dynamical and the correct equations of motion are obtained only weakly, i.e., upon imposing the original Hamiltonian and momentum constraints.) On the other hand, the solutions to (3.11) and (3.12) are not in general solutions to the full set of string equations (2.41) and (2.42). To obtain the physical string solutions we must restrict the admissible initial data in the string sector of Γ' to satisfy the original constraints (2.47) and (2.48). Thus the physically relevent dynamical evolution begins in the subspace $\bar{\Gamma}'$ of Γ' specified by (2.47), (2.48), and (3.5). If initial data are chosen from this subspace, it can be verified that the dynamical evolution remains in $\bar{\Gamma}'$. The verification amounts to checking that the complete set of constraint functions $(\mathscr{H}, \mathscr{H}_1, \mathbf{H}'_\alpha)$ is first class. We have already proven a large part of this in Eqs. (2.54)–(2.56) and (3.8). It only remains to be checked that the original string constraints have weakly vanishing Poisson brackets with the Hamiltonian $\mathbf{H}'(\vec{N})$. Conservation of the Hamiltonian constraint is guaranteed since

$$[\mathscr{H}(\sigma), \mathbf{H}(\vec{N})] = \left[\mathscr{H}(\sigma), \int_{S^1} N^\alpha(-\gamma^{-1/2} n_\alpha \mathscr{H} + X^1_\alpha \mathscr{H}_1)\right].$$

This expression weakly vanishes because the string constraints obey the Dirac algebra, in particular (2.54) and (2.55) [see, also (5.3)]. Similarly,

$$[\mathscr{H}_1(\sigma), \mathbf{H}'(\vec{N})] = \left[\mathscr{H}_1(\sigma), \int_{S^1} N^\alpha(-\gamma^{-1/2} n_\alpha \mathscr{H} + X^1_\alpha \mathscr{H}_1)\right],$$

which weakly vanishes as a result of (2.55) and (2.56) [see, also (5.4)].

To summarize, we have represented the Lie algebra diff(M) at the expense of working in an extended phase space Γ'. The representatives $\mathbf{H}'(\vec{V})$ of the vector fields $\vec{V}$ on M correctly evolve the initial data. To obtain physical string dynamics, we must restrict the admissible initial data to lie in the subspace $\bar{\Gamma}' \subset \Gamma'$. If the dynamical evolution begins in $\bar{\Gamma}'$, it will stay there.

As mentioned earlier, one of the primary advantages of the conformal gauge lies in the ease with which one can solve the equations of motion. Let us conclude our discussion of the conformal gauge by outlining how this occurs in the Hamiltonian formulation based on the extended phase space Γ'.

We begin by observing that the evolution equations for the string are equivalent to a system of functional differential equations.[6] To see this, note that any function on the phase space $F \in C^\infty(\Gamma', R)$ is dynamically evolved according to

$$\dot{F} = [F, \mathbf{H}'(\vec{N})]. \tag{3.13}$$

The dependence of the evolution on the choice of foliation resides in the deformation vector N^α, which is freely specifiable. Both $\mathbf{H}'(\vec{N})$ and $\dot{F}$ are linear functionals of N^α:

$$\mathbf{H}'(\vec{N}) = \int_{S^1} N^\alpha \mathbf{H}'_\alpha,$$

$$\dot{F} = \int_{S^1} \dot{X}^\alpha \frac{\delta F}{\delta X^\alpha} = \int_{S^1} N^\alpha \frac{\delta F}{\delta X^\alpha}.$$

Using the fact that N^α is arbitrary, an equation equivalent to (3.13) is given by

$$\frac{\delta F}{\delta X^\alpha(\sigma)} = [F, \mathbf{H}'_\alpha(\sigma)]. \tag{3.14}$$

In particular, we can let F be one of the canonical coordinates on Γ', thereby establishing a set of functional differential, "many-fingered time" Hamiltonian equations.

We know that the equations of motion will be simple in conformal coordinates, so we choose our coordinates X^α on M to be adapted to the conformal foliation. In particular, define

$$\mathrm{X}^0 = Y^{(0)} =: \mathrm{T}, \quad \mathrm{X}^1 = Y^{(1)} =: \mathrm{S}, \quad \mathrm{X}^\pm := \mathrm{T} \pm \mathrm{S}.$$

Here, T is a coordinate on R^1, S is a coordinate on S^1, and $\mathrm{X}^\pm$ are the associated null coordinates with respect to the metric (3.3). In terms of the latter set of coordinates, the gauge-fixed metric has components given by

$$g_{++} = 0 = g_{--}, \quad g_{+-} = -\tfrac{1}{2}\tilde{\gamma}. \tag{3.15}$$

Given an embedding $X^\alpha(\sigma)$, the metric γ induced on it is

$$\gamma_{11}(\sigma) = -\tilde{\gamma}(\sigma) X^+{}_{,1} X^-{}_{,1}. \tag{3.16}$$

The embedding is to be spacelike, which implies

$$X^+{}_{,1} X^-{}_{,1} < 0. \tag{3.17}$$

The unit normal covector to the embedding has components in the null coordinate basis given by

$$n_\pm = \pm \tfrac{1}{2}\tilde{\gamma}^{1/2}(-X^+{}_{,1}X^-{}_{,1})^{-1/2} X^\mp{}_{,1}. \tag{3.18}$$

Finally, the covector X^1_α (on M), defined in (2.11b), has the components

$$X^1_\pm = \tfrac{1}{2}(X^\pm{}_{,1})^{-1}. \tag{3.19}$$

Assembling the expressions (3.15)–(3.19) together, the many-fingered time Hamiltonian has components given by

$$\mathbf{H}'_+ = P_+ + (X^+{}_{,1})^{-1}\tilde{\alpha}_\mu \tilde{\alpha}^\mu, \tag{3.20a}$$

$$\mathbf{H}'_- = P_- - (X^-{}_{,1})^{-1}\alpha_\mu \alpha^\mu, \tag{3.20b}$$

where we have introduced the usual notation

$$\tilde{\alpha}^\mu := \tfrac{1}{2}(p^\mu + x^\mu{}_{,1}), \tag{3.21}$$

$$\alpha^\mu := \tfrac{1}{2}(p^\mu - x^\mu{}_{,1}). \tag{3.22}$$

Note that α^μ and $\tilde{\alpha}^\mu$ are scalar densities of weight 1 (covectors) on S^1.

By substituting each of the canonical variables into the functional Hamiltonian equation (3.14), we obtain

$$\frac{\delta P_+(\sigma)}{\delta X^+(\sigma')} = (X^+{}_{,1}(\sigma'))^{-2}\tilde{\alpha}_\mu(\sigma')\tilde{\alpha}^\mu(\sigma')\partial_\sigma \delta(\sigma,\sigma'), \tag{3.23}$$

$$\frac{\delta P_-(\sigma)}{\delta X^-(\sigma')} = -(X^-{}_{,1}(\sigma'))^{-2}\alpha_\mu(\sigma')\alpha^\mu(\sigma')\partial_\sigma \delta(\sigma,\sigma'), \tag{3.24}$$

$$\frac{\delta P_\pm(\sigma)}{\delta X^\mp(\sigma')} = 0, \tag{3.25}$$

$$\frac{\delta X^\alpha(\sigma)}{\delta X^\beta(\sigma')} = \delta^\alpha_\beta\, \delta(\sigma,\sigma'), \tag{3.26}$$

$$\frac{\delta x^\mu(\sigma)}{\delta X^+(\sigma')} = (X^+{}_{,1})^{-1}\tilde{\alpha}^\mu \delta(\sigma,\sigma'), \tag{3.27}$$

$$\frac{\delta x^\mu(\sigma)}{\delta X^-(\sigma')} = -(X^-{}_{,1})^{-1}\alpha^\mu \delta(\sigma,\sigma'), \tag{3.28}$$

$$\frac{\delta p_\mu(\sigma)}{\delta X^+(\sigma')} = (X^+{}_{,1}(\sigma'))^{-1}\tilde{\alpha}^\mu(\sigma')\partial_\sigma \delta(\sigma,\sigma'), \tag{3.29}$$

$$\frac{\delta p_\mu(\sigma)}{\delta X^-(\sigma')} = (X^-{}_{,1}(\sigma'))^{-1}\alpha_\mu(\sigma')\partial_\sigma \delta(\sigma,\sigma'). \tag{3.30}$$

Equations (3.23)–(3.30) are to be solved so that they match an initial data set $(\bar{X}^\alpha, \bar{P}_\alpha, \bar{x}^\mu, \bar{p}_\mu)$. The initial data in turn must satisfy the constraints (2.47), (2.48), and (3.5). We do not actually have to solve Eqs. (3.23)–(3.26) directly. The equations for $X^\alpha(\sigma)$ enter as identities and the need to solve them for P_α is obviated once we have imposed the constraints (3.5), which fix the embedding momenta as functionals of the string variables and the embeddings. This can be done in terms of the initial data and we are guaranteed that the constraints will continue to hold throughout the evolution. Thus the initial data set can be viewed as a specification of data for the string variables on some initial Cauchy surface specified by $\bar{X}^\alpha$.

Although the remaining equations (3.27)–(3.30) may look rather formidable, their general solution can be found quite easily from the following considerations. A solution to (3.27)–(3.30) amounts to a specification of the values of the string variables on a Cauchy surface chosen by the embedding variables. We obtain a solution for x^μ by restricting the known solution in conformal coordinates to this embedding. The solutions for the remaining variables, p_μ in particular, can then be extracted from the functional Hamiltonian equations.

The general space-time solution for x^μ is given by

$$x^\mu(X^+,X^-) = f^\mu(X^+) + g^\mu(X^-),$$

where

$$f^\mu(X^+ + 2\pi) + g^\mu(X^- - 2\pi) = f^\mu(X^+) + g^\mu(X^-).$$

Upon restriction of this solution to an embedding, we obtain the canonical solution as a function of σ and a functional of X:

$$x^\mu[X](\sigma) = f^\mu(X^+(\sigma)) + g^\mu(X^-(\sigma)). \tag{3.31}$$

We substitute (3.31) into (3.27) or (3.28) and extract the solution for the string momenta

$$p_\mu[X](\sigma) = \eta_{\mu\nu}[X^+{}_{,1}f^\nu{}_{,+}(X^+(\sigma)) - X^-{}_{,1}g^\nu{}_{,-}(X^-(\sigma))]. \tag{3.32}$$

The solutions for x^μ and p_μ are to be matched to the initial data $(\bar{x}^\mu, \bar{p}_\mu)$ on the initial embedding $\bar{X}^\alpha$. If desired, the resulting solutions can be expressed in terms of an arbitrary system of coordinates $X^{\alpha'} = X^{\alpha'}(X)$ by performing the point canonical transformation:

$$X^\alpha(\sigma) \to X^{\alpha'}(\sigma) := X^{\alpha'}(X(\sigma)),$$

$$P_\alpha(\sigma) \to P_{\alpha'}(\sigma) = \frac{\partial X^\beta}{\partial X^{\alpha'}}\bigg|_{X = X(\sigma)} P_\beta(\sigma).$$

The string solutions, constructed in the manner outlined above, reflect the underlying diffeomorphism invariance of the theory through the appearance of two arbitrary functions, namely, the embeddings $X^\alpha(\sigma)$. From the perspective of Hamiltonian dynamics on Γ', these functions simply serve to select the arbitrary embedding on which we measure the fields. If we *fix* a one-parameter family of embeddings, i.e., a foliation, with which to describe the dynamics, the solutions to the equations of motion become unique. If we wish to fix only the embedding velocities, then there may still exist residual arbitrariness. For example, in the "conformal gauge"

$$N = 1, \quad N^1 = 0,$$

we still have the freedom to redefine the foliation by the induced action of a conformal isometry.

B. Harmonic gauge

Our interest in using the harmonic gauge to represent diff(M) stems from a question posed in Ref. 2 concerning the feasibility of the corresponding approach in canonical gravity. Recently, this gauge has also attracted some interest in studies of string quantum mechanics and string field theory.[7] From the string perspective, the utility of the harmonic gauge lies in the fact that, like the conformal gauge, it respects both target space Poincaré invariance and the ever important Weyl invariance.

The harmonic gauge can be specified in terms of the auxiliary structure as

$$g^{ab}\nabla_a\nabla_b Y^{(\alpha)} = 0. \tag{3.33}$$

For a fixed choice of Y, Eq. (3.33) is again to be interpreted as a pair of restrictions on allowable metrics. This way of gauge fixing differs from the conformal gauge in that the metric is related to the auxiliary structure through an expression involving (in particular time) derivatives of the metric. Hence one can say that metric fixing occurs *dynamically*. As always, we can put (3.33) into a more conventional form by choosing coordinates on M adapted to the auxiliary foliation: This yields

$$(1/\sqrt{-g})\partial_{(\beta)}(\sqrt{-g}g^{(\alpha)(\beta)}) = 0.$$

From this expression of the gauge-fixing condition it is clear that the conformal gauge is a special case of the harmonic gauge.

It would be rather awkward to try and extract the functional dependence of g_{ab} on Y directly from (3.33). Because (3.33) can be interpreted as a dynamical equation, it is most expeditious to incorporate the gauge-fixing condition at the level of the action principle by adding a gauge-fixing term to the action. (This, of course, could also have been done in the case of the conformal gauge, but such a scheme is excessively elaborate.) We thus begin with the following gauge-fixed action:

$$S[x,g,\lambda] = -\int_M \sqrt{-g}g^{ab}\Big(\frac{1}{2}\eta_{\mu\nu}\nabla_a x^\mu \nabla_b x^\nu + \nabla_a\lambda_{(\alpha)}\nabla_b Y^{(\alpha)}\Big), \tag{3.34}$$

where $\lambda_{(\alpha)}$ are Lagrange multipliers to be viewed as a pair of scalar functions on M. This action, while defined in a geometric (i.e., coordinate independent) manner, is not invariant with respect to diffeomorphisms. This is because we are not free to redefine the fixed maps $Y^{(\alpha)}$ by using the pullback action of the diffeomorphisms. This is completely analogous to the lack of diffeomorphism invariance of the action (2.1), where it is the metric which is to be immutable.

The equations of motion associated with (3.34) are

$$\frac{\delta S}{\delta x^\mu} = \eta_{\mu\nu}\sqrt{-g}g^{ab}\nabla_a\nabla_b x^\nu = 0, \tag{3.35}$$

$$\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g_{ab}} = (\delta^a_{(c}\delta^b_{d)} - \tfrac{1}{2}g^{ab}g_{cd})(\eta_{\mu\nu}\nabla^c x^\mu \nabla^d x^\nu + 2\nabla^c\lambda_{(\alpha)}\nabla^d Y^{(\alpha)}) = 0, \tag{3.36}$$

and

$$\frac{\delta S}{\delta\lambda_{(\alpha)}} = \sqrt{-g}g^{ab}\nabla_a\nabla_b Y^{(\alpha)} = 0. \tag{3.37}$$

We cannot obtain a dynamical equation for $\lambda_{(\alpha)}$ directly from the action because $Y^{(\alpha)}$ is not to be varied. Nevertheless, an equation of motion for the Lagrange multipliers can be obtained by differentiating (3.36) and using (3.35) and (3.37). We obtain

$$(\nabla^a\nabla_a\lambda_{(\alpha)})\nabla_b Y^{(\alpha)} = 0. \tag{3.38}$$

The covectors $\nabla_b Y^{(\alpha)}$ constitute a one-form basis for M. Hence (3.38) is satisfied if and only if

$$\nabla^a\nabla_a\lambda_{(\alpha)} = 0. \tag{3.39}$$

Equations (3.35), (3.37), and (3.39) are the dynamical equations in the harmonic gauge; these equations must be solved subject to the constraints (3.36). If the constraints (3.36) are satisfied on a Cauchy surface, they will continue to be satisfied throughout the dynamical evolution. The physical subset of string solutions is obtained by choosing vanishing Cauchy data for the multipliers.

As usual, the first step on the road to the parametrized Hamiltonian formalism is to introduce a foliation $X^\alpha(\sigma,\tau)$. The action is pulled back to $R\times S^1$, yielding

$$S[x,g,\lambda]=\int_{R\times S^1}\Big[\frac{1}{2}N^{-1}(\dot x^\mu-N^1x^\mu{}_{,1})(\dot x_\mu-N^1x_{\mu,1})$$
$$+N^{-1}Y^{(\alpha)}{}_{,\beta}(\dot\lambda_{(\alpha)}-N^1\lambda_{(\alpha),1})\qquad(3.40)$$
$$\times(\dot X^\beta-N^1X^\beta{}_{,1})-N(\tfrac{1}{2}x^\mu{}_{,1}x_{\mu,1}$$
$$+Y^{(\alpha)}{}_{,\beta}\lambda_{(\alpha),1}X^\beta{}_{,1})\Big].$$

The definition of the string momenta that follows from (3.40) is the usual one:

$$p_\mu=N^{-1}(\dot x_\mu-N^1x_{\mu,1}).$$

The momenta $\mu^{(\alpha)}$ conjugate to the multipliers are defined by

$$\mu^{(\alpha)}=N^{-1}Y^{(\alpha)}{}_{,\beta}(\dot X^\beta-N^1X^\beta{}_{,1}).\qquad(3.41)$$

Prior to parametrization, $X^\alpha(\sigma,\tau)$ is a fixed map from $R\times S^1$ to M. Equation (3.41) therefore represents a constraint. If we also define the momenta conjugate to the lapse and shift, we find that they vanish—another pair of constraints. The combined set of constraints is easily verified to be *second class*. We need not, however, go through the full Dirac bracket procedure for these constraints: They arise simply because the action (3.40) is already in Hamiltonian form with respect to the multipliers $\lambda_{(\alpha)}$ and their conjugate momenta which, from (3.41), are combinations of the lapse and shift. Hence we need only perform the Legendre transformation in the string sector to obtain the phase space action:

$$S[x,p,\lambda,\mu]=\int_{R\times S^1}[p_\mu\dot x^\mu+\mu^{(\alpha)}\dot\lambda_{(\alpha)}$$
$$-N(\mathscr{H}+Y^{(\alpha)}{}_{,\beta}\lambda_{(\alpha),1}X^\beta{}_{,1})$$
$$-N^1(\mathscr{H}_1+\mu^{(\alpha)}\lambda_{(\alpha),1})].\qquad(3.42)$$

In (3.42) we are to think of the lapse and shift as being fixed functionals of the momenta $\mu^{(\alpha)}$, obtained by inverting (3.41). We will not need the explicit form of these functionals in what follows. Note that the lapse and shift no longer play the role of multipliers enforcing the constraints (2.47) and (2.48): They now enter as dynamically determined degrees of freedom. Of course, this is because we have fixed the gauge. In fact, there are no constraints associated with (3.42); this corresponds to the expected loss of diffeomorphism invariance.

To regain diffeomorphism invariance and to achieve our goal of representing diff(M), we should parametrize the theory. To do this, we start again with (3.40), but now view $X^{(\alpha)}(\sigma,\tau)$ as a dynamical variable. The definition (3.41) no longer represents a constraint and we have the additional set of momenta conjugate to the embeddings defined by

$$P_\alpha=N^{-1}Y^{(\beta)}{}_{,\alpha}(\dot\lambda_{(\beta)}-N^1\lambda_{(\beta),1}).$$

The phase space of the parametrized string is now doubly extended: We shall denote it by Γ'':

$$\Gamma''=T^*[Emb_\eta(S^1,M^d)$$
$$\times C^\infty(S^1,R^2)\times Emb_g(S^1,M)].$$

The three factors making up Γ'' correspond to the string, multiplier, and embedding subspaces, respectively. The fundamental Poisson brackets are as before for the string and embedding variables. For the multipliers we have

$$[\lambda_{(\alpha)}(\sigma),\mu^{(\beta)}(\sigma')]=\delta^{(\beta)}_{(\alpha)}\delta(\sigma,\sigma').$$

The action functional for the doubly extended phase space takes the form

$$S[x,p,\lambda,\mu,X,P,\vec N]$$
$$=\int_{R\times S^1}(p_\mu\dot x^\mu+\mu^{(\alpha)}\dot\lambda_{(\alpha)}$$
$$+P_{(\alpha)}\dot X^\alpha-N\mathbf{H}''-N^1\mathbf{H}_1''),$$

where

$$\mathbf{H}''=\mathscr{H}+(Y^{-1})^\alpha_{(\beta)}P_\alpha\mu^{(\beta)}+Y^{(\alpha)}{}_{,\beta}\lambda_{(\alpha),1}X^\beta{}_{,1}\approx0,\qquad(3.43a)$$

$$\mathbf{H}_1''=\mathscr{H}_1+\mu^{(\alpha)}\lambda_{(\alpha),1}+P_\alpha X^\alpha{}_{,1}\approx0.\qquad(3.43b)$$

In (3.43a) we have defined $(Y^{-1})^\alpha_{(\beta)}$ to be the inverse of $Y^{(\alpha)}{}_{,\beta}$:

$$(Y^{-1})^\alpha_{(\beta)}Y^{(\beta)}{}_{,\gamma}=\delta^\alpha_\gamma.$$

Both $Y^{(\alpha)}{}_{,\beta}$ and $(Y^{-1})^\alpha_{(\beta)}$ are fixed functionals of the embeddings.

As a result of the parametrization process, the role of the lapse and shift has undergone something of a metamorphosis. This is most easily understood when we observe that through (2.9)–(2.11), the lapse and shift are functionals of both the space-time metric and the foliation X. Prior to the parametrization, the foliation is fixed and hence in (3.42) the lapse and shift are the (dynamically determined) descriptors of the harmonic gauge foliation. By including the foliation (or rather the embeddings) into the phase space, the lapse and shift are no longer locked into the space-time metric, but rather become as freely variable as the foliation. The lapse and shift then resume their role as nondynamical Lagrange multipliers enforcing the constraints (3.43). These constraints (3.43) represent the reintroduction of diffeomorphism invariance—the desired result of the parametrization process. As they should be, the constraints are first class: They in fact obey a Dirac algebra identical to (2.54)–(2.56).

From our work above, we know that in order to represent diff (M) we must Abelianize the constraint algebra by effectively unprojecting the constraint functions in (3.43). This amounts to finding (nondegenerate) combinations of $\mathbf{H}''$ and $\mathbf{H}_1''$ which isolate the embedding momenta as, for example, in (3.5). While we could solve the constraints for the embedding momenta by "brute force," it is more instructive to do this by extracting from Γ'' the "hypersurface basis" and its dual. The hypersurface basis is an orthonormal frame at each point of a given embedding. One leg of the dyad is the unit normal to the embedded circle and the other leg is the unit tangent. The quickest way to find this basis is to use the dynamical equation for X^α:

$$\dot X^\alpha=N(Y^{-1})^\alpha_{(\beta)}\mu^{(\beta)}+N^1X^\alpha{}_{,1}.\qquad(3.44)$$

By comparing (3.44) with (2.9) we can read off the unit normal,

$$n^{\alpha}=\gamma^{-1/2}(Y^{-1})^{\alpha}_{(\beta)}\mu^{(\beta)} \tag{3.45}$$

and the unit tangent,

$$t^{\alpha}=\gamma^{-1/2}X^{\alpha}{}_{,1}. \tag{3.46}$$

In order to obtain the Abelian constraints we will need to construct the corresponding dual basis. This can be done by choosing an *arbitrary* volume form ϵ_{ab} on M. The normal covector is defined as

$$m_{\alpha}:=J^{-1}\gamma^{1/2}\epsilon_{\alpha\beta}X^{\beta}{}_{,1}=-n_{\alpha}, \tag{3.47}$$

where

$$J=\epsilon_{\alpha\beta}(Y^{-1})^{\alpha}_{(\gamma)}X^{\beta}{}_{,1}\mu^{(\gamma)}. \tag{3.48}$$

The spatial leg of the covector basis is then

$$t_{\alpha}:=J^{-1}\gamma^{1/2}\epsilon_{\beta\alpha}(Y^{-1})^{\beta}_{(\gamma)}\mu^{(\gamma)}=\gamma^{1/2}X^{1}_{\alpha}. \tag{3.49}$$

It is straightforward to check that the vector and covector bases are dual.

Note that in the definition of the covector basis it was necessary that J be nonvanishing. Inspection of (3.41) reveals that this simply means that the deformation vector is never tangent to a given embedding. Alternatively, the map $X^{\alpha}(\sigma,\tau)$, defined by the dynamical data (in a given coordinate patch) through (3.44), must be a diffeomorphism. The nonvanishing of J is thus guaranteed by our various hypotheses, but it is important to realize that this requirement does place restrictions on admissible initial data. With this proviso understood, we can continue.

Crucial for the work to follow is the fact that the covector basis is defined independent of the choice of volume form on M. This is because the space of two-forms (at a point) is one-dimensional. The dependence of the definitions (3.47)–(3.49) on the choice of volume form drops out because only ratios of terms linear in ϵ_{ab} are used. The vector basis and its dual are thus fixed functionals on the phase space Γ''. (Strictly speaking, the normalized basis depends also on the induced metric γ. As usual, we need not consider this function as an element of the phase space. Weyl invariance will guarantee its absence in any of our final expressions.) This result, ultimately following from (3.44), relies crucially on the gauge-fixing procedure. Without it, we would have no way of extracting the hypersurface basis.

Armed with the hypersurface covector basis, we construct the unprojected constraints:

$$\mathbf{H}''_{\alpha}=-\gamma^{-1/2}n_{\alpha}\mathbf{H}''+X^{1}_{\alpha}\mathbf{H}''_{1}=P_{\alpha}+\mathscr{H}''_{\alpha}\approx 0, \tag{3.50}$$

where

$$\mathscr{H}''_{\alpha}=\mathscr{H}_{\alpha}-\gamma^{-1/2}n_{\alpha}Y^{(\beta)}{}_{,\gamma}\lambda_{(\beta),1}\times X^{\gamma}{}_{,1}+X^{1}_{\alpha}\mu^{(\beta)}\lambda_{(\beta),1}.$$

Using (3.47)–(3.49) it can be verified that the constraints (3.50) are independent of γ. That the constraints obey an Abelian Poisson algebra

$$[\mathbf{H}''_{\alpha}(\sigma),\mathbf{H}''_{\beta}(\sigma')]=0 \tag{3.51}$$

is seen by using an argument which apparently goes back to Dirac.[8] Since the constraints in (3.50) are combinations of the original first class constraints, they will also be first class. When we compute the Poisson bracket between two of the constraint functions (3.50), the result will be independent of the embedding momenta P_{α} because the embedding momenta only appear in the single term shown in (3.50). This is consistent with the first class nature of $\mathbf{H}''_{\alpha}$ if and only if (3.51) holds.

For completeness, let us point out that there is an alternate route to the unprojected constraints. As we did for the parametrized scalar fields, we could rewrite (3.42) in a form that is linear in the embedding velocites $\dot{X}^{\alpha}$. To do this we would need to "unproject" the lapse and shift by extracting from (3.41) [which is equivalent to (3.44)] the hypersurface basis. Upon adjoining the embeddings to the phase space, the definitions of the conjugate momenta P_{α} would appear as the constraints $\mathbf{H}''_{\alpha}\approx 0$.

By using either of these two equivalent methods, our success in extracting the unprojected constraints means success in representing diff(M). The homomorphic mapping from diff(M) into the Poisson algebra of functions on Γ'' is given by

$$V\rightarrow\vec{V}\rightarrow\mathbf{H}''(\vec{V}),$$

where

$$\mathbf{H}''(\vec{V})=\int_{S^1}V^{\alpha}(X(\sigma))\mathbf{H}''_{\alpha}. \tag{3.52}$$

Given two elements of diff(M), i.e., given two complete vector fields $\vec{U}$ and $\vec{V}$ on M, we have

$$[\mathbf{H}''(\vec{U}),\mathbf{H}''(\vec{V})]=\mathbf{H}''(-[\vec{U},\vec{V}]).$$

Using the fact that conformal gauge metrics are also harmonic gauge metrics, the subalgebra conf(M,g) can be represented as in Sec. III A.

As in the case of the conformal gauge, the dynamical evolution generated by $\mathbf{H}''(\vec{V})$ builds a physically acceptable solution to (3.35)–(3.37) provided that we restrict the allowable initial data. In the present situation this is most easily done by imposing, in addition to (3.50), the following constraint:

$$\lambda_{(\alpha)}\approx 0. \tag{3.53}$$

The demand that (3.53) be preserved by the dynamical evolution implies that

$$P_{\alpha}\approx 0, \tag{3.54}$$

as can be seen, e.g., by taking the Poisson bracket of (3.53) with the Hamiltonian

$$\mathbf{H}''(\vec{N}):=\int_{S_1}(N\mathbf{H}''+N^{1}\mathbf{H}''_{1}).$$

It is easily checked in a similar fashion that (3.54) is automatically preserved in time (modulo the constraints). The complete set of constraints (3.50), (3.53), and (3.54) are thus first class, hence they are satisfied for all time if satisfied initially. Moreover, these constraints imply that the string initial data satisfy the original constraints (2.47) and (2.48). The representatives $\mathbf{H}''(\vec{V})$ thus evolve the physical string initial data into physical string solutions.

IV. REPRESENTATION OF diff (M) ON THE ORIGINAL PHASE SPACE

As the Hamiltonian structure of general relativity began to be unraveled in the late 1950's and early 1960's,[9] it became clear that the contraints that arise in generally covariant theories are due to the fact that somehow time and space coordinates are camouflaged within the original dynamical variables. More precisely, the location of a spacelike hypersurface (time) and coordinates on it (space) are implicitly contained in the phase space variables. The momenta conjugate to these variables must then be identified with energy and momentum densities; these identifications are the Hamiltonian and momentum constraints. A clean example of this structure is provided by the parametized harmonic maps considered previously. There, the time and space coordinates, i.e., the embeddings, were easy to identify since they were essentially put in by hand.

The outstanding challenge in systems that are "already parametrized," such as general relativity and the relativistic string, is to extract the embeddings directly from the original phase space. In general relativity this challenge has never been fully met. It may be that for this theory (and perhaps more generally) there is no single identification of embedding variables suitable for all situations. The relative simplicity of the string permits one to go much further, there *is* a rather natural choice of embeddings. Perhaps not unexpectedly, the rewriting of string theory as a parametrized theory is closely tied to—and rests on the validity of—the light-cone gauge. Having explicitly exposed the parametrized aspects of the string, we once again will have succeeded in representing diff(M).

Our method of expressing the string as a parametrized system is based on the observation made by Kuchař[10] that there is a canonical transformation which takes the (projected) constraints of a parametrized scalar field on a two-dimensional Minkowski space-time into a form identical to those of the string. Here we will essentially run that argument backward. Begin by defining two covariantly constant, linearly independent null vectors k^μ and l^μ in the target space M^d:

$$l_\mu l^\mu = 0 = k^\mu k_\mu;$$

they are to be normalized relative to each other such that

$$\eta_{\mu\nu} k^\mu l^\nu = -1.$$

Light-cone components for any objects with target space (Lorentz) indices are defined using k^μ, l^μ; in particular,

$$x^+ := -\eta_{\mu\nu} x^\mu l^\nu, \quad x^- := -\eta_{\mu\nu} x^\mu k^\nu,$$

and

$$p_+ := k^\mu p_\mu, \quad p_- := l^\mu p_\mu.$$

In terms of light-cone components the string constraints take the form

$$\mathscr{H} = -p_+ p_- - x^+{}_{,1} x^-{}_{,1} + \tfrac{1}{2}(p_i p^i + x^i{}_{,1} x_{i,1}) \approx 0, \quad (4.1)$$

$$\mathscr{H}_1 = p_+ x^+{}_{,1} + p_- x^-{}_{,1} + p_i x^i{}_{,1} \approx 0. \quad (4.2)$$

Latin indices from the middle of the alphabet are taken to label spacelike directions in M^d which are orthogonal to the null vectors, i.e., $i = 1, 2, \ldots, d-2$.

Now, consider the transformation defined by

$$T_{,1} = -(p_-)_0^{-1} x^+{}_{,1}, \quad (4.3)$$

$$S_{,1} = (p_-)_0^{-1} p_-, \quad (4.4)$$

$$P_T = -(p_-)_0 p_+, \quad (4.5)$$

$$P_S = (p_-)_0 x^-{}_{,1}, \quad (4.6)$$

where

$$(p_\mu)_0 := \frac{1}{2\pi} \int_{S_1} p_\mu .$$

Our reuse of the symbols T and S will be justifed shortly. In terms of the relabeled variables, (4.1) and (4.2) become

$$\mathscr{H} = P_T S_{,1} + P_S T_{,1} + \tfrac{1}{2}(p_i p^i + x^i{}_{,1} x_{i,1}) \approx 0, \quad (4.7)$$

$$\mathscr{H}_1 = P_T T_{,1} + P_S S_{,1} + p_i x^i{}_{,1} \approx 0. \quad (4.8)$$

Comparing with (2.36) and (2.37), we see that Eqs. (4.7) and (4.8) are precisely the form of the constraints that arise in the parametrized formalism for a set of $d-2$ scalar fields x^i in two dimensions. This interpretation of (4.7) and (4.8) has the variables $T(\sigma)$ and $S(\sigma)$ playing the role of spacelike embeddings with respect to conformal coordinates on M. In contrast to our previous work with the conformal gauge (III), no gauge fixing is needed here to provide a privileged set of coordinates. The string itself determines a conformal coordinate system (up to a choice of origin; see the discussion below).

At this point, our identification of the string as a set of $d-2$ parametrized scalar fields has been purely formal. A closer examination of the transformation (4.3)–(4.6) is warranted. To begin, the transformation, strictly speaking, is not canonical on Γ. On the circle the operator ∂_σ has a nonempty kernel given by the constant functions. In particular, the mapping $(x^-, p_-) \leftrightarrow (S, P_S)$ is not bijective. (This issue did not arise in Ref. 10 because there M was taken to be an *infinite* two-dimensional Minkowski space.) Thus for example, P_S as defined in (4.6) has a vanishing constant mode:

$$\int_{S^1} P_S = (p_-)_0 \int_{S^1} x^-{}_{,1} = 0.$$

As a result, the variables $S(\sigma)$ and $P_S(\sigma)$ defined through (4.4) and (4.6) cannot be canonically conjugate; if they were, we would obtain a contradiction:

$$1 = \int_0^{2\pi} d\sigma' \, \delta(\sigma, \sigma') = \left[S(\sigma), \int_{S^1} P_S \right] = 0.$$

There is, however, a natural subspace of Γ for which the transformation (4.3)–(4.6) *is* canonical. This subspace is obtained by effectively factoring out the "center of mass" portion of Γ. Denote by $\tilde{\Gamma}$ the cotangent bundle over the space $\widetilde{Emb}_\eta(S^1, M^d)$ of *based loops* in M^d, i.e., embeddings of S^1 into M^d which begin and end at a fixed point. As a differentiable manifold, $\widetilde{Emb}_\eta(S^1, M^d)$ has many nice properties, e.g., it is a Kähler manifold.[11] We can identify $\widetilde{Emb}_\eta(S^1, M^d)$ with the space of exact one-forms on S^1 taking their values in Minkowski vector space. The string phase space can then be decomposed as

$$\Gamma \simeq T^*M^d \times \tilde{\Gamma}$$
$$\simeq M^d \times M^d \times T^*\widetilde{Emb}_\eta(S^1, M^d).$$

For a given value of $(p_\mu)_0$ local (noncanonical) coordinates for $\tilde{\Gamma}$ are provided by $(x^\mu{}_{,1}, p_\nu)$ or, alternatively, $(\alpha^\mu, \tilde{\alpha}^\nu)$ [defined in (3.21) and (3.22)]. It is a key feature of the string constraints [e.g., (4.1) and (4.2) or (4.7) and (4.8)] that they are essentially only functions on $\tilde{\Gamma}$. In any computations we will perform which involve only the constraints, the center of mass momenta $(p_\mu)_0$ will simply "go along for the ride."

The transformation (4.3)–(4.6) can be interpreted as a one-parameter family of canonical transformation on $\tilde{\Gamma}$, where the parameter is $(p_-)_0$. This parameter represents the discontinuity which $S(\sigma)$ must possess if it is to be the spatial coordinate location of an embedded circle. For this reason $-(p_-)_0$ is often called the "string length" in the light-cone gauge; see the discussion below. We have rescaled the would-be embedding variables so that this discontinuity has the value 2π. Thus if $\sigma \in [0, 2\pi]$, with $\sigma = 0$ and $\sigma = 2\pi$ identified, we have

$$S(2\pi) - S(0) = \int_0^{2\pi} d\sigma\, S_{,1} = 2\pi.$$

We must then make the identification

$$S \sim S + 2\pi \tag{4.9}$$

to produce embeddings with the topology of a circle.

While we have uncovered the natural space on which the transformation is bijective, there is still the problem that the embeddings and their conjugate momenta as identified in (4.3)–(4.6) do not constitute a canonical chart on $T^*Emb(S^1, M)$. At the root of this difficulty is the degeneracy of the transformations (4.4) and (4.6), which reflects the fact that the light-cone string variables do not contain all of the "pure gauge" degrees of freedom. To see this, notice that (4.4) leaves a mode of $S(\sigma)$ unspecified. This mode can be taken to represent a choice of origin for the S coordinate relative to the σ origin. It is not fixed by the light-cone variables, nor is it fixed by the constraints (4.7) and (4.8). The freedom to choose the origin for S is a miniature gauge freedom introduced by the degenerate nature (on Γ) of the transformations. Indeed, the corresponding first class constraint on $T^*Emb(S^1, M)$ is simply

$$\int_{S^1} P_S \approx 0. \tag{4.10}$$

Alternatively, by virtue of the constrainsts (4.7) and (4.8), (4.10) can be viewed as a restriction on allowable initial data for the *transverse* string variables:

$$\int_{S^1} [(S_{,1})^2 - (T_{,1})^2]^{-1} \Big[\frac{1}{2} T_{,1} (p_i p^i + x^i{}_{,1} x_{i,1}) - S_{,1} (p_i x^i{}_{,1}) \Big] \approx 0. \tag{4.11}$$

As we shall see shortly, this constraint is the precursor of the familiar shift of origin constraint which remains after going to the light-cone gauge.[4]

The single degree of freedom that is fixed by the constraint (4.11) can be thought of as the missing constant mode of the embedding momentum P_S. Its conjugate coordinate is to be a mode of $S(\sigma)$ such that the embeddings and their momenta provide a true canonical chart. Evidently, extracting this single remaining degree of freedom from the transverse subspace of the phase space is an unpleasant task and we have not succeeded in doing this in any nice way. We can sidestep this difficulty by using the same strategy that served us so well in Sec. III: If there is difficulty in isolating the embeddings from the original phase space, simply introduce them explicity. We do this as follows. Fix the origin of S relative to σ by integrating (4.4) as

$$S(\sigma) = (p_-)_0^{-1} \Big(-q + \int_0^\sigma dx\, p_-(x) \Big).$$

We now treat q as a dynamical variable. Its conjugate momentum, denoted p, must be constrained to vanish:

$$p \approx 0,$$

so that q is pure gauge, i.e., not dynamically determined; hence the enlarged phase space has the same physical content as before. The momentum is incorporated into the definition of P_S via

$$P_S = (p_-)_0 (x^-{}_{,1} - p\delta(\sigma))$$

P_S is thus weakly equal to its original definition. (Note that P_S is now a distribution. This feature of the formalism can be avoided by introducing an auxiliary prescribed measure on S^1. For simplicity we will retain the delta function "measure" in what follows.) Using the Poisson bracket

$$[q, p] = 1,$$

it is easily verified that $S(\sigma)$ and $P_S(\sigma)$ are canonically conjugate variables:

$$[S(\sigma), P_S(\sigma')] = \delta(\sigma, \sigma'). \tag{4.12}$$

The remaining degree of freedom in the light-cone subspace of Γ corresponds to the coordinate conjugate to the string length $(p_-)_0$. Intuitively, this single mode should represent the constant map from S^1 into the "$-$" direction of M^d. To extract this mode from $x^-(\sigma)$ we need to choose a measure on the circle. A natural choice is given by $S_{,1}$, as defined in (4.4). This choice of measure "weighs" functions using the metric induced on the circles $T = $ const by a flat metric on M. Using this measure to define the homogeneous mode of x^- and then extending the definition to provide a *canonical* coordinate on the extension of Γ, we obtain

$$(x^-)_0 := (p_-)_0^{-1} \int_{S_1} \Big(p_- x^- - p_+ x^+ - \frac{qp}{2\pi} \Big).$$

Straightforward computation confirms that $(x^-)_0$ has vanishing Poisson brackets with all variables except $(p_-)_0$: This bracket is

$$[(x^-)_0, (p_-)_0] = 1.$$

To summarize, if we work on a slightly extended phase space Γ^* obtained by including the canonical pair q,p, then from the light-cone string variables we can isolate a natural set of embedding variables (relative to a conformal coordinate chart on M) by the *canonical* transformation on Γ^*:

$$T = -(p_-)_0^{-1} x^+, \tag{4.13}$$

$$S = -(p_-)_0^{-1}\left(q - \int_0^\sigma dx\, p_-(x)\right), \qquad (4.14)$$

$$P_T = -(p_-)_0 p_+, \qquad (4.15)$$

$$P_S = (p_-)_0 (x^-_{,1} - p\delta(\sigma)), \qquad (4.16)$$

$$(x^-)_0 = (p_-)_0^{-1}\int_{S^1}\left(p_- x^- - p_+ x^+ - \frac{qp}{2\pi}\right), \qquad (4.17)$$

$$(p_-)_0 = \frac{1}{2\pi}\int_{S^1} p_-. \qquad (4.18)$$

In the new canonical variables, the constraints take the form (4.7) and (4.8), along with

$$\int_{S^1} P_S \approx 0. \qquad (4.19)$$

In (4.7) and (4.8) P_S is to be thought of as having all of its modes intact. The transformation (4.13)–(4.18), when complemented with the identity transformation on the transverse phase space, is bijective.

As always, the easiest way to exhibit the desired homomorphisms is through the unprojected (Abelian) constraints. These have already been written (for d scalar fields) in terms of the null combinations of T and S in Sec. III A. Translating (3.20) into the original string variables, we obtain

$$\begin{aligned}\mathbf{H}_+ :&= \tfrac{1}{2}(P_T + P_S) + (T_{,1} + S_{,1})^{-1}\tilde{\alpha}_i\tilde{\alpha}^i \\ &= (p_-)_0(\tilde{\alpha}^- - p\,\delta(\sigma) - (2\tilde{\alpha}^+)^{-1}\tilde{\alpha}_i\tilde{\alpha}^i) \approx 0 \qquad (4.20)\end{aligned}$$

and

$$\begin{aligned}\mathbf{H}_- :&= \tfrac{1}{2}(P_T - P_S) - (T_{,1} - S_{,1})^{-1}\alpha_i\alpha^i \\ &= (p_-)_0(\alpha^- + p\,\delta(\sigma) - (2\alpha^+)^{-1}\alpha_i\alpha^i) \approx 0; \qquad (4.21)\end{aligned}$$

these functions obey an Abelian Poisson algebra.

To construct the representatives of diff(M) we proceed as follows. Fix a set of coordinates $\mathrm{X}^\pm$ on M by letting $\mathrm{X}^\pm \in (-\infty, \infty)$ and making the identification

$$\mathrm{X}^+ - \mathrm{X}^- \sim \mathrm{X}^+ - \mathrm{X}^- + 4\pi.$$

If we define

$$X^\pm(\sigma) = (T(\sigma) \pm S(\sigma)), \qquad (4.22)$$

$$P_\pm = \tfrac{1}{2}(P_T \pm P_S), \qquad (4.23)$$

then we can identify $X^\pm(\sigma)$ as an embedding of S^1 into M expressed parametrically in the coordinates $\mathrm{X}^\pm$. The corresponding embedding deformation generators are given in (4.20) and (4.21). Given a vector field $\vec{V}$ on M representing $V \in \mathrm{diff}(M)$, we take its components in the coordinate basis provided by $\mathrm{X}^\pm$ and pull the resulting functions $V^\pm$ back to S^1 along the embedding (4.22). We use these two functions on S^1 to smear the generators $\mathbf{H}_\pm$:

$$\mathbf{H}(\vec{V}) = \int_{S^1}(V^+\mathbf{H}_+ + V^-\mathbf{H}_-), \qquad (4.24)$$

where the functions $V^\pm(\sigma)$ are to be viewed as fixed functionals of the canonical variables (x^+, p_-, q) through the combinations (4.22), (4.13), and (4.14). Either from direct computation or as a consequence of our previous work (III), the map

$$V \to \vec{V} \to \mathbf{H}(\vec{V})$$

is once again a homomorphism:

$$[\mathbf{H}(\vec{V}), \mathbf{H}(\vec{W})] = \mathbf{H}(-[\vec{V}, \vec{W}]).$$

Notice that the enlargement of Γ to include q and p is crucial for this result.

To obtain a representation of conf(M,g), we need to know how to define conformal Killing vectors. This is slightly subtle since the metric on M is only determined (up to conformal rescalings) by the dynamical evolution of the string initial data. However, having evolved the data, the functions $X^\pm(\sigma)$ have the interpretation of embeddings relative to a null coordinate system on M. Hence the restriction of the components of a conformal Killing vector $\vec{v}$ to such an embedding takes the form

$$v^\pm(\sigma) = v^\pm(X^\pm(\sigma)). \qquad (4.25)$$

The homomorphism from conf(M,g) into $C^\infty(\Gamma^*, R)$ is given by

$$v \to \vec{v} \to \mathbf{H}(\vec{v}),$$

where

$$\mathbf{H}(\vec{v}) = \int_{S^1}(v^+\mathbf{H}_+ + v^-\mathbf{H}_-). \qquad (4.26)$$

As before, rather than view conf(M,g) as a subgroup of the dynamical group Diff(M), we can emphasize its role as a symmetry group for the string via the homomorphism

$$v \to \vec{v} \to \mathbf{P}(\vec{v}),$$

$$\begin{aligned}\mathbf{P}(\vec{v}) = \int_{S^1}(p_-)_0[&v^+(\tilde{\alpha}^- - p\,\delta(\sigma)) \\ &+ v^-(\alpha^- + p\,\delta(\sigma))], \qquad (4.27)\end{aligned}$$

and the antihomomorphism

$$v \to \vec{v} \to \mathbf{h}(\vec{v}),$$

$$\begin{aligned}\mathbf{h}(\vec{v}) = \int_{S^1} -(p_-)_0[&v^-(2\alpha^+)^{-1}\alpha_i\alpha^i \\ &+ v^+(2\tilde{\alpha}^+)^{-1}\tilde{\alpha}_i\tilde{\alpha}^i]. \qquad (4.28)\end{aligned}$$

Using the fact that $v^+(\sigma)$ and $v^-(\sigma)$ are fixed functionals of x^+, p_-, and q through the combinations $X^+(\sigma)$ and $X^-(\sigma)$, respectively, we have

$$[\mathbf{P}(\vec{v}), \mathbf{P}(\vec{w})] = \mathbf{P}(-[\vec{v}, \vec{w}])$$

and

$$[\mathbf{h}(\vec{v}), \mathbf{h}(\vec{w})] = \mathbf{h}([\vec{v}, \vec{w}]).$$

As in Sec. III A, the functions $\mathbf{h}(\vec{v})$ are symmetry generators, i.e., constants of motion, while the functions $\mathbf{P}(\vec{v})$ are generators of a conditional symmetry, i.e., they are constants of motion modulo (4.20) and (4.21).

The "diffeomorphism Hamiltonian" $\mathbf{H}(\vec{V})$ generates the dynamical evolution associated with the embedding deformation

$$X^\alpha(\sigma) \to X^\alpha(\sigma) + V^\alpha(X(\sigma)),$$

exactly as in Sec. III A. It is therefore a simple matter to make the identification

$$X^\pm(\sigma) = -(p_-)_0^{-1}\left(x^+ \pm q \mp \int_0^\sigma dy\, p_-(y)\right)$$

and write the solutions to the many-fingered time Hamiltonian equations for the transverse variables. Thus

$$x^i[x^+,p_-,q](\sigma) = f^i[X^+(\sigma)] + g^i[X^-(\sigma)] \tag{4.29}$$

and

$$p_i[x^+,p_-,q](\sigma) = \delta_{ij}(X^+{}_{,1} f^j{}_{,+} - X^-{}_{,1} g^j{}_{,-}), \tag{4.30}$$

where $f^i(X^+)$ and $g^i(X^-)$ are any two functions which satisfy

$$f^i(X^+ + 2\pi) + g^i(X^- - 2\pi) = f^i(X^+) + g^i(X^-),$$

are solutions to

$$\frac{\delta x^i}{\delta X^\pm(\sigma)} = [x^i,\mathbf{H}_\pm(\sigma)], \quad \frac{\delta p_i}{\delta X^\pm(\sigma)} = [p_i,\mathbf{H}_\pm(\sigma)].$$

The solutions for the remaining degrees of freedom $(x^-)_0$ and $(p_-)_0$ are easily obtained by virtue of the fact that they are constants of motion. This can be seen by observing that the coordinates on Γ^* defined in (4.13)–(4.18) are canonical and in this chart $\mathbf{H}_\pm$ are functionals of the embedding and transverse variables only. Thus the functional differential equations for the two remaining phase space coordinates take the simple form

$$\frac{\delta(x^-)_0}{\delta X^\pm(\sigma)} = [(x^-)_0,\mathbf{H}_\pm(\sigma)] = 0,$$

$$\frac{\delta(p_-)_0}{\delta X^\pm(\sigma)} = [(p_-)_0,\mathbf{H}_\pm(\sigma)] = 0.$$

Solutions to these equations are, of course,

$$(x^-)_0 = a, \tag{4.31}$$

$$(p_-)_0 = b, \tag{4.32}$$

where a and b are constants.

To complete the specification of the solutions, we again should match the solutions to an initial data set. This is done as follows. Choose an initial data surface by specifying the initial values $\overline{X}^\alpha(\sigma)$ for the embeddings. On that surface specify the initial data $[(\overline{x^-})_0,(\overline{p_-})_0,\bar{x}^i,\bar{p}_i]$. Initial data for the embedding momenta are fixed by the constraints

$$\overline{P}_+ = -(\overline{X}^+{}_{,1})^{-1}\bar{\bar{\alpha}}_i\bar{\bar{\alpha}}^i, \tag{4.33a}$$

$$\overline{P}_- = (\overline{X}^-{}_{,1})^{-1}\overline{\alpha}_i\overline{\alpha}^i. \tag{4.33b}$$

Since the constraints are first class, they continue to hold throughout the dynamical evolution; hence (4.33) fixes the evolution of the embedding momenta in terms of the transverse dynamics. The tranverse initial data are also constrained [via (4.10) or (4.11)] [From (4.28), this constraint function represents the action on the transverse string variables of a conformal isometry which is generated by the conformal Killing vector field tangent to the circles $T=\text{const}$]:

$$\int_{S^1}((\overline{X}^+{}_{,1})^{-1}\bar{\bar{\alpha}}_i\bar{\bar{\alpha}}^i + (\overline{X}^-{}_{,1})^{-1}\overline{\alpha}_i\overline{\alpha}^i) = 0. \tag{4.34}$$

Again, this constraint is first class and need only be satisfied initially. In terms of the original string coordinates, (4.33) and (4.34) are equivalent to

$$\bar{\bar{\alpha}}^- = (2\bar{\bar{\alpha}}^+)^{-1}\bar{\bar{\alpha}}_i\bar{\bar{\alpha}}^i, \quad \overline{\alpha}^- = (2\overline{\alpha}^+)^{-1}\overline{\alpha}_i\overline{\alpha}^i, \quad \bar{p} = 0.$$

With the constraints taken care of, we can now match the solutions to the initial data. For example, let the initial embedding be an element of the preferred foliation associated with the coordinates $X^\alpha = (T,S)$ i.e., let

$$\overline{T}(\sigma) = 0, \quad \overline{S}(\sigma) = \sigma.$$

Then the solutions (4.29)–(4.32) become uniquely determined functionals of the embeddings via

$$x^i[X](\sigma) = \frac{1}{2}\Big[\bar{x}^i(X^+(\sigma)) + \bar{x}^i(-X^-(\sigma)) + \int_{-X^-(\sigma)}^{X^+(\sigma)} dy\,\bar{p}^i(y)\Big], \tag{4.35}$$

$$\begin{aligned} p_i[X](\sigma) &= \tfrac{1}{2}[X^+{}_{,1}(\sigma)(\bar{x}_{i,+}(X^+(\sigma)) + \bar{p}_i(X^+(\sigma))) \\ &\quad + X^-{}_{,1}(\sigma)(\bar{x}_{i,-}(-X^-(\sigma)) \\ &\quad - \bar{p}_i(-X^-(\sigma)))], \end{aligned} \tag{4.36}$$

$$(x^-)_0 = (\overline{x^-})_0, \tag{4.37}$$

$$(p_-)_0 = (\overline{p_-})_0. \tag{4.38}$$

We can, if we like, specify the solutions by incorporating the initial data surface into a (freely specifiable) one-parameter family of embeddings:

$$X^\alpha = X^\alpha(\sigma,\tau), \quad \text{where} \quad X^\alpha(\sigma,0) = \overline{X}^\alpha(\sigma).$$

A preferred choice is, of course, given by

$$T(\sigma,\tau) = \tau, \tag{4.39}$$

$$S(\sigma,\tau) = \sigma. \tag{4.40}$$

Translating expressions (4.39) and (4.40) back into the original string variables, we find that this choice of foliation corresponds to working in the light-cone gauge. Thus (4.39) becomes

$$x^+ = -(p_-)_0\tau,$$

while the derivative of (4.40) yields

$$p_- = (p_-)_0.$$

The solutions (4.35) and (4.36) then take the familiar form

$$x^i(\sigma,\tau) = \frac{1}{2}\Big(\bar{x}^i(\sigma+\tau) + \bar{x}^i(\sigma-\tau) + \int_{\sigma-\tau}^{\sigma+\tau} dy\,\bar{p}^i(y)\Big), \quad p_i = \dot{x}_i,$$

while the constraint on the transverse initial data becomes

$$\int_{S^1} \bar{p}_i\bar{x}^i{}_{,1} = 0.$$

The reduction to the light-cone gauge can be viewed as the final step in a sequence of reductions beginning with the phase space Γ' of Sec. III A. In this reduction we first eliminate the embeddings by identifying them with the light-cone variables, as we have just described (this can be thought of as gauge fixing). The embedding momenta are then eliminated by solving (3.5). The remaining constraints can be taken as $P_\alpha \approx 0$ which, having solved (3.5), become constraints on the string data of exactly the form (4.20) and (4.21). The final step in the reduction amounts to imposing the light-cone gauge conditions (4.39) and (4.40), whereby we fix the allowable embeddings to be leaves of a foliation adapted to conformal coordinates. These coordinates are (effectively) defined by the string itself.

Of course, the initial identification of the embedding variables from Γ' with the string variables is permissible only if the combinations (4.13) and (4.14) [with the identification (4.9)] truly define spacelike embeddings. This requirement effectively amounts to the validity of the inequality [see (3.17)]

$$(S_{,1})^2 - (T_{,1})^2 > 0. \tag{4.41}$$

As we shall see, the inequality (4.41) is not satisfied precisely when the light-cone gauge fails to be admissible. By admissible we mean that every dynamical trajectory in Γ^* can be deformed by the induced action of a diffeomorphism on M into a trajectory which satisfies (4.39) and (4.40). Thus consider the equations of motion for the putative embedding variables:

$$\dot{X}^{\pm} = (N^1 \pm N) X^{\pm}{}_{,1}. \tag{4.42}$$

Representing the diffeomorphism locally as a coordinate transformation, the Jacobian for the transformation from the (almost) global coordinates (σ,τ) to the (conformal) light-cone gauge coordinates is given by

$$\mathscr{J} = \dot{X}^{+} X^{-}{}_{,1} - \dot{X}^{-} X^{+}{}_{,1}.$$

Using (4.42) we have

$$\mathscr{J} = 2N X^{+}{}_{,1} X^{-}{}_{,1}.$$

By assumption, $N \neq 0$; thus the coordinate transformation is a good one only when $X^{+}{}_{,1} X^{-}{}_{,1} \neq 0$. Moreover, given a nonvanishing Jacobian, we can pass to the light-cone gauge only if $X^{+}{}_{,1} X^{-}{}_{,1} < 0$, for if $X^{+}{}_{,1} X^{-}{}_{,1} > 0$ we have

$$(T_{,1})^2 - (S_{,1})^2 > 0.$$

Setting $T = \tau$ then leads to a contradiction. Hence the light-cone gauge is admissible only if

$$X^{+}{}_{,1} X^{-}{}_{,1} < 0, \tag{4.43}$$

which is just (4.41). It has been noted in the literature[12] that for the *open* string, the transformation to light-cone gauge coordinates is singular. We will now indicate how similar difficulties arise (in a somewhat more severe manner) for the closed string which is of interest here. To do this, we should be quite specific on how the closed string phase space is to be defined. As mentioned in Sec. II B, a (slightly generous) definition of Γ is the cotangent bundle over the space of smooth embeddings of a circle into d-dimensional Minkowski space such that the metric induced on the circle by pulling back the Minkowski metric on M^d is positive definite. (Here we ignore the q, p extension of Γ as it is irrelevant for the discussion to follow.) This restriction on the string variables translates into

$$x^{\mu}{}_{,1} x_{\mu,1} > 0. \tag{4.44}$$

Further, at least classically, we can restrict our attention to the constraint surface $\overline{\Gamma} \subset \Gamma$ obtained by imposing (2.47) and (2.48) or equivalently (4.1) and (4.2). These requirements are most transparent when they are written in terms of the $\tilde{\alpha}^{\mu}$, α^{μ} variables introduced previously. Thus $\overline{\Gamma}$ is defined as that subspace of Γ that satisfies

$$\tilde{\alpha}^{\mu} \tilde{\alpha}_{\mu} = 0, \tag{4.45a}$$

$$\alpha^{\mu} \alpha_{\mu} = 0. \tag{4.45b}$$

The requirements (4.45) are just a restatement of the constraints. Geometrically, (4.45) tell us that at each point of the string (labeled by σ) α^{μ}, $\tilde{\alpha}^{\mu}$ are a pair of null vectors in M^d. Modulo (4.45), (4.44) can be written as

$$-\alpha^{\mu} \tilde{\alpha}_{\mu} > 0. \tag{4.46}$$

Equation (4.46) implies that the two null vectors at each σ are linearly independent and are given the same orientation in M^d. Without loss of generality, we can assume that both $\alpha^{\mu}(\sigma)$ and $\tilde{\alpha}^{\mu}(\sigma)$ are future oriented. Similarly, we will assume that the vectors k^{μ}, l^{μ} are future oriented. Now, the admissibility of the light-cone gauge (4.41) translates into

$$\alpha^{+} \tilde{\alpha}^{+} > 0. \tag{4.47}$$

Off of $\overline{\Gamma}$, that is, if we only impose (4.44), it is easy to see that (4.47) need not be satisfied. If we restrict attention to $\overline{\Gamma}$, i.e., impose (4.45), it can be shown that the requirements (4.45) and (4.46) are inadequate to guarantee the strict inequality (4.47).

Let us first show that we cannot reverse the inequality in (4.47). To see this, note that since α^{μ} and $\tilde{\alpha}^{\mu}$ are future pointing null vectors, they have a negative definite scalar product with the timelike vector $t^{\mu} := k^{\mu} + l^{\mu}$:

$$\tilde{\alpha}_{\mu} t^{\mu} < 0, \quad \alpha_{\mu} t^{\mu} < 0.$$

Thus we have

$$\tilde{\alpha}^{+} + \tilde{\alpha}^{-} > 0, \quad \alpha^{+} + \alpha^{-} > 0.$$

Furthermore, from (4.45) the products $\alpha^{+}\alpha^{-}$ and $\tilde{\alpha}^{+}\tilde{\alpha}^{-}$ are positive semidefinite. Hence we conclude that the set of functions $(\alpha^{\pm}, \tilde{\alpha}^{\pm})$ are all greater than or equal to zero; in particular

$$\tilde{\alpha}^{+} \alpha^{+} \geqslant 0.$$

Unfortunately, we cannot make the above weak inequality a strong one. A simple example will be sufficient to demonstrate this.

Consider the following point(s) in $\overline{\Gamma}$:

$$\tilde{\alpha}^{\mu} = \tfrac{1}{2} k^{\mu} + l^{\mu} + (\cos \sigma) s_1^{\mu} + (\sin \sigma) s_2^{\mu},$$

$$\alpha^{\mu} = (\sin^2 \sigma) k^{\mu} + l^{\mu} + (\sqrt{2} \sin \sigma) s_3^{\mu},$$

where s_i^{μ} are a set of $d-2$ spacelike orthonormal vectors, labeled by i, orthogonal to k^{μ} and l^{μ}. (Note that this example requires $d \geqslant 5$.) These data uniquely determine $p_{\mu}(\sigma)$ and fix $x^{\mu}(\sigma)$ up to an irrelevant additive constant. It is easily verified that this initial data set satisfies (4.45) and (4.46) at each σ. As for (4.47), we have

$$\alpha^{+} \tilde{\alpha}^{+} = \tfrac{1}{2} \sin^2 \sigma \geqslant 0.$$

Thus the transformation to the light-cone gauge from this point in phase space is singular at $\sigma = 0$. Notice that the intrinsic geometry generated by the above choice of initial data is perfectly regular. If we evolve the data into a string solution using conformal coordinates for σ and τ, we find

$$\tilde{\alpha}^{\mu}(\sigma,\tau) = \tfrac{1}{2} k^{\mu} + l^{\mu} + \cos(\sigma + \tau) s_1^{\mu} + \sin(\sigma + \tau) s_2^{\mu},$$

$$\alpha^{\mu}(\sigma,\tau) = \sin^2(\sigma - \tau) k^{\mu} + l^{\mu} + \sqrt{2} \sin(\sigma - \tau) s_3^{\mu}.$$

The induced metric on M is then given by

$$ds^2 = (1 + 2\sin^2(\sigma - \tau))(-d\tau^2 + d\sigma^2).$$

This metric is completely nonsingular. Moreover, the corresponding curvature vanishes on all of M. Thus in terms of the intrinsic geometry, the initial data consist merely of a nonstandard slice of a flat cylinder.

The above example illustrates the fact that strictly speaking, the functions $\mathbf{H}(\vec{V})$ are not globally defined on Γ^* or even $\overline{\Gamma}$. It may be that the necessary weakening of the inequality in (4.47) is rather harmless: If we work with the constraint surface as defined by (4.7) and (4.8) one can certainly imagine a limiting procedure in which the diffeomorphism Hamiltonians remain defined at the points of $\overline{\Gamma}$ where the light-cone gauge fails. Alternatively, we can simply redefine the phase space such that for a given choice of k^μ and l^μ the inequality (4.47) holds. This is being done implicitly whenever one uses the light-cone gauge in string theory. The redefined phase space, being a dense subset of $\overline{\Gamma}$, may differ insignificantly from $\overline{\Gamma}$ itself, e.g., the light-cone gauge only fails on sets of measure zero in $\overline{\Gamma}$. If this were true, it would seem to imply that the corresponding quantum theory would be unaltered by using the more restricted phase space. However, the inequivalence of light-cone gauge and "covariant" quantization away from the critical dimension suggests otherwise. Clearly, a complete understanding of the difficulties associated with the light-cone gauge will require a separate investigation; we hope to pursue this in a future publication.

To summarize, insofar as the light-cone gauge is admissible, we have succeeded in finding a relatively simple representation of diff(M) on (a slight extension of) the usual string phase space. Other, more complicated representations can, of course, be contemplated. For example, one could use the component of x^μ along a *timelike* vector in M^d to define the "many-fingered time." The corresponding gauge conditions would certainly be admissible. And it would seem that one could represent diff(M) globally on Γ (or at least $\overline{\Gamma}$). We have refrained from following this avenue for pragmatic reasons: The corresponding preferred coordinates are not conformal and one must contend in the quantum theory with unwieldy (square root) operators.

V. BRST EXTENSIONS

We hardly need emphasize the central role that the BRST formalism has played in string quantum mechanics and string field theory. More generally, it is currently being revealed[13] that this way of dealing with constrained systems provides an elegant unification between the classical Hamiltonian structure of such systems and their canonical or path integral methods of quantization. Consequently, it is useful to extend our method of representing diff(M) to include the phase spaces that are enlarged by the introduction of the ghost variables that feature in the BRST formalism. Our goal in this section will be to obtain BRST extensions of the functions (3.5) and (3.50) and (4.20) and (4.21). Since these "diffeomorphism Hamiltonians" are constrained to vanish, their extended versions should arise as cohomologically trivial functions on the BRST phase space.

Let us begin with the phase space Γ' obtained through the use of the conformal gauge. The specific form of the BRST charge Ω' that we construct will depend on how we choose to represent the constraint surface $\overline{\Gamma}'$. Of course, there will be an infinity of choices, but all the resulting BRST charges can be shown to be related by canonical transformations on the BRST-extended phase space (this is one of the most beautiful features of the classical BRST formalism). For our purposes, it will be most convenient to represent $\overline{\Gamma}'$ by

$$\mathbf{H}'_\alpha \approx 0 \tag{5.1}$$

and

$$\mathscr{H} \approx 0 \approx \mathscr{H}_1. \tag{5.2}$$

Local coordinates on the extended phase space, which will be denoted $\hat{\Gamma}'$, consist of the string and embedding variables introduced previously along with the Grassmann-valued ghost coordinates $(\eta^\alpha,\eta,\eta^1)$ and their respective conjugate momenta $(\mathscr{P}_\alpha,\mathscr{P},\mathscr{P}_1)$. The triplet $(\eta^\alpha,\eta,\eta^1)$ consists, respectively, of a pair of scalar functions, a scalar density of weight minus 1, and a vector (all with respect to S^1). The corresponding conjugate momenta are, respectively, a pair of scalar densities of weight 1, a density of weight 2, and a covector density of weight 1. The only nonvanishing (symmetric) Poisson brackets are

$$[\eta^\alpha(\sigma),\mathscr{P}_\beta(\sigma')] = -\delta^\alpha_\beta\delta(\sigma,\sigma'),$$
$$[\eta(\sigma),\mathscr{P}(\sigma')] = -\delta(\sigma,\sigma'),$$
$$[\eta^1(\sigma),\mathscr{P}_1(\sigma')] = -\delta(\sigma,\sigma').$$

To obtain Ω', we will need the Poisson brackets between the constraint functions chosen in (5.1) and (5.2). We have already seen that the functions in (5.1) obey an Abelian algebra, while the functions of (5.2) obey the Dirac algebra (2.54)–(2.56), which is isomorphic to diff(S^1) $\oplus$ diff(S^1). The rest of the Poisson brackets are given by

$$[\mathbf{H}'(\vec{N}),H(M)]$$
$$= \int_{S^1} -\gamma^{-1/2}n_\alpha N^\alpha[(\partial_1 M)\mathscr{H}_1 + \partial_1(M\mathscr{H}_1)]$$
$$+ \int_{S^1} X^1_{\ \alpha} N^\alpha[(\partial_1 M)\mathscr{H} + \partial_1(M\mathscr{H})] \tag{5.3}$$

and

$$[\mathbf{H}'(\vec{N}),H(M^1)]$$
$$= \int_{S^1} -\gamma^{-1/2}n_\alpha N^\alpha[(\partial_1 M^1)\mathscr{H} + \partial_1(M^1\mathscr{H})]$$
$$+ \int_{S_1} X^1_{\ \alpha} N^\alpha[(\partial_1 M^1)\mathscr{H}_1 + \partial_1(M^1\mathscr{H}_1)]. \tag{5.4}$$

Relations (5.3) and (5.4) define the remaining first-order structure functions. Notice that these functions are not simply constants, i.e., the Poisson algebra of the constraints (5.1) and (5.2) does not represent a Lie algebra. Hence it is possible that there exist higher-order structure functions. Fortunately, it can be verified by direct computation that these functions can be set to zero. This can be seen without undue labor by noticing that the nontrivial first-order structure functions are functionals of the embeddings only. Using this fact, it is easy to infer from the definition of the second-order structure functions (B8) that the second-order (and hence all higher-order) structure functions can be set to zero. [The crux of the argument is that there is no combination of constraints which involves only $X^\alpha(\sigma)$.] The BRST charge then takes the form

$$\Omega' = \mathscr{U}' + \int_{S^1} (\eta^\alpha \mathbf{H}'_\alpha + \eta\mathscr{H} + \eta^1\mathscr{H}_1), \qquad (5.5a)$$

where

$$\mathscr{U}' = \int_{S^1} \left(\frac{1}{2}\eta - \gamma^{-1/2}n_\alpha\eta^\alpha\right)[(\partial_1\eta)\mathscr{P}_1 + \partial_1(\eta\mathscr{P}_1) + (\partial_1\eta^1)\mathscr{P} + \partial_1(\eta^1\mathscr{P})] + \int_{S^1}\left(\frac{1}{2}\eta^1 + X^1_{\ \alpha}\eta^\alpha\right)[(\partial_1\eta)\mathscr{P} + \partial_1(\eta\mathscr{P}) + (\partial_1\eta^1)\mathscr{P}_1 + \partial_1(\eta^1\mathscr{P}_1)]. \qquad (5.5b)$$

It is straightforward, if somewhat painful, to check that the BRST transformation—the canonical transformation generated by Ω'—is a nilpotent transformation on $\widehat{\Gamma}'$:

$$[\Omega',\Omega'] = 0.$$

Given the (trivial) observables $\mathbf{H}'_\alpha$, there is a systematic way of obtaining their BRST invariant extensions (modulo the addition of cohomologically trivial functions).[13] However, with our choice of Ω', it is easy enough to guess their construction. The cohomologically trivial functions defined by

$$\widehat{\mathbf{H}}'_\alpha(\sigma) := [-\mathscr{P}_\alpha(\sigma),\Omega'] \qquad (5.6)$$

have all the properties that we desire: They are BRST invariant and equivalent (in the sense of BRST cohomology) to zero. Moreover, by direct calculation or using the results of Appendix B, they satisfy an Abelian Poisson algebra:

$$[\widehat{\mathbf{H}}'_\alpha(\sigma),\widehat{\mathbf{H}}'_\beta(\sigma')] = 0. \qquad (5.7)$$

Equation (5.7) is obtained thanks to the Abelian algebra satisfied by $\mathbf{H}'_\alpha$, from which the cubic ghost terms in Ω' are no more than linear in η^α.

Upon computation of the bracket in (5.6), we find

$$\widehat{\mathbf{H}}'_\alpha = \mathbf{H}'_\alpha - \gamma^{-1/2}n_\alpha\mathscr{H}^{gh} + X^1_{\ \alpha}\mathscr{H}^{gh}_1, \qquad (5.8)$$

where

$$\mathscr{H}^{gh} = (\partial_1\eta)\mathscr{P}_1 + \partial_1(\eta\mathscr{P}_1) + (\partial_1\eta^1)\mathscr{P} + \partial_1(\eta^1\mathscr{P}), \qquad (5.9)$$

$$\mathscr{H}^{gh}_1 = (\partial_1\eta)\mathscr{P} + \partial_1(\eta\mathscr{P}) + (\partial_1\eta^1)\mathscr{P}_1 + \partial_1(\eta^1\mathscr{P}_1). \qquad (5.10)$$

From (5.8) we see that $\widehat{\mathbf{H}}'_\alpha$ consists of the original diffeomorphism Hamiltonian plus the ghost corrections $\mathscr{H}^{gh}$ and $\mathscr{H}^{gh}_1$, which are generators for the ghost dynamical evolution associated with normal and tangential deformations of a given embedding. As such, the functions (5.9) and (5.10) obey the two-dimensional Dirac algebra. This fact guarantees (5.7) and we again have the homomorphic mapping from diff(M) into the Poisson algebra of functions on $\widehat{\Gamma}'$ given by

$$V \to \vec{V} \to \widehat{\mathbf{H}}'(\vec{V}),$$
$$[\widehat{\mathbf{H}}'(\vec{V}),\widehat{\mathbf{H}}'(\vec{U})] = \widehat{\mathbf{H}}'(-[\vec{V},\vec{U}]).$$

Given a choice of a space-time vector field $\vec{V}$ [or alternatively, given a choice of lapse and shift (N,N^1)], the Hamiltonian $\widehat{\mathbf{H}}'(\vec{V})$ [or $\widehat{\mathbf{H}}'(\vec{N})$] generates the appropriate dynamical evolution in $\widehat{\Gamma}'$. To recover the standard conformal gauge evolution in the BRST-extended phase space, we can choose the foliation to be that given by the auxiliary structure

$$X^\alpha(\sigma,\tau) = \widetilde{Y}^{(\alpha)}(\sigma,\tau),$$

which implies

$$N = 1, \quad N^1 = 0.$$

Any foliation described by this choice of lapse and shift will yield the conventional conformal gauge evolution of the string and ghost variables.

What has been done for Γ' can also be done for Γ''. The constraint surface is conveniently specified by

$$\mathbf{H}''_\alpha \approx 0, \qquad (5.11)$$

$$\mathscr{H} \approx 0 \approx \mathscr{H}_1, \qquad (5.12)$$

$$\lambda_{(\alpha)} \approx 0. \qquad (5.13)$$

The first-order structure functions associated with (5.11) and (5.12) are formally identical to (2.54)–(2.56) and (5.3) and (5.4) (with $\mathbf{H}'_\alpha \to \mathbf{H}''_\alpha$) except now $\gamma^{-1/2}n_\alpha$ and $X^1_{\ \alpha}$ are functionals of both $X^\alpha(\sigma)$ and $\mu^{(\alpha)}(\sigma)$ via (3.47)–(3.49). New structure functions arise from Poisson brackets between $\lambda_{(\alpha)}$ and $\mathbf{H}''_\alpha$. If we define

$$\lambda(\nu) := \int_{S^1} \nu^{(\alpha)}\lambda_{(\alpha)},$$

then

$$[\lambda(\nu),\mathbf{H}''(\vec{M})] = \int_{S^1} \nu^{(\alpha)}M^\beta[-\gamma^{-1}n_\beta n_\gamma (Y^{-1})^\gamma_{(\alpha)} \times(\mathscr{H} + Y^{(\epsilon)}{}_{,\delta}X^\delta{}_{,1}\lambda_{(\epsilon),1}) + (Y^{-1})^\gamma_{(\alpha)}\gamma^{-1/2}n_\beta X^1_{\ \gamma}(\mathscr{H}_1 + \mu^{(\delta)}\lambda_{(\delta),1}) + X^1_{\ \beta}\lambda_{(\alpha),1}]. \qquad (5.14)$$

The first-order structure functions include non-constant functions on Γ'' due to the appearance of $\gamma^{-1/2}n_\alpha$, $X^1_{\ \alpha}$, $Y^{(\alpha)}{}_{,\beta}$, and $(Y^{-1})^\alpha_{(\beta)}$. Nevertheless, as before, the second-order structure functions can be set to zero. This is essentially because there is no combination of the constraints that is a functional of $X^\alpha(\sigma)$ and $\mu^{(\alpha)}(\sigma)$ only.

The BRST charge will be a function on the phase space $\widehat{\Gamma}''$, which has local coordinates given by those chosen earlier for $\widehat{\Gamma}'$ along with $(\lambda_{(\alpha)}, \mu^{(\beta)})$ and the additional ghosts $(\eta^{(\alpha)}, \mathscr{P}_{(\beta)})$. The ghost $\eta^{(\alpha)}$ is a pair of (Grassman valued) scalar densities of weight 1, while its conjugate momentum $\mathscr{P}_{(\beta)}$ consists of two scalars: They satisfy the Poisson brackets

$$[\eta^{(\alpha)}(\sigma),\mathscr{P}_{(\beta)}(\sigma')] = -\delta^{(\alpha)}_{(\beta)}\delta(\sigma,\sigma').$$

In these coordinates Ω'' takes the form

$$\Omega'' = \mathscr{U}'' + \mathscr{U}' + \int_{S^1} (\eta^\alpha\mathbf{H}''_\alpha + \eta\mathscr{H} + \eta^1\mathscr{H}_1 + \eta^{(\alpha)}\lambda_{(\alpha)}), \qquad (5.15)$$

where

$$\mathscr{U}'' = \int_{S^1} \eta^{(\alpha)}\eta^{\beta} [-\gamma^{-1} n_\beta n_\gamma (Y^{-1})^{\gamma}_{(\alpha)} \times (\mathscr{P} + Y^{(\epsilon)}{}_{,\delta} X^{\delta}{}_{,1} \mathscr{P}_{(\epsilon),1}) + (Y^{-1})^{\gamma}_{(\alpha)} \gamma^{-1/2} n_\beta X^1_\gamma (\mathscr{P}_1 + \mu^{(\delta)} \mathscr{P}_{(\delta),1}) + X^1_\beta \mathscr{P}_{(\alpha),1)}]. \tag{5.16}$$

It is important to keep in mind that in (5.15) $\mathscr{U}'$ is given by (5.5b), but n_α and X^1_α are to be thought of as functionals of $X^\alpha(\sigma)$ and $\mu^{(\alpha)}(\sigma)$ via (3.47)–(3.49).

As before, a BRST invariant extension of $\mathbf{H}''_\alpha$ is obtained from the following Poisson bracket:

$$\widehat{\mathbf{H}}''_\alpha(\sigma) := [-\mathscr{P}_\alpha(\sigma), \Omega''] = \mathbf{H}''_\alpha(\sigma) - \gamma^{-1/2} n_\alpha \mathscr{H}^{gh} + X^1_\alpha \mathscr{H}^{gh}_1 + V''_\alpha, \tag{5.17}$$

where

$$V''_\alpha = -\gamma^{-1/2} n_\alpha \eta^{(\beta)} [-\gamma^{-1/2} n_\gamma (Y^{-1})^{\gamma}_{(\beta)} \times (\mathscr{P} + Y^{(\epsilon)}{}_{,\delta} X^{\delta}{}_{,1} \mathscr{P}_{(\epsilon),1}) + (Y^{-1})^{\gamma}_{(\beta)} X^1_\gamma (\mathscr{P}_1 + \mu^{(\delta)} \mathscr{P}_{(\delta),1})] + X^1_\alpha \eta^{(\beta)} \mathscr{P}_{(\beta),1}. \tag{5.18}$$

Once again, by virtue of the Abelian Poisson bracket algebra obeyed by $\mathbf{H}''_\alpha$, we have

$$[\widehat{\mathbf{H}}''_\alpha(\sigma), \widehat{\mathbf{H}}''_\beta(\sigma')] = 0.$$

The mapping

$$V \to \vec{V} \to \widehat{\mathbf{H}}''(\vec{V})$$

then serves to represent diff(M) in terms of cohomologically trivial observables on $\widehat{\Gamma}''$.

Finally, we turn to the phase space Γ^* of Sec. IV. The constraints on this phase space consist of the vanishing of the diffeomorphism Hamiltonians (4.20) and (4.21):

$$\mathbf{H}_\alpha \approx 0,$$

along with the "shift of origin" constraint (4.34):

$$R := \int_{S^1} ((X^+{}_{,1})^{-1} \tilde{a}_i \tilde{a}^i + (X^-{}_{,1})^{-1} a_i a^i) \approx 0.$$

These constraint functions satisfy an Abelian Poisson bracket algebra; hence the BRST charge takes its simplest possible form. In terms of canonical variables on $\widehat{\Gamma}^*$ consisting of the string variables of Sec. IV and the ghost canonical pairs $(\eta^\alpha, \mathscr{P}_\alpha)$, $(\eta_0, \mathscr{P}_0)$, where the latter variables are constants on S^1, we have

$$\Omega = \eta_0 R + \int_{S^1} \eta^\alpha \mathbf{H}_\alpha. \tag{5.19}$$

The BRST extensions of the constraint functions are just the functions themselves:

$$\widehat{\mathbf{H}}_\alpha = [-\mathscr{P}_\alpha, \Omega] = \mathbf{H}_\alpha, \tag{5.20}$$

$$\widehat{R} = [-\mathscr{P}_0, \Omega] = R. \tag{5.21}$$

We see that the representation of diff(M) on $\widehat{\Gamma}^*$ is a trivial extension of what was done in Sec. IV for Γ^*.

The investigations of McMullan and Paterson[14] have revealed that the Abelian constraint functions play an important role in arriving at consistent factor orderings for the non-Abelian (i.e., projected) constraint functions by using the BRST formalism in the quantum theory. It is therefore of interest to relate Ω defined above to the BRST charge $\widetilde{\Omega}$ associated with the projected constraints:

$$\widetilde{\Omega} = \mathscr{U} + \eta_0 R + \int_{S^1} (\eta \mathscr{H} + \eta^1 \mathscr{H}_1), \tag{5.22}$$

where

$$\mathscr{U} = \int_{S^1} [\eta((\partial_1 \eta)\mathscr{P}_1 + (\partial_1 \eta^1)\mathscr{P}) + \eta^1((\partial_1 \eta^1)\mathscr{P}_1 + (\partial_1 \eta)\mathscr{P})]. \tag{5.23}$$

The charges Ω and $\widetilde{\Omega}$ are related by a canonical transformation on $\widehat{\Gamma}^*$. Define

$$\eta^{(\pm)} = (\eta \pm \eta^1), \quad \mathscr{P}_{(\pm)} = \tfrac{1}{2}(\mathscr{P} \pm \mathscr{P}_1);$$

the canonical transformation is obtained by first performing (4.13)–(4.18) and then making the transformation

$$\eta^{\pm} = (X^{\pm}{}_{,1})\eta^{(\pm)}, \tag{5.24}$$

$$\mathscr{P}_{\pm} = (X^{\pm}{}_{,1})^{-1} \mathscr{P}_{(\pm)}, \tag{5.25}$$

$$P_{\pm} = \widetilde{P}_{\pm} + \partial_1 [(X^{\pm}{}_{,1})^{-1} \eta^{(\pm)} \mathscr{P}_{(\pm)}], \tag{5.26}$$

along with the identity transformation on the remaining variables.

In contrast to gauge theories, where constraints are *linear* in momenta, this transformation is not a point transformation; it necessarily mixes the original string coordinates and momenta through the functionals $X^{\pm}(\sigma)$. Note, however, that once the transformation (4.13)–(4.18) has been made, the subsequent Abelianizing transformation (5.24)–(5.26) *is* a point transformation on $\widehat{\Gamma}^*$. In other words, in a manifestly parametrized theory, the transformation from the projected constraint functions to the unprojected constraint functions is a point canonical transformation in the BRST phase space.

VI. THE ROLE OF THE CONFORMAL GROUP

We have now exhibited several string phase space representations of the Lie algebra diff(M). However, as pointed out in Sec. II, we cannot expect to be able to "exponentiate" the corresponding infinitesimal canonical transformations to yield a canonical representation of the full group Diff(M). The difficulty stems from the way in which the diffeomorphisms act on the spacelike embeddings [as in (2.28)] to produce, in general, nonspacelike embeddings.

One is naturally led to ask if there exist interesting subgroups of Diff(M) which *do* preserve the spacelike character of the embeddings. Generalizing for a moment to an n-dimensional manifold M^n with Cauchy surface Σ, the answer is essentially negative. It is true that if a metric g_{ab} on M^n admits Killing vectors, then the corresponding conserved charges do generate a subgroup which is representable on $Emb_g(\Sigma, M^n)$. To see this, consider two embeddings related by $\phi \in$ Diff(M^n):

$$X: \Sigma \to M^n, \quad \overline{X}: \Sigma \to M^n, \quad \overline{X} = \phi \circ X. \tag{6.1}$$

If the original embedding is spacelike, the induced metric

$$\gamma = X^*g$$

is positive definite. The metric induced on the new embedding is obtained via

$$\bar{\gamma} = \bar{X}^*g = (\phi\circ X)^*g = X^*(\phi^*g).$$

If ϕ is generated by a Killing vector, then

$$\phi^*g = g$$

and the new embedding is also spacelike. In fact, we see that all we really need is for ϕ to be a *conformal* isometry since then

$$\phi^*g = e^{2\Psi}g$$

for some nonvanishing function $\Psi: M^n \to R$. In this case the dynamical field of interest must have a Weyl invariant action functional in order to produce the appropriate conserved charge, which then serves to represent the conformal group of isometries.

For a generic space-time (M^n,g) there are no conformal isometries or at best, a finite-dimensional group of them. In our two-dimensional cylindrical space-time the situation is much more interesting. Indeed, the dimension and topology of $M = R \times S^1$ admit an infinite-dimensional group of conformal isometries irrespective of the metric (see Appendix A). The phase space of the Weyl invariant string will then carry a representation of this rather large group. As we have already seen, the conformal group serves as a symmetry group for the string. What is a bit more surprising is that this group can also play the role of a *dynamical* group. If $\mathrm{Conf}(M,g)$ is to allow this reinterpretation, we must show that *any* two spacelike embeddings can be linked by the action of a conformal isometry. This result, implicitly contained in the fact that the two-dimensional Dirac algebra is isomorphic to $\mathrm{conf}(M,g)$, can be directly demonstrated as follows.

Consider two spacelike embeddings related by a diffeomorphism as in (6.1). We will show that there exists a conformal isometry $\psi \in \mathrm{Conf}(M,g)$ such that

$$\bar{X} = \psi\circ X. \tag{6.2}$$

The proof is by construction. Introduce a conformal coordinate system $\mathrm{X}^{\pm}$ on M:

$$ds^2 = -e^{2\omega(X)}d\mathrm{X}^+ \, d\mathrm{X}^-;$$

for a manifold homeomorphic to R^2 this is a global coordinate chart. For our cylindrical space-time some care is needed at the points that are identified to produce circles. As summarized in Appendix A, within this chart conformal isometries are characterized by a pair of one-dimensional diffeomorphisms

$$\psi^{\pm} = \psi^{\pm}(\mathrm{X}^{\pm}) \tag{6.3}$$

[in the notation of Appendix A this is case (i); case (ii) is handled similarly]. Next, we use the fact that the null geodesics $\mathrm{X}^{\pm} = \mathrm{const}$ intersect each spacelike embedding (Cauchy surface) once and only once, so that both X^+ and X^- serve as good coordinates on the embedding. In particular, the functions $\mathrm{X}^{\pm} = X^{\pm}(\sigma)$ are invertible, i.e., we can form

$$\sigma = \sigma^+(\mathrm{X}^+)$$

or

$$\sigma = \sigma^-(\mathrm{X}^-).$$

Given any $\phi \in \mathrm{Diff}(M)$ that preserves the spacelike character of an embedding, we construct the corresponding $\psi \in \mathrm{Conf}(M,g)$ via

$$\psi^{\pm}(\mathrm{X}^{\pm}) = \phi^{\pm}(\sigma(\mathrm{X}^{\pm})), \tag{6.4}$$

where

$$\phi(\sigma) = \phi(X(\sigma)).$$

That the conformal isometry (6.4) does the job (6.2) is guaranteed by the identity

$$\bar{X}^{\pm}(\sigma) = \psi^{\pm}(X^{\pm}(\sigma)) = \phi^{\pm}(\sigma X^{\pm}(\sigma))$$
$$= \phi^{\pm}(\sigma) = \bar{X}^{\pm}(\sigma).$$

Thus in two dimensions, the symmetry group $\mathrm{Conf}(M,g)$ can in fact also be taken as the dynamical group. That this is a somewhat remarkable occurrence can be seen by comparing the present situation with that arising in canonical geometrodynamics: There, the Poisson algebra of constraints ("hypersurface deformation algebra") is not a true Lie algebra. With some labor (analogous to what was done in Sec. III) one can extract a realization of the dynamical algebra $\mathrm{diff}(M)$, but no subgroup of $\mathrm{Diff}(M)$ exists which can play the role of the dynamical group. In string theory, the hypersurface deformation algebra (2.54)–(2.56) is a true Lie algebra isomorphic to $\mathrm{conf}(M,g)$ and, as we have just seen, is associated with a group representation on the string phase space. It has, of course, been known for some time (and from various perspectives) that the conformal group plays a fundamental role in string theory. We now see that from the perspective of hypersurface dynamics, $\mathrm{Conf}(M,g)$ is a valid substitute for the would-be dynamical group $\mathrm{Diff}(M)$.

VII. DISCUSSION

The multiplicity of ways that we have uncovered to represent $\mathrm{diff}(M)$ for the canonical string each has its own advantages and disadvantages. It is amusing to note that all the ways of making the canonical string manifestly covariant rest on the validity of associated gauge choices, in particular the conformal, harmonic, and light-cone gauges. Using the conformal and harmonic gauges we were able to construct $\mathrm{diff}(M)$ comoments which were globally defined on the respective extended phase spaces Γ' and Γ''. Moreover, manifest (target space) Poincaré covariance is maintained throughout. The price we pay is that we must extend the usual string phase space. Whether or not these extended phase spaces have a useful role in string theory remains to be seen. We should point out that Γ' has featured in one approach to string field theory.[15] We have not investigated whether a corresponding approach using Γ'' is viable. From the gravitational perspective, Γ' is analogous to the extended phase space based on the Gaussian coordinate conditions used in Ref. 2. The space Γ'' serves as a useful model for the generalization of the techniques of Ref. 2 to the harmonic gauge in general relativity. Here the string is serving its purpose as a gravitational paradigm quite well: The techniques developed in Sec. III B have a straightforward generalization to geometrodynamics in the harmonic gauge.[16] We are thus able to tie up a loose end left dangling in Ref. 2.

When we try to represent diff(M) on the usual string phase space Γ, a fairly large price must be paid. First, strictly speaking, the use of an enlarged phase space (Γ^*) is necessary owing to the "shift of origin" symmetry which remains after passing to the light-cone gauge (this problem is absent for the open string). The definition of the embedding momentum P_S that accompanies the enlargement is somewhat cumbersome: From (4.16) we see that P_S is a distribution (off the constraint surface). Ultimately, this feature of the formalism has its origin in our use of local coordinates on the circle. A global, coordinate-independent treatment would make the extraction of the embedding phase space a bit more elegant. We will present this improvement elsewhere. A more serious drawback of using Γ^* to represent diff(M) is that the diffeomorphism Hamiltonians are not globally defined on Γ^*. As was shown, this difficulty is intimately associated with the failure of the light-cone gauge. Finally, it is necessary to break manifest Lorentz invariance in order to extract the embedding phase space from Γ^*. Classically this is of no real consequence. By translating the usual Poincaré group generators using (4.13)–(4.18), one can show that the Lie algebra is still realized. A corresponding statement may not survive quantization, as will be discussed below.

A preliminary investigation of the quantum mechanical representation of diff(M) is given in Ref. 5. Let us summarize here the gist of this work. The strategy of Ref. 5 is to use the fact that, as shown here, the canonical string can be explicitly reformulated as a parametrized field theory. Quantization of parametrized scalar fields on cylindrical space-times has been extensively studied in Ref. 17. Using the techniques developed there, we have reached the following results. The quantization based on Γ' does allow for an anomaly-free representation of diff(M) for any value of the target space dimension d. Thus it is consistent to select physical states by requiring that they be annihilated by the (suitably ordered) quantum mechanical diffeomorphism Hamiltonians. However, one still must impose the original constraints (2.47) and (2.48). To do this consistently, one must therefore also represent the conformal *symmetry* group on the string Hilbert space. This can only be done projectively. Thus within the canonical framework we recover the results of Polyakov[3]: diff(M) covariance is maintained at the expense of conformal symmetry. As expected, by incorporating the BRST ghosts, according to Sec. V, Conf(M,g) *can* be represented without anomaly in the critical dimension $d = 26$. The quantization based on Γ'' has not been followed in Ref. 5.

When we study the quantization based on Γ^*, we find again that diff(M) can be represented without anomaly irrespective of the value of d. Since the diffeomorphism Hamiltonians are now just combinations of the original constraint functions (2.47) and (2.48) [*modulo* (4.11)], this means that a true "Dirac quantization" is possible within this framework. The symmetry group Conf(M,g) is still projectively represented (now only for the transverse variables), but this is of no consequence for quantization *à la* Dirac. Of course, it is necessary to sacrifice manifest Lorentz invariance to achieve this result. If it could be shown that Lorentz invariance is still present, we would have a new, nontrivial quantization of the string away from the critical dimension. (Of course, we do not maintain that this quantization needs to be useful from the point of view of elementary particle physics.) From the quantum gravity perspective, the role of the Lorentz group in the quantization based on Γ^* is equally interesting. The quantization procedure that allows diff(M) to be represented without anomaly necessarily treats the kinematic variables, i.e., the embeddings, quite differently from the dynamical variables, i.e., the transverse variables [*modulo* (4.11)]. The Lorentz group, if it does act in the quantum theory, mixes the kinematical "many-fingered time" variables with the dynamical fields. Thus we have, in a rather simple setting, a good model for exploring the role of time in quantum gravity. At this writing it is unknown whether the Lorentz group is represented in the critical dimension or otherwise. This issue will be investigated in a future publication.

ACKNOWLEDGMENT

This work was supported in part by NSF Grant No. PHY 8503653 to the University of Utah.

APPENDIX A: THE TWO-DIMENSIONAL CONFORMAL GROUP

Here we present some needed results from two-dimensional (Lorentzian) geometry as applied to $M = R \times S^1$. Here M can be defined as the Cartesian product of $(-\infty, \infty)$ and $[0,2\pi]$, with the points 0 and 2π identified:

$$M = (-\infty, \infty) \times [0,2\pi].$$

Coordinates $X^0 = T \in (-\infty, \infty)$ and $X^1 = S \in [0,2\pi]$ will be referred to as "standard coordinates" on M.

As is well known, because the Euler number of M is zero, every metric on M is conformal to a flat metric:

$$g_{ab} = e^{2\omega}\eta_{ab}, \quad R(\eta) = 0.$$

Here ω is a function on M. Without further assumptions, neither ω nor η_{ab} is unique. Given a choice for ω and η_{ab}, we can introduce conformal coordinates, i.e., standard coordinates on M such that the line element takes the form

$$ds^2 = e^{2\omega(X)}(-dT^2 + dS^2).$$

Alternatively, we can introduce null coordinates $X^{\pm} = T \pm S$ such that

$$ds^2 = -e^{2\omega(X)} dX^+ \, dX^-. \tag{A1}$$

The curves $X^+ = \text{const}$ and $X^- = \text{const}$ are null geodesics. (*Every* null curve in two dimensions is a geodesic.) Nonaffine (in general) parameters for $X^{\pm} = \text{const}$ are provided by $X^{\mp}$.

The group Conf(M,g) of conformal isometries is that subgroup of the diffeomorphism group Diff(M) which has the effect of rescaling the metric by a function. More precisely, if $\psi \in \text{Conf}(M,g)$, then

$$(\psi^* g)_{ab} = e^{2\Psi} g_{ab} \tag{A2}$$

for some $\Psi: M \to R$. Clearly, $\psi \in \mathrm{Conf}(M,g)$ must also serve as a conformal isometry for the conformally related flat metric, i.e., $\psi \in \mathrm{Conf}(M,\eta)$. A coordinate expression ψ^α for ψ is most easily obtained in terms of the null coordinates $X^\pm$. The condition (A2) for the metric (A1) becomes

$$\psi^+{}_{,+}\, \psi^-{}_{,+} = 0, \tag{A3}$$

$$\psi^+{}_{,-}\, \psi^-{}_{,-} = 0, \tag{A4}$$

$$\psi^+{}_{,+}\, \psi^-{}_{,-} + \psi^+{}_{,-}\, \psi^-{}_{,+} = e^{2\Psi(X)}. \tag{A5}$$

The transformation is nonsingular provided that

$$J = \psi^+{}_{,+}\, \psi^-{}_{,-} - \psi^+{}_{,-}\, \psi^-{}_{,+} \neq 0.$$

From these requirements we have either

$$\text{(i)}\ \psi^\pm = \psi^\pm(X^\pm), \quad \psi^+{}_{,+}\, \psi^-{}_{,-} > 0 \quad (J>0)$$

or

$$\text{(ii)}\ \psi^\pm = \psi^\pm(X^\mp), \quad \psi^+{}_{,-}\, \psi^-{}_{,+} > 0 \quad (J<0).$$

Infinitesimal transformations in $\mathrm{Conf}(M,g)$ are generated by conformal Killing vectors which, as usual, form the corresponding Lie algebra $\mathrm{conf}(M,g)$. In the null coordinates introduced above, the conformal Killing equation

$$\mathscr{L}_v g_{ab} = 2\alpha(X) g_{ab}$$

implies

$$v^+{}_{,-} = 0 = v^-{}_{,+}.$$

Thus $v^\pm = v^\pm(X^\pm)$ and we conclude that $\mathrm{Conf}(M,g)$ is disconnected: Type (i) transformations are connected to the identity and the corresponding Lie algebra is isomorphic to $\mathrm{diff}(S^1) \oplus \mathrm{diff}(S^1)$. Type (ii) transformations are not connected to the identity.

Finally, we note that a conformal isometry is completely determined by its action on the Cauchy surfaces $T = \text{const}$ in M. This is obvious from the coordinate forms we have presented for ψ; given $\psi^\pm(S)$, type (i) transformations are determined by

$$\psi^\pm = \psi^\pm(S = X^\pm)$$

and type (ii) transformations are given by

$$\psi^\pm = \psi^\pm(S = X^\mp).$$

This result can be generalized to any Cauchy surface

$$X^\pm = X^\pm(\sigma) \tag{A6}$$

by using the techniques of Sec. VI. Thus given the restriction of a conformal isometry to the circle (A6), i.e., given $\psi^\pm(\sigma)$, we invert the functions in (A6) and obtain the form of ψ valid on all M:

$$\psi^\pm(X^\pm) = \psi^\pm(\sigma(X^\pm))$$

or

$$\psi^\pm(X^\mp) = \psi^\pm(\sigma(X^\mp)).$$

APPENDIX B: BACKGROUND MATERIAL ON THE BRST FORMALISM

In this Appendix we give background material on the BRST formalism needed for Sec. V. The Hamiltonian approach to BRST results from the work of Batalin *et al.*[18] The formalism is described in considerable detail by Henneaux[13] and for the most part we will follow his notation and conventions. For the purpose of this Appendix it will be necessary to suspend the notation developed in the main part of the present paper.

At the classical level, the BRST formalism is a mathematically elegant way of capturing the additional structure that arises for Hamiltonian systems when one introduces first class constraints.[1] Recall that the basic ingredients of such systems consist of a phase space Γ and a constraint surface $\overline{\Gamma} \subset \Gamma$ which can be specified (locally) by the vanishing of a set of functions on Γ:

$$\phi_\alpha = 0; \tag{B1}$$

these functions satisfy a closed Poisson bracket algebra

$$[\phi_\alpha, \phi_\beta] = C^\gamma_{\alpha\beta} \phi_\gamma, \tag{B2}$$

where in general, $C^\gamma_{\alpha\beta}$ are nonconstant functions on Γ. All physical dynamics take place on $\overline{\Gamma}$. However, as a result of (B2) the Hamiltonian vector fields associated with ϕ_α are both tangent and normal to $\overline{\Gamma}$; hence $\overline{\Gamma}$ cannot be a symplectic submanifold of Γ. The physical symplectic manifold can be defined by a complete set of "observables," which are functions on Γ which are weakly "gauge invariant" in the sense that the variation of the observables associated with the infinitesimal canonical transformations generated by the constraint functions vanishes upon restriction to $\overline{\Gamma}$. Thus if $A: \Gamma \to R$ is an observable, we have

$$[A, \phi_\alpha] = \Lambda^\beta_\alpha \phi_\beta \tag{B3}$$

for some functions Λ^β_α. Any two observables that differ by a combination of constraint functions are considered equivalent:

$$A \sim A + \Lambda^\alpha \phi_\alpha, \tag{B4}$$

where Λ^α are functions on Γ. In particular, the functions ϕ_α are equivalent to zero. In gauge theories the observables are simply the (weakly) gauge invariant functions. In parametrized theories, which are of central interest here, the "observables" can be identified with the constants of motion.

All the pertinent information contained in the constraints (e.g., the constraint algebra, etc.) can be neatly stored in a single function on an extended phase space $\hat{\Gamma} \supset \Gamma$; this function is the BRST charge Ω. Here $\hat{\Gamma}$ can be obtained by introducing the Grassmann-valued "ghosts" η^α and their conjugate momenta $\mathscr{P}_\alpha$. The only nonvanishing Poisson brackets involving these new variables are

$$[\eta^\alpha, \mathscr{P}_\beta] = -\delta^\alpha_\beta.$$

Brackets between two Grassmann-odd functions on $\hat{\Gamma}$ are symmetric; otherwise, the Poisson brackets are, as usual, antisymmetric in their arguments. The BRST charge can be specified by two conditions:

$$\Omega(\mathscr{P}_\alpha = 0) = \eta^\alpha \phi_\alpha, \tag{B5}$$

$$[\Omega, \Omega] = 0. \tag{B6}$$

The requirement (B5) simply informs Ω about the existence of $\overline{\Gamma}$. The requirement (B6) then summarizes the algebra (B2), as well as all the identities that arise by taking repeated Poisson brackets of (B2) with the constraint functions and using the Jacobi identity. The solution of (B5) and (B6) is given by[13]

$$\Omega = \sum_{n=0}^{R} U^{\beta_1\cdots\beta_n}_{\alpha_0\cdots\alpha_n}\eta^{\alpha_n}\cdots\eta^{\alpha_0}\mathscr{P}_{\beta_n}\cdots\mathscr{P}_{\beta_1}. \tag{B7}$$

The integer R is known as the *rank* of the constraint functions ϕ_α. The coefficients U are functions on Γ and are known as the *structure functions* for the constraint algebra. In particular, we have the zeroth-order structure functions

$$U_\alpha = \phi_\alpha$$

and the first-order structure functions

$$U^\gamma_{\alpha\beta} = -\tfrac{1}{2}C^\gamma_{\alpha\beta}.$$

The general expressions for the higher-order structure functions are given in Ref. 13. We will only need to know the definition of the second-order structure functions. They can be defined via

$$2U^{\delta\epsilon}_{\alpha\beta\gamma}\phi_\epsilon = D^\delta_{[\alpha\beta\gamma]}, \tag{B8a}$$

where

$$D^\delta_{\alpha\beta\gamma} = [U^\delta_{\alpha\beta},\phi_\gamma] + 2U^\kappa_{\alpha\beta}U^\delta_{\gamma\kappa}. \tag{B8b}$$

All BRST charges constructed in this paper are of rank 1. This means that the second-order—and all higher-order—structure functions can be set to zero. Note that for the rank to be 1 it is sufficient, but not necessary that the first-order structure functions are constants on Γ.

It can be shown[13] that any observable A defined by (B3) and (B4) can be extended to a function $\hat{A}$ on $\hat{\Gamma}$ which is BRST invariant, i.e., invariant under the nilpotent canonical transformation generated by Ω:

$$[\hat{A},\Omega] = 0. \tag{B9}$$

The extension of A is unique up to the addition of terms of the form

$$B = [\Lambda,\Omega], \tag{B10}$$

where Λ is a function on $\hat{\Gamma}$. Thus it is natural to define BRST observables as equivalence classes of functions which are BRST invariant; two functions belong to the same equivalence class if they differ by a "cohomologically trivial" function of the form (B10):

$$\hat{A} \sim \hat{A} + [\Lambda,\Omega]. \tag{B11}$$

The cohomologically trivial functions are observables, but they are equivalent to zero. Note that (B9) is the BRST generalization of (B3) and (B11) is the generalization of the equivalence relation (B4).

In Sec. V we are interested in computing the BRST extensions of the diffeomorphism Hamiltonians which, being constraint functions, are trivial observables in the usual framework. Their extension should likewise be trivial, i.e., of the form (B10). The key feature of the BRST-extended diffeomorphism Hamiltonians is that they take the form of the original diffeomorphism Hamiltonians plus ghost representatives of diff(M). The extended functions must still satisfy an Abelian Poisson bracket algebra. To see how this comes about, consider partitioning the constraints ϕ_α as

$$\phi_\alpha = (H_A,\chi_a),$$

where H_A are to play the role of the unextended diffeomorphism Hamiltonians and χ_a are any remaining constraint functions. Assuming ϕ_α are of rank 1, we have

$$\Omega = \eta^\alpha\phi_\alpha + U^\gamma_{\alpha\beta}\eta^\beta\eta^\alpha\mathscr{P}_\gamma. \tag{B12}$$

The BRST extension of H_A is obtained via

$$\hat{H}_A = -[\mathscr{P}_A,\Omega] = H_A - 2U^\gamma_{A\beta}\eta^\beta\mathscr{P}_\gamma.$$

We are interested in computing the Poisson brackets between $\hat{H}_A$. Using the (graded) Jacobi identity we find

$$\begin{aligned}[\hat{H}_A,\hat{H}_B] &= [\mathscr{P}_B,[\Omega,\hat{H}_A]] - [\Omega,[\hat{H}_A,\mathscr{P}_B]] \\ &= 2U^\gamma_{AB}\hat{\phi}_\gamma + 2\mathscr{P}_\gamma[\Omega,U^\gamma_{AB}] \\ &\quad + [\mathscr{P}_B,[\Omega,\hat{H}_A]],\end{aligned} \tag{B13}$$

where

$$\hat{\phi}_\gamma := -[\mathscr{P}_\gamma,\Omega].$$

The third term in (B13) can be eliminated, since $\hat{H}_A$ is cohomologically trivial, by using the Jacobi identity

$$[\Omega,\hat{H}_A] = \tfrac{1}{2}[\mathscr{P}_A,[\Omega,\Omega]] = 0.$$

The second term in (B13) vanishes if U^γ_{AB} are constants. For our purposes, we are interested in the situation where these functions actually vanish. In this case, the first term also vanishes and we therefore find

$$[\hat{H}_A,\hat{H}_B] = 0.$$

This result is crucial for our representation of diff(M) on $\hat{\Gamma}'$ and $\hat{\Gamma}''$ in Sec. V.

APPENDIX C: NOTATION AND CONVENTIONS

A. Indices

Latin indices from the beginning of the alphabet (a, b, c, etc.) are abstract indices for tensors on M. Two-dimensional coordinate basis indices are denoted by Greek letters from the beginning of the alphabet (α, β, γ, etc.); they take the values zero and 1. Indices for preferred coordinates associated with a choice of gauge are further distinguished by parentheses [(α), (β), etc.]. General target space indices are denoted as upper case Latin letters. When we have a flat Minkowski target space, as for the string, we use Greek letters from the latter part of the alphabet to denote components with respect to an inertial coordinate frame. The spacelike directions orthogonal to a pair of null directions in the target Minkowski space are labeled by the Latin indices i, j, k, etc.

B. Coordinates

Generic coordinates on M are denoted X^α. Preferred coordinates are labeled $Y^{(\alpha)}$. Coordinates on $R\times S^1$ are denoted as $\sigma^\alpha = (\tau,\sigma)$ where $\tau \in (-\infty,\infty)$ and $\sigma \in [0,2\pi]$, with $\sigma = 0$ and $\sigma = 2\pi$ identified.

C. Metrics and derivative operators

We use g_{ab} to denote a metric on M: It has the signature $(-\,+)$. Note that this metric is *a priori* independent of the induced metric on the string world sheet. The two-dimensional Minkowski metric in inertial coordinates is denoted $\eta_{\alpha\beta}$. The general target space metric is denoted G_{AB}. The target space Minkowski metric in inertial coordinates is denoted $\eta_{\mu\nu}$. The signature here is $(-\,+\,+\cdots+)$. We use the symbols γ and $\tilde{\gamma}$ to represent one-dimensional metrics induced on spacelike embeddings of circles in M.

The derivative operator ∇_a on M is torsion-free and compatible with g_{ab}. The Lie derivative with respect to the vector V^a is denoted $\mathscr{L}_V$. We use ∂_1 and the associated comma notation to denote partial derivatives on the circle with respect to the coordinate σ.

D. Groups and Lie algebras

The diffeomorphism group of the manifold M is denoted Diff(M). The generic element is $\phi\in$Diff(M). The associated Lie algebra is diff(M), with the typical elements U, V, $W\in$diff(M). Similarly, the group of conformal isometries of the metric g_{ab} on M is denoted Conf(M,g), with the generic element $\psi\in$Conf(M,g). The Lie algebra of Conf(M,g) is conf(M,g), with the typical elements v, $w\in$conf(M,g).

E. Deformation vectors

Unless otherwise stated, the symbol N^α stands for an externally prescribed ("c number") smearing field. The lapse function $N^\perp$ and shift vector N^1, obtained by projecting N^α, are, respectively, a scalar function and a vector on S^1. The symbol N denotes the lapse density, which is $N^\perp$ rescaled to be a density of weight minus 1 on the circle. Smearing fields that are obtained by restriction of a space-time vector V^a to an embedding retain the same root letter; these fields are functionals of the embeddings (i.e., "q numbers").

F. Phase spaces

We use Γ to denote the usual string phase space. Its slight extension, as defined in Sec. IV, is denoted Γ^*. We obtain Γ' by adding to Γ the embedding phase space in conjunction with the conformal gauge. We obtain Γ'' by adding embedding and Lagrange multiplier phase spaces to Γ in conjunction with the harmonic gauge. We denote the phase space for harmonic maps from M into M^d by Γ_0 and Γ_0' is its embedding-extended counterpart. A caret (e.g., $\widehat{\Gamma}$) over any of the above phase spaces denotes their BRST extensions. An overbar (e.g., $\overline{\Gamma}$) denotes the submanifold obtained by imposing all relevant constraints.

G. Brackets

The symbol [,] is used to represent both Poisson brackets and vector field commutators. Which meaning is intended in a given expression should be clear from the context.

H. The symbol X

The symbol X is used in a variety of ways and its meaning should be clear from the context in which it is being used. An embedding of a circle into M is denoted abstractly as X. Its coordinate form is $X^\alpha(\sigma)$. A one-parameter family of such embeddings, i.e., a foliation of M, is also denoted abstractly as X: Its coordinate form is $X^\alpha(\sigma,\tau)$. As described above, X^α represent generic coordinates on M.

[1]See, e.g., A. J. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Academia Nazionale dei Lincei, Rome, 1976); K. Sundermeyer, *Constrained Dynamics* (Springer, New York, 1982) and references therein.
[2]C. J. Isham and K. V. Kuchař, Ann. Phys. (NY) **164**, 288, 316 (1985).
[3]A. Polyakov, Phys. Lett. B **103**, 207, 211 (1981).
[4]See, e.g., M. Green, J. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U.P., Cambridge, 1987); L. Brink and M. Henneaux, *Principles of String Theory* (Plenum, New York, 1988) and references therein.
[5]K. V. Kuchař and C. G. Torre, "Strings as Poor Relatives of Relativity," University of Utah preprint (1988).
[6]K. V. Kuchař, J. Math. Phys. **13**, 768 (1972).
[7]L. Balieu, W. Siegel, and B. Zwiebach, Nucl.Phys. B **287**, 93 (1987).
[8]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva Univ., New York, 1964).
[9]P. A. M. Dirac, Proc. R. Soc. London Ser. A **246**, 333 (1958); R.Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L.Witten (Wiley, New York, 1962).
[10]K. V. Kuchař, J. Math. Phys. **19**, 390 (1978).
[11]M. Bowick and S. G. Rajeev, Phys. Rev. Lett. **58**, 535 (1987).
[12]A. Patracioiu, Nucl. Phys. B **81**, 525 (1974); see, also, the second reference of Ref. 4.
[13]M. Henneaux, Phys. Rep. **126**, 1 (1985).
[14]D. McMullan and J. Paterson, Phys. Lett. B **202**, 358 (1988).
[15]S. R. Das and M. A. Rubin, Phys. Lett. B **169**, 182 (1986).
[16]K. V. Kuchař and C. G. Torre, in preparation.
[17]K. V. Kuchař, "Parametrized Scalar Field on $R\times S^1$: Dynamical Pictures, Space-time Diffeomorphisms and Conformal Isometries" and "The Dirac Constraint Quantization of a Parametrized Field Theory by Anomaly-Free Operator Representations of Spacetime Diffeomorphisms," University of Utah preprints (1988).
[18]E. S. Fradkin and G. A. Vilkovisky, CERN Rep. TH2332 (1977); I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977).

Geodesic deviation in the Schwarzschild space-time

Stanisław L. Bażański
Institute of Theoretical Physics, Warsaw University, ul. Hoża 69, 00-681 Warsaw, Poland
Piotr Jaranowski
Institute of Physics, Warsaw University Division, ul. Lipowa 41, 15-424 Białystok, Poland

(Received 17 March 1989; accepted for publication 12 April 1989)

A general solution of the geodesic deviation equations in the Schwarzschild space-time is presented. It is found in a rigorous form with the aid of a recently formulated method of integration of the Hamilton–Jacobi type, and is applied to a study of the behavior of a cloud of particles freely falling into the Schwarzschild singularity. Presuming that the cloud of such particles provides one with a measure of strength of the singularity of a field, a similar problem of freely falling particles is also solved in the Newtonian theory. It has turned out that the behavior of the cloud of particles freely falling into the center of attraction is, contrary to what one would rather expect, in each of the two theories exactly the same. A possible reason for obtaining such a result is briefly discussed.

I. INTRODUCTION

Recently a new method of finding solutions to the geodesic deviation equations in general relativity has been formulated.[1] The method is based on theorems that are a generalization of the Jacobi theorem on the complete integral of the Hamilton–Jacobi equation for geodesics. As a result of these theorems, the knowledge of a complete integral of the Hamilton–Jacobi equation for geodesics enables one to find the general solution of the geodesic deviation equations, with no necessity of any additional integration.

The procedure of finding such a solution can be described in the following way. Let $(M,g_{\alpha\beta})$ be a four-dimensional space-time parametrized in a local neighborhood by the coordinates x^α. Suppose that the function $U(x^\alpha,a^k)$, where a^k are some three ($k=1,2,3$) additional parameters, is a complete integral of the Hamilton–Jacobi equation,

$$g^{\alpha\beta}\frac{\partial U}{\partial x^\alpha}\frac{\partial U}{\partial x^\beta}=1, \tag{1.1}$$

for timelike geodesics on M. Suppose further that one is interested in finding the general solution of the system of geodesic deviation differential equations,

$$\frac{D^2r^\alpha}{ds^2}+R^\alpha{}_{\beta\gamma\delta}u^\beta r^\gamma u^\delta=0, \tag{1.2}$$

with the first integral

$$u_\alpha r^\alpha=\beta_0. \tag{1.3}$$

In Eqs. (1.2) and (1.3), all the coefficients are evaluated along a known timelike geodesic Γ, called also the basic geodesic, which is described by equations of the form

$$x^\alpha=\xi^\alpha(s,a^k,\alpha_l), \tag{1.4}$$

where s is the proper time parameter along Γ, and a^k and α_l are six integration constants determined by initial conditions (the seventh integration constant is hidden in the choice of the initial value s_0 of the proper time). The quantities u^α in Eqs. (1.2) are components of the vector tangent to Γ,

$$u^\alpha=\frac{d\xi^\alpha}{ds}, \tag{1.5}$$

and the four unknowns r^α in these equations, treated as functions of s, are components of the geodesic deviation vector. The kinematic interpretation of this vector and the meaning of the condition (1.3) are well known. Among others, they were discussed in a former paper by one of us.[2]

In Ref. 1 an algorithm has been formulated which permits one to find a general solution to Eqs. (1.2), provided that a complete integral $U(x^\alpha,a^k)$ of Eq. (1.1) is known, and known as well are the functions ξ^α in Eqs. (1.4) which are describing a given basic geodesic Γ. Furthermore, the values of the parameters a^k in the complete integral must be set equal to the constants a^k in Eq. (1.4). In the case when the basic geodesic and the deviation vector are parametrized by the proper time s, the procedure of finding the solution r^α of the deviation equations (1.2), described in Ref. 1, reduces to solving the following system of four inhomogeneous linear algebraic equations for four unknown quantities r^α,

$$\frac{\partial^2U}{\partial a^k\partial x^\alpha}r^\alpha+b^n\frac{\partial^2U}{\partial a^k\partial a^n}=\beta_k,\quad k=1,2,3, \tag{1.6}$$

$$\frac{\partial U}{\partial x^\alpha}r^\alpha=\beta_0, \tag{1.7}$$

where β_0 is the same constant as in Eq. (1.3) and β_k are some new arbitrary constants. It is understood that the derivatives of the function $U(x^\alpha,a^k)$ in Eqs. (1.6) and (1.7) are evaluated along the given basic geodesic Γ, i.e., for x^α given by Eq. (1.4). Equations (1.6) and (1.7) thus constitute a set of four linear algebraic equations whose solution is of the form

$$r^\alpha=\rho^\alpha(s,a^k,\alpha_l,b^m,\beta_n,\beta_0), \tag{1.8}$$

depending on 13 constants a^k, α_l, b^m, β_n, β_0. In accordance with the procedure formulated in Ref. 1, the functions ρ^α found in such a way are the general solution of Eqs. (1.2).

Due to Eq. (1.7), this solution satisfies the subsidiary condition

$$u_\alpha\frac{Dr^\alpha}{d\tau}=0, \tag{1.9}$$

which means (cf. Refs. 1 and 2) that the scalar product of

 0022-2488/89/081794-10$02.50

the four-velocity u and of the deviation vector r is a constant of the motion

$$g_{\alpha\beta} \frac{dx^\alpha}{ds} r^\beta = \pm \beta_0. \tag{1.10}$$

One of the objectives of the present paper is to apply the above procedure of finding solutions to the geodesic deviation equation to timelike geodesics in the Schwarzschild space-time of general relativity. This is preceded in Sec. II by a general discussion of how a restriction of all the geodesics from a neighborhood of the basic geodesic Γ to those belonging only to a subfamily, which is realized by introducing constraints on the integration constants a^k and α_l in Eq. (1.4), induces constraints on the constants b^k and β_l in Eq. (1.8). This discussion is rather a completion to the general formalism presented in Ref. 1. Its results are employed in the following sections of the paper.

The explicit form of the general solution of the geodesic deviation equations in the Schwarzschld space-time is derived in Sec. III. This section also contains the analysis of the behavior of the deviation field on the horizon.

Section IV deals with a restriction of the general solution to two more special cases. In the first case, the general solution is reduced to the solution describing deviations between geodesics lying in the equatorial plane. Such a specialized solution is in Sec. IV compared to a solution found by Fuchs[3] by a direct integration of the geodesic deviation equations in the Schwarzschild space-time.

In the second case considered in Sec. IV, the general solution is restricted to the solution describing deviations between radial geodesics which represent the motion of particles falling freely and radially onto the singularity. This solution is then used in Sec. V in the discussion of the asymptotic behavior, for small values of r, of a cloud of such particles. The result obtained here is in full accordance with that obtained by a quite different approach in §32.6 in the textbook by Misner, Thorne, and Wheeler.[4]

The same problem of the free and radial fall of a cloud of particles onto the singularity is, in Sec. VI, discussed in the framework of the Newtonian theory of gravitation, giving basically the same asymptotic behavior as in the relativistic case. Since the behavior of such a cloud could be regarded as a kind of measure of strength of the singularity, and intuitively the relativistic singularity seems to be stronger than the Newtonian one, the result obtained in Sec. VI might be considered to be a rather unexpected one. A discussion of possible reasons for obtaining such a result concludes this section.

As was already observed in Ref. 2, the knowledge of the deviation field along a geodesic enables one to discuss several optical effects, and in particular the frequency shift, in the case when the moving adjacent particles exchange light signals. Thus the solution representing geodesic deviation between radial geodesics, which has been found in Sec. IV, can be immediately applied to a discussion of the Doppler tracking from the basic geodesic of adjacent particles which are falling freely and radially onto the singularity. A discussion of such a kind, in both the relativistic and the Newtonian cases, is presented in Sec. VII.

In the computations performed in this paper, we additionally assume that the constant β_0 in Eq. (1.7) vanishes, which is an insignificant restriction on the generality of our consideration and can always be achieved by an appropriate choice of the initial condition imposed on the field $r^\alpha(s)$.

In general relativity, one considers geodesic deviation vectors that are defined along a timelike geodesic parametrized by its proper time s and satisfying the condition (1.10). Vectors of such a kind describe the relative motion of geodesic observers that employ ideal clocks (i.e., their geodesics are also parametrized by the proper time) and move in a close neighborhood of the basic geodesic. The assumption that the time constant β_0 in condition (1.10) vanishes means that the vector $r^\alpha(s)$ describes the position of a neighboring observer in the local inertial comoving frame of the basic observer. This local inertial frame is determined by a tetrad of vectors consisting of the four-velocity vector u^α and three linearly independent spacelike vectors which are normal to the four-velocity and are propagated parallel along the basic geodesic.

II. GEODESIC DEVIATIONS OF GEODESICS BELONGING TO A SELECTED SUBFAMILY OF ALL NEIGHBORING GEODESICS

The solution (1.8) of the system of Eqs. (1.6) and (1.7) determines the components of the geodesic deviation field which is defined along an arbitrary timelike basic geodesic Γ and describes deviations pointing to all geodesics from the full six-parameter family of geodesics in a neighborhood of Γ. Sometimes one might be interested only in deviations pointing to geodesics which belong to some subfamilies of the set of all geodesics. We shall now show how the general solution of geodesic deviation equations can be reduced to a solution that contains only deviations pointing to geodesics from a selected subfamily. We start with the case of a one-parameter family of geodesics.

Let $U = U(x^\alpha, a^k)$ be a complete integral of the geodesic Hamilton–Jacobi equation. Then, in virtue of Theorem 4.1 from Ref. 1, the system of equations

$$\frac{\partial U}{\partial a^k}(x^\beta, a^k) = \alpha_k, \quad k = 1,2,3, \tag{2.1}$$

together with the normalization condition

$$g_{\alpha\beta} \frac{d\xi^\alpha}{ds} \frac{d\xi^\beta}{ds} = 1,$$

determines four functions ξ^α such that the world lines

$$x^\alpha = \xi^\alpha(s, a^k, \alpha_l), \tag{2.2}$$

where s is the proper time along the world line (2.2), are geodesics for all fixed values of a^k and α_l taken from an admissible domain. Substituting (2.2) into (2.1), we obtain an identity with respect to the variables s, a^k, and α_l,

$$\frac{\partial U}{\partial a^k}(\xi^\beta(s, a^l, \alpha_m), a^n) = \alpha_k, \quad k = 1,2,3. \tag{2.3}$$

Let us now take a curve in the six-dimensional space of parameters (a^k, α_l),

$$\mathbb{R} \supset [\rho_1, \rho_2] \ni \rho \to (a^k(\rho), \alpha_l(\rho)) \in \mathbb{R}^6. \tag{2.4}$$

The curve (2.4) is used to select, from the set of all geodesics described by Eqs. (2.2), a one-parameter subfamily. Geodesics being members of this subfamily are labeled by values of the parameter ρ and are described, because of (2.2) and (2.4), by equations of the form

$$x^\beta = \xi^\beta(s,a^k(\rho),\alpha_l(\rho)) =: \Xi^\beta(s,\rho). \tag{2.5}$$

The basic geodesic Γ can be chosen to be labeled by the value $\rho = 0$. Then the deviation field r^α along Γ is defined as

$$r^\beta := \left.\frac{\partial \Xi^\beta(s,\rho)}{\partial \rho}\right|_{\rho=0}. \tag{2.6}$$

Substituting Eqs. (2.4) and (2.5) into (2.3), we have

$$\frac{\partial U}{\partial a^k}(\Xi^\beta(s,\rho),a^l(\rho)) = \alpha_k(\rho).$$

We differentiate this with respect to ρ and evaluate the result for $\rho = 0$. After making use of (2.6), we obtain

$$\frac{\partial^2 U}{\partial a^k \partial x^\sigma}(\xi^\alpha,a^l) r^\sigma + \left.\frac{\partial^2 U}{\partial a^k \partial a^m}(\xi^\alpha,a^l)\frac{\partial a^m}{\partial \rho}\right|_{\rho=0} = \left.\frac{\partial \alpha_k}{\partial \rho}\right|_{\rho=0}, \quad k=1,2,3. \tag{2.7}$$

If we now compare Eqs. (2.7) with Eqs. (1.6), which are valid also along the basic geodesic, we obtain the relations

$$b^k = \left.\frac{\partial a^k}{\partial \rho}\right|_{\rho=0}, \quad \beta_l = \left.\frac{\partial \alpha_l}{\partial \rho}\right|_{\rho=0}. \tag{2.8}$$

Thus if a deviation vector is defined by (2.6), i.e., is a vector tangent to the curve determined by (2.5) for $s=\text{const}$, then in accordance with (2.8) the parameters b^k and β_l introduced in Eq. (1.6) can be interpreted as pieces of information that determine relative changes of the values of the parameters a^k and α_l between the basic geodesic and the neighboring ones.

This consideration can easily be generalized to an f-parameter family of geodesics $(0<f<6)$. If in the space of parameters (a^k, α_l) we have found an f-cube,

$$\mathbb{R}^f \ni (\rho_1,\dots,\rho_f) \to (a^k(\rho_1,\dots,\rho_f),\alpha_l(\rho_1,\dots,\rho_f)) \in \mathbb{R}^6,$$

then in analogy to (2.5) we obtain

$$x^\beta = \xi^\beta(s,a^k(\rho_1,\dots,\rho_f),\alpha_l(\rho_1,\dots,\rho_f)) =: \Xi^\beta(s,\rho_1,\dots,\rho_f). \tag{2.9}$$

The deviation vector is now defined as the linear combination

$$r^\sigma := \left.\frac{\partial \Xi^\sigma}{\partial \rho_a}\right|_{\rho_a=0} A^a, \tag{2.10}$$

in which A^a, $a=1,\dots,f$, are some arbitrary coefficients. In analogy to (2.8), it can easily be shown that the parameters b^k, β_l, which appeared in (1.6), can be interpreted as

$$b^k = \left.\frac{\partial a^k}{\partial \rho_a}\right|_{\rho_a=0} A^a, \quad \beta_l = \left.\frac{\partial \alpha_l}{\partial \rho_a}\right|_{\rho_a=0} A^a. \tag{2.11}$$

Now from the set of all geodesics we want to select a subfamily parametrized by $f=6-N$ parameters, $0<N<6$. One way of doing this is to impose N equations constraining the six parameters a^k and α_l

$$f_a(a^k,\alpha_l) = 0, \quad a=1,\dots,N. \tag{2.12}$$

Equations (2.12) determine an f-surface (or more rigorously speaking, a local part of it forming an f-cube) in the space of all parameters (a^k,α_l). This surface can in an equivalent way be described in the parameter form

$$a^k = a^k(\rho_1,\dots,\rho_f), \quad \alpha_l = \alpha_l(\rho_1,\dots,\rho_f), \tag{2.13}$$

where $\rho_1,\dots,\rho_f$ are parameters which take their values from certain intervals of $\mathbb{R}$. The functions $a^k(\cdot,\dots,\cdot)$ and $a_l(\cdot,\dots,\cdot)$ are such that

$$f_a(a^k(\rho_1,\dots,\rho_f),\alpha_l(\rho_1,\dots,\rho_f)) \equiv 0 \tag{2.14}$$

for all admissible values of the parameters.

The question is how one should take into account the existence of the constraints (2.12) in the general solution (1.8) of the geodesic deviation equations.

After differentiating Eqs. (2.14) with respect to ρ_a, $a=1,\dots,f$, evaluating the result for $\rho_a=0$, and making use of (2.11), we obtain

$$\frac{\partial f_a}{\partial a^k} b^k + \frac{\partial f_a}{\partial \alpha_l}\beta_l = 0, \quad a=1,\dots,N. \tag{2.15}$$

These relations are constraints on the parameters b^k and β_l that should be added to the constraints (2.12) on a^k and α_l. Thus if in the general solution (1.8) the superfluous degrees of freedom represented by a now too large number of all the parameters a^k, α_l, b^m, and β_n will be eliminated by means of the constraint relations (2.12) and (2.15), then one obtains a solution that represents deviations which point from the basic geodesic to geodesics belonging to a selected subfamily determined by the constraints (2.12).

III. GENERAL SOLUTION OF THE GEODESIC DEVIATION EQUATIONS IN THE SCHWARZSCHILD SPACE-TIME

The Schwarzschild space-time in the standard coordinates is described by the well-known metric

$$ds^2 = \left(1-\frac{r_g}{r}\right)c^2\,dt^2 - \left(1-\frac{r_g}{r}\right)^{-1} dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\phi^2, \tag{3.1}$$

where c is the velocity of light, $r_g = 2GMc^{-2}$ is the gravitational radius, M is the total mass of the source of the gravitational field, and G is the Newtonian gravitational constant. The coordinates t, r, θ, and ϕ take their values from the following intervals: $-\infty<t<\infty$, $0<\theta<\pi$, $0<\phi<2\pi$, and $r_g<r<\infty$ or $0<r<r_g$. For $r=r_g$, i.e., on the horizon, we have a singularity of the coordinates systems which we use. This singularity is unphysical and can be removed by using other coordinates, e.g., those found by Kruskal.[5] In the sequel we will consider both $0<r<r_g$ and $r_g<r<r_g$ coordinates patches simultaneously. In these coordinate systems, the Hamilton–Jacobi equation reads

$$\frac{1}{c^2}\left(r-\frac{r_g}{r}\right)^{-1}\left(\frac{\partial U}{\partial t}\right)^2 - \left(1-\frac{r_g}{r}\right)\left(\frac{\partial U}{\partial r}\right)^2 - \frac{1}{r^2}\left(\frac{\partial U}{\partial \theta}\right)^2 - \frac{1}{r^2\sin^2\theta}\left(\frac{\partial U}{\partial \phi}\right)^2 = mc^2, \tag{3.2}$$

where m is the mass of the test particle, and $U=U(t,r,\theta,\phi)$ is the Hamilton–Jacobi function for geodesics. The complete integral of Eq. (3.2) can be found in a standard way,[6] by separation of variables

$$U(t,r,\theta,\phi;\mathscr{E},\mathscr{K},\mathscr{J}) = -\mathscr{E}t - \int\left(\frac{B(r)}{A(r)}\right)^{1/2}dr + \int C(\theta)^{1/2}\,d\theta + \mathscr{J}\phi, \tag{3.3}$$

where $\mathscr{E}$, $\mathscr{K}$, and $\mathscr{J}$ are separation constants, and where the following notation has been introduced:

$$\begin{aligned} A(r) &:= 1 - r_g/r,\\ B(r) &:= \mathscr{E}^2/c^2A(r) - \mathscr{K}^2/r^2 - m^2c^2,\\ C(\theta) &:= \mathscr{K}^2 - \mathscr{J}^2/\sin^2\theta. \end{aligned} \tag{3.4}$$

Evidently, the relation $-\mathscr{K}\leqslant\mathscr{J}\leqslant\mathscr{K}$ is necessary for $C(\theta)\geqslant 0$. The constants, $\mathscr{E}$, $\mathscr{K}$, and $\mathscr{J}$ have a well-known physical interpretation (cf. Refs. 4 and 6). In the case of open geodesics (i.e., of those "starting" from the spaceline infinity), which are the only ones that will be considered here in some detail, the constant $\mathscr{E}$ is the total energy of the test particle at the spacelike infinity (therefore $\mathscr{E}\geqslant mc^2$), whereas $\mathscr{K}$ can be interpreted as the magnitude and $\mathscr{J}$ as the z-component of the angular momentum of the particle there.

The system of equations $\partial U/\partial\mathscr{E} = \alpha_1$, $\partial U/\partial\mathscr{K} = \alpha_2$, $\partial U/\partial\mathscr{J} = \alpha_3$ [cf. Ref. 1, Eqs. (4.2)] has the following solution:

$$\begin{aligned} t(r) &= -\alpha_1 - \frac{\mathscr{E}}{c^2}\int\frac{1}{A(r)^2}\left(\frac{A(r)}{B(r)}\right)^{1/2}dr,\\ \theta(r) &= \arccos\left[\frac{(\mathscr{K}^2-\mathscr{J}^2)^{1/2}}{\mathscr{K}}\cos D(r)\right],\\ \phi(r) &= \alpha_3 - \arcsin\left[\frac{\mathscr{J}\cos D(r)}{[\mathscr{K}^2-(\mathscr{K}^2-\mathscr{J}^2)\cos^2 D(r)]^{1/2}}\right], \end{aligned} \tag{3.5}$$

where

$$D(r) := \alpha_2 - \mathscr{K}\int r^{-2}[A(r)B(r)^{-1}]^{1/2}A(r)^{-1}\,dr.$$

Equations (3.5) describe geodesics in the Schwarzschild space-time as parametrized by the radial coordinate r. Since, in general, any parameter along a non-null curve must be a monotonic function of its proper time s, Eqs. (3.5) describe a curve only for $r_1 < r < r_2$, where r_1 and r_2 are some turning points (in particular r_2 could be taken equal to $+\infty$). Such an approach, based on the separation of spherical variables, must in particular exclude from the consideration the circular orbits, both in the relativistic and in the Newtonian dynamics.

Let us observe that the equation $\partial U/\partial\mathscr{J} = \alpha_3$ can easily be integrated and rewritten in the form

$$-(\mathscr{K}^2-\mathscr{J}^2)^{1/2}\sin\alpha_3\cos\phi\sin\theta + (\mathscr{K}^2-\mathscr{J}^2)^{1/2}\cos\alpha_3\sin\phi\sin\theta + \mathscr{J}\cos\theta = 0. \tag{3.6}$$

Obviously Eq. (3.6) describes a two-parameter family of planes through the origin of the spherical coordinate system (r,θ,ϕ) in the space $\mathbb{R}^3$.

Knowing the Hamilton–Jacobi function U for geodesics, determined here by (3.3), we can in accordance with Eq. (6.3) from Ref. 1 construct a complete integral S of the Hamilton–Jacobi equations for the geodesic deviation,

$$\begin{aligned} S &= \left(\frac{\partial U}{\partial x^\sigma}\right)\rho^\sigma + E\left(\frac{\partial U}{\partial\mathscr{E}}\right) + K\left(\frac{\partial U}{\partial\mathscr{K}}\right) + I\left(\frac{\partial U}{\partial\mathscr{J}}\right)\\ &= -\mathscr{E}\rho^t - \left(\frac{B(r)}{A(r)}\right)^{1/2}\rho^r + C(\theta)^{1/2}\rho^\theta + \mathscr{J}\rho^\phi - Et\\ &\quad - \frac{\mathscr{E}E}{c^2}\int\frac{1}{A(r)^2}\left(\frac{A(r)}{B(r)}\right)^{1/2}dr\\ &\quad + \mathscr{K}K\int\frac{1}{r^2A(r)}\left(\frac{A(r)}{B(r)}\right)^{1/2}dr\\ &\quad + \mathscr{K}K\int\frac{1}{C(\theta)^{1/2}}\,d\theta + I\phi\\ &\quad - I\mathscr{J}\int\frac{1}{[C(\theta)]^{1/2}\sin^2\theta}\,d\theta, \end{aligned} \tag{3.7}$$

where the parameters b^s ($s = 1,2,3$) which entered Eq. (6.3) in Ref. 1 are now denoted by $b^1 = E$, $b^2 = K$, and $b^3 = I$. Next we have to solve the system of algebraic equations (1.6) and (1.7) for the components ρ^σ of the deviation field. Introducing in this system the notation $\beta_1 = \alpha$, $\beta_2 = \beta$, $\beta_3 = \gamma$, and assuming that $\beta_0 = 0$, after some simple although lengthy calculations we derive the following general solution of the geodesic deviation equations along a geodesic described by Eqs. (3.5):

$$\rho^t(r) = -\alpha + m^2EX_1(r) - \frac{\mathscr{K}}{c^2}(\mathscr{E}K - \mathscr{K}E)X_2(r) + \frac{\mathscr{E}\Omega(r)}{m^2c^4A(r)}, \tag{3.8a}$$

$$\rho^r(r) = -\frac{\Omega(r)}{m^2c^2}(A(r)B(r))^{1/2}, \tag{3.8b}$$

$$\rho^\theta(r) = \frac{\mathscr{J}(\mathscr{K}I - \mathscr{J}K)\cos D(r)}{\mathscr{K}[(\mathscr{K}^2-\mathscr{J}^2)\Gamma(r)]^{1/2}} + \frac{(\mathscr{K}^2-\mathscr{J}^2)^{1/2}\sin D(r)}{\Gamma(r)^{1/2}}\left(\Sigma(r) + \frac{\mathscr{K}\Omega(r)}{m^2c^2r^2}\right), \tag{3.8c}$$

$$\rho^\phi(r) = \gamma + \frac{(\mathscr{J}K - \mathscr{K}I)\sin D(r)\cos D(r)}{\Gamma(r)} + \frac{\mathscr{K}\mathscr{J}}{\Gamma(r)}\left(\Sigma(r) + \frac{\mathscr{K}\Omega(r)}{m^2c^2r^2}\right), \tag{3.8d}$$

where

$$\Omega(r) := (\mathscr{E}_\alpha + \mathscr{K}\beta + \mathscr{J}\gamma) - m^2c^2\left(\frac{\mathscr{E}EX_1(r)}{c^2} - \mathscr{K}KX_3(r)\right), \tag{3.8e}$$

$$\Sigma(r) := \beta - \frac{\mathscr{E}}{c^2}(\mathscr{E}K - \mathscr{K}E)X_2(r) + m^2c^2KX_3(r), \tag{3.8f}$$

$$\Gamma(r) := \mathscr{K}^2 - (\mathscr{K}^2 - \mathscr{J}^2)\cos^2 D(r), \tag{3.8g}$$

$$X_1(r) := \int\frac{1}{A(r)^3}\left(\frac{A(r)}{B(r)}\right)^{3/2}dr,$$

$$X_2(r) := \int\frac{1}{r^2A(r)^3}\left(\frac{A(r)}{B(r)}\right)^{3/2}dr,$$

$$X_3(r) := \int \frac{1}{r^2 A(r)^2}\left(\frac{A(r)}{B(r)}\right)^{3/2} dr,$$

and where $D(r)$ is given by Eq. (3.5).

Equations (3.5) and (3.8) determine the solution of our problem in the two coordinate patches for $0<r<r_g$ and $r_g<r<+\infty$.

It can be easily shown that in the limit at the horizon $r=r_g$, only the components ρ^r, ρ^θ, and ρ^ϕ, and not ρ^t, are nonsingular. This last property is a consequence of the fact that an arbitrary geodesic which describes a free fall into the singularity reaches the horizon after an infinite coordinate time t. The singularity of ρ^t is however weak enough to permit the square of the length of the deviation vector, i.e., the invariant $\rho^2 = g_{\alpha\beta}\rho^\alpha\rho^\beta$, to be nonsingular at the horizon. This is just another example of the pathological behavior of the classical Schwarzschild coordinates.

The question which should yet be clarified is that of the correspondence between the two solutions obtained in the two coordinate patches, $0<r<r_g$ and $r_g<r<+\infty$, used here. Some geodesics in the exterior region, namely those which have their endpoints at the horizon, will have their prolongation inside the horizon, as it can be seen by considering the motion, for example, in the Kruskal coordinates. If one solves the geodesic and geodesic deviation equations at first in the Kruskal coordinates, and rewrites this solution later in terms of the Schwarzschild coordinates both outside and inside the horizon, then one will obtain exactly the same formulas as those given by (3.5) and (3.8). Thus it is the identity of all the values of the parameters $(a^k, \alpha_1, b^m, \beta_n)$ in the two coordinate patches that ensures the correspondence between a solution determined in the outside region with that being its prolongation into the inside of the horizon.

We conclude this section by quoting the solution of the geodesic deviation equations in the Kruskal coordinates (v,u), which can be introduced by means of the transformation (cf. Ref. 5),

$$\left.\begin{aligned} v &= \left(\frac{r}{r_g}-1\right)^{1/2}\exp\left(\frac{r}{2r_g}\right)\sinh\left(\frac{ct}{2r_g}\right),\\ u &= \left(\frac{r}{r_g}-1\right)^{1/2}\exp\left(\frac{r}{2r_g}\right)\cosh\left(\frac{ct}{2r_g}\right),\end{aligned}\right\}\quad \text{for } r>r_g, \tag{3.9a}$$

and

$$\left.\begin{aligned} v &= \left(1-\frac{r}{r_g}\right)^{1/2}\exp\left(\frac{r}{2r_g}\right)\cosh\left(\frac{ct}{2r_g}\right),\\ u &= \left(1-\frac{r}{r_g}\right)^{1/2}\exp\left(\frac{r}{2r_g}\right)\sinh\left(\frac{ct}{2r_g}\right),\end{aligned}\right\}\quad \text{for } r<r_g. \tag{3.9b}$$

The inverse transformation is implicitly determined by the equations

$$\begin{aligned} u^2 - v^2 &= \left(\frac{r}{r_g}-1\right)\exp\left(\frac{r}{r_g}\right),\\ t &= -\frac{r_g}{c}\ln\left|\frac{u-v}{u+v}\right|. \end{aligned} \tag{3.10}$$

In these coordinates, the components ρ^v and ρ^u of the deviation field are

$$\rho^v = \frac{c}{2r_g}Q(r)u(r) + \frac{\Omega(r)}{2r_g m^2c^2A(r)}\left[\frac{\mathscr{E}u(r)}{c} - (A(r)B(r))^{1/2}v(r)\right], \tag{3.11a}$$

$$\rho^u(r) = -\frac{c}{2r_g}Q(r)v(r) + \frac{\Omega(r)}{2r_g m^2c^2A(r)}\left[\frac{\mathscr{E}v(r)}{c} - (A(r)B(r))^{1/2}u(r)\right], \tag{3.11b}$$

where

$$Q(r) := \alpha - m^2EX_1(r) + \frac{\mathscr{K}}{c^2}(\mathscr{E}K - \mathscr{K}E)X_2(r), \tag{3.11c}$$

and $v(r)$ and $u(r)$ are coordinates of the point running along the basic geodesic. The angular components are of course the same as in Eqs. (3.5) and (3.8), respectively. Evidently, the two components ρ^v and ρ^u are regular at the horizon, which is described here by the equations $v=\pm u$.

IV. DEVIATIONS BETWEEN GEODESICS FROM A SUBFAMILY IN THE SCHWARZSCHILD SPACE-TIME

In this section, the general procedure discussed in Sec. II of restricting the general solution of geodesic deviation equations to a solution describing deviations between geodesics belonging only to a subfamily of a set of all geodesics will be applied to two specific cases. In the first case we restrict ourselves to geodesics lying in the equatorial plane $\theta = \pi/2 = \text{const}$. This is motivated by the fact that due to the symmetry of the Schwarzschild geometry, it is sufficient to deal with geodesics from this plane only. In the second case, we will consider radial geodesics, i.e., for which θ and ϕ are constants. This second case will be used in the discussion which is carried on in Sec. V.

Let us take a plane, arbitrarily chosen from the family (3.6), and kept fixed during the whole consideration afterwards. The family (3.6) of planes through the origin depends obviously on two parameters. In the case of $\mathscr{J} \neq 0$ (which is in particular true for the equatorial plane) one can take for such parameters the quantities

$$C_1 = \frac{(\mathscr{K}^2 - \mathscr{J}^2)^{1/2}\sin\alpha_3}{\mathscr{J}},\qquad C_2 = \frac{(\mathscr{K}^2 - \mathscr{J}^2)^{1/2}\cos\alpha_3}{\mathscr{J}}. \tag{4.1}$$

Selecting a plane from the family (3.6) amounts then to setting C_1 and C_2 to be equal to some constants. After the plane is fixed, it follows from (4.1) that there is some freedom of choice of the parameters $\mathscr{K}$ and $\mathscr{J}$ in Eqs. (3.5) and (3.8), though restricted by the constraint condition

$$\mathscr{K} - (1 + C_1^2 + C_2^2)^{1/2}\mathscr{J} = 0 \tag{4.2}$$

(without any essential loss of generality, it has been assumed here that both $\mathscr{K}$ and $\mathscr{J}$ are non-negative), whereas there is no freedom of choice at all of the quantity α_3, since due to (4.1),

$$\alpha_3 = \arctan\left(\frac{C_1}{C_2}\right). \tag{4.3}$$

Relations (4.2) and (4.3) are examples of the constraint conditions on the parameters a^k and α_1 of the form given by Eqs. (2.12). Now we intend to derive the constraints which are consequences of Eqs. (4.2) and (4.3) and restrict the parameters of the type of b^k, β_l. The constraint (4.2) on $\mathscr{K}$ and $\mathscr{J}$ implies the constraint corresponding to (2.15), which in the present case assumes the form

$$K - (1 + C_1{}^2 + C_2{}^2)^{1/2} I = 0. \tag{4.4}$$

Similarly from (4.3) it follows that

$$\gamma = 0. \tag{4.5}$$

Now from (4.2) and (4.4) it immediately follows that in the present case, the constants $\mathscr{K}$, $\mathscr{J}$, K, and I must necessarily satisfy the relation

$$\mathscr{K} I - \mathscr{J} K = 0. \tag{4.6}$$

We are now prepared to reduce the general geodesic deviation field (3.8) to deviations between the geodesics lying in a fixed plane from the family (3.6). First we make use in Eqs. (3.8) of (4.5) and (4.6), and then eliminate from there the parameters $\mathscr{J}$ and I, expressing them by $\mathscr{K}$ and K in accordance with Eqs. (4.2) and (4.4). As a result, the deviation field will depend on the parameters C_1 and C_2 characterizing the choice of the plane. In the special case of the equatorial plane, one has $C_1 = C_2 = 0$ (i.e., $\mathscr{K} = \mathscr{J}$ and $K = I$), and the geodesic deviation field describing deviations between the geodesics lying only in that plane is of the form

$$\rho^t(r) = -Q(r) + \frac{\mathscr{E}\Omega(r)}{m^2c^4A(r)}, \tag{4.7a}$$

$$\rho^r(r) = -\frac{[A(r)B(r)]^{1/2}\Omega(r)}{m^2c^2}, \tag{4.7b}$$

$$\rho^\theta(r) = 0, \tag{4.7c}$$

$$\rho^\phi(r) = \sum(r) + \frac{\mathscr{K}\Omega(r)}{m^2c^2r^2}, \tag{4.7d}$$

where $\Omega(r)$, $\Sigma(r)$, and $Q(r)$ are given by Eqs. (3.8e), (3.8f), and (3.11c), respectively.

For an arbitrary spherical symmetric and static space-time, the geodesic deviation equations were for the first time solved by Fuchs.[3] The solution was found as a result of a direct integration of these equations with the aid of the first integrals which existed because of the assumed symmetries. In Ref. 3, it is asserted that one takes into account only geodesics lying in the equatorial plane. It is, however, not stated explicitly that it is only the basic geodesic that lies in the equatorial plane, whereas the paper evidently also admits deviations pointing to geodesics which are lying in planes different from the equatorial one. The expressions derived in Ref. 3 for the components ρ^t, ρ^r, and ρ^ϕ of the deivation field are identical, although in a different notation, with the expressions (4.7a), (4.7b), and (4.7d) obtained here. For the component ρ^θ, however, one obtains in Ref. 3 the expression

$$\rho^\theta = -C_5 \sin\phi + C_6 \cos\phi, \tag{4.8}$$

where C_5 and C_6 are constants, while in the case of deviations between geodesics lying in the equatorial plane only, in accordance with (4.7c), one should obtain $\rho^\theta = 0$. Elementary geometric consideration enables one to find the interpretation of the parameters C_5 and C_6 in (4.8). If the geodesic to which the deviation field is pointing lies in a plane forming with the equatorial plane an angle ψ, $\psi \ll 1$, and if ϕ_0 $(0 \leqslant \phi_0 \leqslant \pi)$ is the azimuthal angle of the straight line being the line of intersection of the two planes, then

$$C_5 = \psi\cos\phi_0, \quad C_6 = \psi\sin\phi_0. \tag{4.9}$$

One should note that in Ref. 3 the geodesic deviation equations were solved for the Schwarzschild metric restricted to the equatorial plane, i.e., for a three-dimensional submanifold. It can be checked by a direct calculation that if one starts from the Hamilton–Jacobi equation (1.1) written in the three-dimensional metric, then one obtains for the components ρ^t, ρ^r, and ρ^ϕ of the deviation field expressions that are identical with those given here by Eqs. (4.7). The difference between the approach presented here and that employed in Ref. 3 relies basically upon two various ways of restricting the solution to a submanifold. In our case it is the solution found in the full, four-dimensional space-time that is restricted to a three-dimensional submanifold, and in Ref. 3 the solution of the problem restricted to a submanifold is found. The fact that the two approaches have led in the case of the equatorial plane in the Schwarzschild space-time to the same result is to a certain extent obvious, because this plane is a totally geodesic submanifold, i.e., all the geodesics of the internal geometry on this submanifold are also geodesics of the full four-dimensional metric. However, a procedure analogous to that of Ref. 3 in the case of an arbitrary hypersurface of a general space-time would obviously be incorrect.

Let us consider now the second case of restricting the general solution (3.8) to radial geodesics. To derive an expression for deviations in this particular case, one must pass with the general solution (3.8) to the limit $\mathscr{J} = 0$ and $\mathscr{K} = 0$. Since the solution is determined only for $-\mathscr{K} \leqslant \mathscr{J} \leqslant \mathscr{K}$, the order of taking the limits is not arbitrary and one should take into account first the condition $\mathscr{J} = 0$ and its implication $I = 0$, and then the conditions $\mathscr{K} = 0$ and $K = 0$ [cf. Eqs. (2.12) and (2.15)]. We obtain

$$\rho^t(r) = \frac{rF + H}{m^2c^2(r - r_g)}[\alpha - m^2EX_1(r)], \tag{4.10a}$$

$$\rho^r(r) = -\frac{(rF + H)^{1/2}}{m^2c^2r^{1/2}}\{\alpha - m^2EX_1(r)\}, \tag{4.10b}$$

$$\rho^\theta(r) = \beta, \tag{4.10c}$$

$$\rho^\phi(r) = \gamma, \tag{4.10d}$$

where

$$F := \frac{\mathscr{E}^2}{c^2} - m^2c^2, \quad H := m^2c^2r_g$$

and

$$X_1(r)=\begin{cases}\dfrac{1}{F^2}\dfrac{r^{1/2}(rF+3H)}{(rF+H)^{1/2}}+\dfrac{3H}{2F^{5/2}}\ln\left|\dfrac{(rF+H)^{1/2}-(rF)^{1/2}}{(rF+H)^{1/2}+(rF)^{1/2}}\right|, & F>0 \quad (\mathscr{E}>mc^2),\\ \dfrac{2r^{5/2}}{5H^{3/2}}, & F=0 \quad (\mathscr{E}=mc^2).\end{cases}$$

V. THE FATE OF A MAN WHO FALLS INTO THE SINGULARITY AT $r=0$

The title of this section is that of §32.6, pp. 860–862, in the textbook by Misner, Thorne, and Wheeler,[4] in which one discusses the fate of an astrophysicist who stands on the surface of a freely collapsing star. As a result of an analysis of only the form of the deviation equations written in the astrophysicist's local inertial frame, one arrives in Ref. 4 at the conclusion that on his body act stresses which stretch him in the longitudinal direction and compress him in the transversal direction. As the distance to the singularity decreases, the stresses increase. Since the compressing stresses exceed the stretching ones, the astrophysicist's body must finally decay into a cloud of freely falling baryons, and the volume of this cloud tends to zero for $r\to 0$. As a conclusion from the analysis of the asymptotic behavior of radial geodesics for $r\to 0$ carried on in Ref. 4, one also finds a more precise result stating that the volume decreases to zero as $r^{3/2}$.

Now, the rigorous solution of the geodesic deviation equations restricted to deviations between radial geodesics only, given by Eqs. (4.10), enables us to propose another method of discussing the astrophysicist's fate. Let us suppose that after its decay, the astrophysicist's body forms a cloud of pointlike, freely and radially falling particles—baryons. Since before the decay, the baryons constituting the astrophysicist's body were at relative rest, we can assume that the cloud of baryons was "inserted" into the space-time at spatial infinity, where the relative velocities of baryons vanish,

$$\lim_{r\to\infty}\frac{D\rho^\sigma}{ds}=0,\quad \sigma=0,1,2,3.$$

This equality holds if and only if

$$E=0. \tag{5.1}$$

In order to compute the derivative $D\rho^\sigma/ds$, we make use of the relation

$$\frac{D\rho^\alpha}{ds}=\pm g^{\alpha\beta}\left(\frac{\partial S}{\partial x^\alpha}-\Gamma^\sigma{}_{\beta\gamma}\rho\gamma\frac{\partial S}{r^\sigma}\right),$$

where S is the function (3.7), cf. Ref. 1. Substituting (5.1) into (4.10) leads to

$$\begin{aligned}\rho^t(r)&=\frac{\alpha}{m^2c^2A(r)}\left(\frac{\mathscr{E}^2}{c^2}-m^2c^2A(r)\right),\\ \rho^r(r)&=-\frac{\alpha}{m^2c^2}\left(\frac{\mathscr{E}^2}{c^2}-m^2c^2A(r)\right)^{1/2},\\ \rho^\theta(r)&=\beta,\\ \rho^\phi(r)&=\gamma.\end{aligned} \tag{5.2}$$

The volume V of the astrophysicist's body is defined as

$$V(r):=|\epsilon_{\alpha\beta\gamma\delta}(-g)^{1/2}u^\alpha\rho_1^\beta\rho_2^\gamma\rho_3^\delta|, \tag{5.3}$$

where $\epsilon_{\alpha\beta\gamma\delta}$ is the completely antisymmetric Levi-Civita symbol of the fourth order ($\epsilon_{0123}=-1$), $g=\det(g_{\alpha\beta})$, u^α are components of the four-velocity vector along the basic geodesic, and ρ_1^β, ρ_2^γ, and ρ_3^δ are components of three linearly independent solutions of the geodesic deviation equations taken along the basic geodesic. The basic geodesic is understood as the world line of the baryon at the geometric center of the cloud, and the four-vectors ρ_k, $k=1,2,3$, are deviations pointing from the basic geodesic to geodesics being the world lines of the baryons at the boundary of the cloud.

The following argument can be used to support the view that the quantity V defined above really has the physical interpretation of the volume of the baryonic cloud. To demonstrate this, let us endow the basic baryon with an orthonormal tetrad constructed of the four-velocity vector $u=\lambda_{(0)}$ and of three mutually orthonormal spacelike vectors $\lambda_{(a)}$, $(a)=1,2,3$ (indices in brackets are tetrad indices) which are orthogonal to the four-velocity vector. Thus the tetrad vectors satisy the orthonormality conditions

$$\lambda^\mu_{(\alpha)}g_{\mu\sigma}\lambda^\sigma_{(\beta)}=\eta_{(\alpha)(\beta)},$$

where $\eta_{(\alpha)(\beta)}$ is the Minkowski matrix. The vectors $\lambda_{(a)}$ can be interpreted as forming a Cartesian coordinate system which is used by the basic observer.[7] They span the three-space of instantaneous rest of the observer. The physical components of the deviation vectors ρ_k in this tetrad, which are defined to be the scalars

$$\rho_k(\alpha):=\lambda^\sigma_{(\alpha)}\rho_{k\sigma},\quad k=1,2,3,\quad (\alpha)=0,1,2,3$$

(obviously $\rho_{k(0)}=0$ for $k=1,2,3$), can be considered to be Cartesian components in the three-space of the basic observer of vectors which determine the size of the baryonic cloud. Since ρ_k are three linearly independent four-vectors, the three-vectors with the components $(\rho_{k(1)},\rho_{k(2)},\rho_{k(3)})$, $k=1,2,3$, will also be linearly independent. The volume $\widetilde{V}$ measured by the basic observer is equal to the volume of the parallelepiped spanned by these vectors,

$$\widetilde{V}:=|\epsilon^{(b)(c)(d)}\rho_{1(b)}\rho_{2(c)}\rho_{3(d)}|,$$

where $\epsilon^{(b)(c)(d)}$ is the Levi-Civita symbol of the third order ($\epsilon^{(1)(2)(3)}=1$). One can easily show that $\widetilde{V}=V$.

In accordance with the solution (5.2), the three linearly independent and mutually orthogonal solutions of the deviation equations can be chosen in the form

$$\begin{aligned}\rho_1^\mu&=\frac{\alpha}{m^2c^2}\Bigg(A(r)^{-1}\left(\frac{\mathscr{E}^2}{c^2}-m^2c^2A(r)\right),\\ &\qquad -\mathscr{E}\left(\frac{\mathscr{E}^2}{c^2}-m^2c^2A(r)\right)^{1/2},0,0\Bigg),\\ \rho_2^\mu&=(0,0,\beta,0),\\ \rho_3^\mu&=(0,0,0,\gamma).\end{aligned} \tag{5.4}$$

The components of the four-velocity vector along the basic radial geodesic ($\mathscr{K} = \mathscr{J} = 0$) are equal to

$$u^\mu = \frac{1}{mc}\left(\frac{\mathscr{E}}{c^2 A(r)}, -\left(\frac{\mathscr{E}^2}{c^2} - m^2c^2A(r)\right)^{1/2}, 0, 0\right). \tag{5.5}$$

Substituting Eqs. (5.4) and (5.5) into (5.3), we obtain ($g = -c^2r^4 \sin^2\theta$):

$$V(r) = \frac{\alpha\beta\gamma}{m} r^2 \sin\theta\left(\frac{\mathscr{E}^2}{c^2} - m^2c^2A(r)\right)^{1/2}, \tag{5.6}$$

thus in the limit for $r \to 0$,

$$V(r) \sim c\alpha\beta\gamma \sin\theta (r_g)^{1/2} r^{(3/2)}, \quad \text{for } r \to 0.$$

This equation reveals the same type of the asymptotic r dependence for $r \to 0$ as in Ref. 4. There, however, the result followed from a rather qualitative discussion of the behavior of geodesics in the vicinity of $r = 0$, mainly from the observation that in the Schwarzschild coordinates, the coordinate lines in the neighborhood of $r = 0$ are nearly geodesics, whereas our treatment is manifestly geometric. It is based on a covariant definition of the three-volume of the cloud of falling particles, as well as on the exact solution (5.2) of the deviation equations describing the relative motion of these particles. Besides, the final formula (5.6) contains more information than a mere statement about the asymptotic behavior of the three-volume of the cloud, and could, for instance, be used for making a comparison with analogous situations in other theories.

VI. COMPARISON WITH THE NEWTONIAN CASE

It is not too difficult to analyze a similar problem in the Newtonian theory of gravitation. Let us therefore consider a cloud of pointlike test particles which are falling freely and radially (in accordance with the conditions $\mathscr{K} = \mathscr{J} = 0$) into the center of a Newtonian spherically symmetric gravitational field of the potential $-GM/r$ (G is here the gravitational constant, and M is the mass of the pointlike source of the field). It is assumed that all particles in the cloud have the same total energy $\mathscr{E}_0$ (which corresponds to the condition $E = 0$ assumed in the relativistic case). Due to this assumption, all particles are falling into the center within a solid angle of measure $\Delta\Omega$ and, moreover, if at the instant of time $t = t_0$ the cloud is filling a section of a spherical shell of thickness $h(r_0)$ with the radius $r(t_0) = r_0$, then the three-volume of the cloud at the instant of time when its geometrical center is at the distance r to the center of the field will be given by the formula

$$V(r) = \Delta\Omega h(r) r^2, \tag{6.1}$$

where $h(r)$ is the thickness of the section of the spherical shell occupied by the cloud at that instant of time.

Making use of the mechanical energy conservation law in a gravitational field,

$$\frac{1}{2} m\left(\frac{dr}{dt}\right)^2 - \frac{GMm}{r} = \mathscr{E}_0,$$

where m is the mass of a single test particle, we obtain the following relation between the radial coordinate r of a falling particle and the time t,

$$f(r) - f(r_0) = -\mathscr{E}_0 (2/m)^{1/2}(t - t_0), \tag{6.2}$$

where

$$f(r) = (\mathscr{E}_0 r^2 + GMmr)^{1/2} - \frac{GMm}{\mathscr{E}_0^{1/2}} \operatorname{Arsinh}\left[\left(\frac{\mathscr{E}_0 r}{GMm}\right)^{1/2}\right],$$

and where the initial condition $r(t_0) = r_0$ was assumed. Let us consider now the free fall of two particles with the same energy $\mathscr{E}_0$. At the time $t = t_0$ we have $r_1(t_0) = r_0$, $r_2(t_0) = r_0 + \delta r_0$, $\delta r_0 \ll r_0$, where r_1 and r_2 are distances of the particles 1 and 2, respectively, to the center of the field. With the help of Eq. (6.2) we can write

$$f(r_2) - f(r_0 + \delta r_0) = f(r_1) - f(r_0), \tag{6.3}$$

which leads to the observation that

$$r_2(t) = r_1(t) + h(t), \tag{6.4}$$

where $h(t)$ is a small quantity of the same order as δr_0. After substituting (6.4) into (6.3) and expanding the functions $f(r_1 + h)$ and $f(r_0 + \delta r_0)$ with respect to small quantities h and δr_0, with the accuracy up to linear terms we obtain

$$h(r) = \left(\frac{df}{dr}(r_0)\right)\left(\frac{df}{dr}(r)\right)^{-1} \delta r_0 = \frac{(\mathscr{E}_0 r^2 + GMmr)^{1/2} r_0 \delta r_0}{(\mathscr{E}_0 r_0^2 + GMmr_0)^{1/2} r}, \tag{6.5}$$

which after substituting into (6.1) gives

$$V(r) = \Delta\Omega r^2 \frac{(\mathscr{E}_0 + GMm/r)^{1/2}}{(\mathscr{E}_0 + GMm/r_0)^{1/2}} \delta r_0. \tag{6.6}$$

A comparison of Eq. (6.6) with its relativistic counterpart given by Eq. (5.6) shows that each of the two relations is described by the same function, up to some yet to be determined constant coefficients, of the radial coordinate r. This will enable us, after a comparison of corresponding coefficients is made, to obtain an interpretation of the parameters α, β, and γ occurring in the relativistic formula (5.6). Before we start making this comparison, let us observe that in the Newtonian limit,

$$\mathscr{E}^2/c^2 - m^2c^2 = 2m\mathscr{E}_0, \tag{6.7}$$

where $\mathscr{E}$ is the total energy of the relativistic test particle at spatial infinity where the special theory of relativity is valid. Therefore $\mathscr{E}^2/c^2 - m^2c^2 = p_\infty{}^2$, where p_∞ is the momentum of the particle at infinity. The Newtonian limit corresponds to the case in which the velocity v_∞ of the particle at infinity is much smaller than the velocity of light c: $v_\infty \ll c$. Thus $p_\infty{}^2 = m^2 v_\infty{}^2$. On the other hand, $\mathscr{E}_0$ is the kinetic energy of the Newtonian particle at infinity, which implies that $2m\mathscr{E}_0 = m^2 v_\infty{}^2$. The comparison of the coefficients in the formulas (5.6) and (6.6), after Eq. (6.7) is taken into account, leads to the relations

$$\alpha = \frac{m\delta r_0}{[2m(\mathscr{E}_0 + GMm/r_0)]^{1/2}}, \tag{6.8a}$$

$$\beta\gamma \sin\theta = \Delta\Omega. \tag{6.8b}$$

The right-hand side of Eq. (6.8a) can be written in the form $\delta r_0/v_0$, where v_0 is the velocity of the particle situated at the center of the cloud, i.e., for $r = r_0$, and this means that α is a characteristic time related to the initial relative radial distance between the falling particles. On the other hand, from Eq. (6.8b) it follows that the constants β and γ determine the angular, transversal distance between the particles.

As we have shown, a cloud of particles freely falling into the singularity at $r = 0$ in the Schwarzschild space-time is crushed by the gravitational field, as is described by Eq. (5.6), in exactly the same way as a similar clould of particles falling into the center of a spherically symmetric Newtonian gravitational potential, being crushed in accordance with Eq. (6.6).

At first sight this result might be regarded as being rather a surprise. Intuitively we might expect that the rate at which the volume of a cloud of freely falling particles is compressed might be taken for a measure of strength of the singularity into which the particles are falling. If it is so, the result obtained here will mean that the two singularities, that of the Schwarzschild space-time and that of the Newtonian potential, are of exactly the same strength. Such a conclusion seems, however, to be in disagreement with the intuition developed as a result of all the work on relativistic gravitational collapse in the past, which prompts us to think that due to relativistic effects, the Schwarzschild singularity is considerably stronger than the Newtonian one. Thus we are left with a choice between two possibilities. Either we have to accept the result that the two singularities have the same strength, or we must look for some more subtle measures which would help distinguish between them. The second possibility remains, in our opinion, still open. The concept of geodesic deviation is an approximate one, for it describes only the first neighborhood of a basic geodesic.[2] It is thus possible that an approach based on this concept suppresses some genuine relativistic effects and a better approach based on a better approximation is needed to show them up. It is likely that such a better approximation could be, for instance, obtained by employing the concept of the geodesic deviation of the second or even higher order which was introduced in Ref. 2. To determine, however, which one of the two possibilities should be taken more seriously, requires further study.

VII. FREQUENCY SHIFT OF LIGHT SIGNALS EXCHANGED BETWEEN TWO FREELY FALLING OBSERVERS

We shall now study the effect of the frequency shift of light which two neighboring geodesic observers in a Schwarzschild field can send to one another, and we shall compare it with a Doppler shift observed by similar observers freely falling in the corresponding Newtonian, spherically symmetric gravitational field of a point source. In other words, we shall compare the results of the Doppler tracking between two freely falling, infinitesimally close observers in the two theories.

Let us first consider two infinitesimally close observers, Γ and $\tilde{\Gamma}$, freely falling in a relativistic gravitational field. We assume that along the geodesic Γ of a basic observer we are given a deviation field ρ^μ satisfying the condition $g_{\alpha\beta}u^\alpha\rho^\beta = 0$, where u^α is the four-velocity. Let $\gamma(s)$ be a

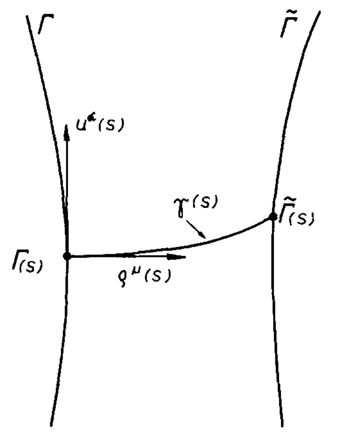


FIG. 1. To find the point $\tilde{\Gamma}(s)$ determined on the neighboring geodesic $\tilde{\Gamma}$ by a deviation vector $\rho^\mu(s)$, which is given at a point $\Gamma(s)$ on the basic geodesic Γ, one must construct at $\Gamma(s)$ a geodesic $\gamma(s)$ tangent to $\rho^\mu(s)$. The point $\tilde{\Gamma}(s)$ is situated at the intersection of $\gamma(s)$ and $\tilde{\Gamma}$, and the length of the arc $\Gamma(s)\tilde{\Gamma}(s)$, measured along $\gamma(s)$, is the spatial distance between the two observers, whereas Eq. (7.1) gives the linear approximation of the distance.

spacelike geodesic sent from the point $\Gamma(s)$ on the basic geodesic Γ (see Fig. 1) in the direction of $\rho^\mu(s)$ (s is here the proper time measured along Γ). As was shown in Ref. 2, if the adjacent geodesic $\tilde{\Gamma}$ intersects $\gamma(s_0)$ at a point $\tilde{\Gamma}(s_0)$ corresponding to the value κ of an affine parameter along γ, then for every s, $\gamma(s)$ will "nearly" intersect $\tilde{\Gamma}$, for the same value κ of the affine parameter missing it by terms of the order κ^2. Up to terms of the order κ^2, the expression

$$l_{\rm rel}(s) = \kappa(-g_{\alpha\beta}\rho^\alpha\rho^\beta)^{1/2} \tag{7.1}$$

can be considered to be the spatial distance between the basic and the adjacent geodesics Γ and $\tilde{\Gamma}$. In Ref. 2 it was shown that the frequency shift of light signals exchanged between the observers Γ and $\tilde{\Gamma}$ is described by the formula

$$z_{\rm rel}(s) = \frac{d}{ds}(l_{\rm rel}(s)). \tag{7.2}$$

It turns out (cf. Ref. 2) that up to terms of the order κ^2, it is irrelevant which of the observers emits and which receives the light signals.

Similarly one can show that in an analogous Newtonian case the relative frequency shift of light is given by the formula

$$z_{\rm newt}(t) = \frac{1}{c}\frac{d}{dt}(l_{\rm newt}(t)), \tag{7.3}$$

where $l_{\rm newt}(t)$ is an instantaneous spatial distance between two falling, infinitestimally close particles, and c is the velocity of light.

Now we apply Eq. (7.2) to the free fall in the Schwarzschild field. The components of the deviation field describing the relative motion of observers who fall freely into the center of the field and whose angular momentum as well as the relative energy at spatial infinity vanish are given by Eqs. (5.2). Substituting (5.2) into (7.1) leads to

$$l_{\rm rel}(s) = \kappa\left\{\frac{\alpha^2}{m^2}\left[\frac{\mathscr{E}^2}{c^2} m^2c^2\left(\frac{r_g}{r}-1\right)\right] + r^2(\beta^2+\gamma^2\sin^2\theta)\right\}^{1/2},$$

and this expression, after the relationship

$$\frac{dr}{ds} = u^r = -\frac{1}{mc}\left[\frac{\mathscr{E}^2}{c^2} + m^2c^2\left(\frac{r_g}{r}-1\right)\right]^{1/2}$$

[see (5.5)] is taken into account, we substitute in turn into Eq. (7.2). As a result, we obtain

$$z_{\rm rel}(s) = \frac{-\kappa}{2mc}\frac{[\mathscr{E}^2/c^2 + m^2c^2(r_g/r-1)]^{1/2}[2(\beta^2+\gamma^2\sin^2\theta)r - \alpha^2c^2r_g/r^2]}{\{(\alpha^2/m^2)[\mathscr{E}^2/c^2+m^2c^2(r_g/r-1)] + (\beta^2+\gamma^2\sin^2\theta)R^2\}^{1/2}}. \tag{7.4}$$

It can be easily seen that for $r > r_1$, we have $z<0$ (i.e., a blueshift), and for $r<r_1$, we obtain $z>0$ (i.e., a redshift), where r_1 is the solution of the equation $z(r) = 0$, that is

$$r_1 = \left(\frac{\frac{1}{2}\alpha^2c^2r_g}{\beta^2+\gamma^2\sin^2\theta}\right)^{1/3}.$$

In the limit for $r\to 0$, $z(r)$ tends to infinity, and its asymptotic behavior depends on the values of the integration constants,

$$z_{\rm rel}(r) \sim \kappa\alpha c r_g/r^2 \quad \text{for } r\to 0 \text{ and } \alpha\neq 0,$$

$$z_{\rm rel}(r) \sim -\kappa m c r_g(\beta^2+\gamma^2\sin^2\theta)^{1/2}/r$$

$$\begin{cases}\text{for } r\to 0 \text{ and } \alpha = 0,\\ \text{while } \beta \text{ or } \gamma\neq 0.\end{cases}$$

From Eq. (7.4) it follows immediately that in the two extremal cases corresponding to $\alpha = 0$ and $\beta=\gamma=0$, respectively, one obtains always $z(r)<0$ in the first and always $z(r)>0$ in the second case.

Similar calculations were performed in the Newtonian case by making use of Eqs. (6.4) and (6.5). After taking into account the relation (6.8a) and a modification of the relation (6.8b), one finds that the Newtonian expression for the quantity (7.3) is identical with the relativistic one given by Eq. (7.4).

Thus the frequency shift of the light signals exchanged between two neighboring freely falling observers both in the Schwarzschild and in the Newtonian gravitational field of a point source is described by the same function of the radial coordinate r, provided that one takes into account only terms which are of the first order with respect to small quantities determining the size of the cloud of such falling particles. This supports the result obtained in Sec. V that the relative motion of freely falling particles, when described in terms of the first geodesic deviation only, looks very much alike both in the Schwarzschild and in the corresponding Newtonian space-time.

ACKNOWLEDGMENT

This work has been partly supported by the Polish Research Program CPBP 01.03.

[1] S. L. Bażański, J. Math. Phys. **30**, 1018 (1989); S. L. Bażański, in *Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1986), p. 1615.
[2] S. L. Bażański, Ann. Inst. H. Poincaré A **27**, 115 (1977).
[3] H. Fuchs, Ann. Physik (Leipzig) **40**, 231 (1983).
[4] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), §32.6, pp. 860–862.
[5] M. D. Kruskal, Phys. Rev. **110**, 1743 (1960).
[6] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, translated by M. Hamermesh, 3rd English ed. (Pergamon, Oxford, 1971).
[7] F. A. E. Pirani, Acta Phys. Pol. **15**, 389 (1956).

Irreducible tensor description. I. A classical gas

Zbigniew Banach
Institute of Fundamental Technological Research, Department of Fluid Mechanics, Polish Academy of Sciences, Swietokrzyska 21, 00-049 Warsaw, Poland
Slawomir Piekarski
Institute of Fundamental Technological Research, Department of Acoustoelectronics, Polish Academy of Sciences, Swietokrzyska 21, 00-049 Warasw, Poland

(Received 7 June 1988; accepted for publication 15 March 1989)

A classical, moderately rarefied, simple, monatomic gas is considered, and it is supposed that its behavior is governed by Boltzmann's equation. Then, after a brief review of the fundamental properties of the symmetric traceless tensors, the irreducible equations of transfer for the relative symmetric traceless moments of the one-point distribution function f are systematically derived and related to those of Johnston. Subsequently, Grad's expansion of the distribution function in terms of reducible three-dimensional Hermite "polynomials," which does not fit in together and with the equations of transfer just mentioned, is rigorously transformed into its irreducible counterpart fashioned by mathematical apparatus such as one-dimensional Laguerre polynomials and Ikenberry's tensorial harmonics. Finally, some useful conversion formulas between the relative symmetric traceless moments and the tensorial Laguerre–Ikenberry expansion coefficients of f are deduced and the irreducible variant of Grad's moment truncation procedure is discussed. The conclusions that have so far been reached concerning Grad's method are quite specific in that they apply to one-dimensional (classical or quasiparticle) gases. In many cases of interest, the treatment of certain aspects of the kinetic theory of "actual" gases requires, as a prerequisite, a comprehensive discussion of some complicated tensorial problems. In this work the so-called irreducible tensor description of three-dimensional gaseous systems is exploited, and this subject is developed only insofar as it relates to Grad's moment procedure and to those *universal* questions which have already been formulated in previous papers. [Z. Banach, J. Stat. Phys. **48**, 813 (1987); Arch. Mech. (to be published); Physica A **129**, 95 (1984); **145**, 105 (1987).]

I. INTRODUCTION

Consider as a starting point of this work Boltzmann's equation describing a classical, moderately rarefied, simple, monatomic gas. One of the basic problems before us, then, is to derive from it, by means of the elementary operations of multilinear algebra,[1] i.e., without the necessity of handling spherical tensors, the infinite hierarchy of the equations of transfer for the *relative irreducible symmetric traceless* moments of the one-point distribution function f.

Working in a somewhat different direction, Coope and Snider[2] proposed a very general formulation of the method of reduction of the tensors, one whose beginnings may be found in the papers by Grad,[3,4] Ikenberry,[5,6] Ikenberry and Truesdell,[7] as well as in Hamermesh's book.[8] With respect to the relative symmetric traceless moments of the distribution function, however, the theoretical results for the irreducible equations of transfer outlined earlier, especially those in the careful texts on Maxwellian iteration[7,9] and in the pioneering work by Johnston,[10] are the strongest yet obtained, being in fact the *only* ones presently available; for more details, see also the important Truesdell–Muncaster monograph.[11]

Although Truesdell *et al.*[7,9,11] have no difficulties in principle to calculate explicitly and for increasing order as many of the irreducible equations of transfer as are needed, they do not provide in their papers the *compact* and *single* expression for an infinite hierarchy of the tensorial symmetric traceless balance equations. To the best of our knowledge, Johnston[10] was the first to invent the general method of deriving it. But for reasons that will become clear later, he did not solve the problem completely and the fundamental result (12) written down on p. 1457 in Ref. 10, although of course compatible with our equation, i.e., the relatively simple result (3.21) of Sec. III, is not equivalent or similar in every respect to it. We shall have more to say about this in Sec. IV A, especially due to the fact that, as elementary inspection reveals, we arrive at the counterpart to Johnston's proposition differently and by the method of straightforward interest also for other gaseous systems and kinetic equations.

In order to account for the potential applicability of the irreducible equations of transfer (in the solution scheme regarding Boltzmann's equation, for instance), one must necessarily exhibit, in terms of the relative symmetric traceless moments, an explicit formula for the irreducible collision integrals—they occur on the rhs of those equations of transfer and depend functionally upon the distribution function f. To this end, apart from a few special cases, e.g., Maxwellian molecules, the very hard particle model, etc., for which, if required, the alternative methods may be developed, a large amount of theoretical work[12–22] has been done, treating only such molecular densities f as can be expanded in a series of the complete set of Hermite or Laguerre polynomials.[23–25]

Obviously, the expansion for the distribution function f

 0022-2488/89/081804-12$02.50

that fits in *together* and *with* the irreducible equations of transfer is now of great interest. It has long been known that the choice of Laguerre polynomials[24,25] for classical gases is "more appropriate" than any other. The new outcome of the present work, at least we hope it to be so, is a direct demonstration of the crucial aspect of this fact in consequence of applying the ordinary transformations of multilinear algebra[1] and, unlike many previous efforts, without the necessity of referring to spherical harmonics.

Although the symmetric traceless tensors have been considered from the mathematical point of view by Coope and Snider,[2] insofar as we have seen they have not been appreciated and systematically used in the literature on the kinetic theory, except for the relevant Truesdell–Ikenberry–Muncaster framework[7,9,11] and the interesting Johnston's approach[10] tending toward a general formulation. It is, then, no surprise that a serious discussion of the above-mentioned problems must be delayed until Sec. III, in order that the essential ideas of the irreducible tensor description for a classical gas may be presented unaccompanied by any marginal arguments of purely technical importance.

Here we proceed as follows. Section II A deals with the simplest properties of the symmetric traceless tensors, whereas Sec. II B introduces the notion of the ∇ operator as well as of its various modifications. Our presentation of some of the auxiliary concepts is necessarily very brief, being a subject of Appendices A and B only for completeness. In Sec. III and Appendices C and D the algebraic difficulties in calculating the irreducible equations of transfer are faced once more.[10] Section IV A achieves the object of transforming Grad's expansion of the distribution function f in terms of three-dimensional Hermite polynomials[23] into its irreducible counterpart fashioned by mathematical apparatus such as Laguerre polynomials[24,25] and Ikenberry's tensorial harmonics.[5,6] In the older view the irreducible moment representation of the distribution function f was postulated rather than logically deduced, and Ikenberry's harmonics were replaced by their spherical analogues; see, however, Ref. 26 and the literature quoted there. Section IV B discusses, after demonstrating that the tensorial Laguerre–Ikenberry expansion coefficients of f show up in a very particular way in the approximate expression for the entropy density h, the irreducible variant of Grad's moment truncation procedure. Given a gas of Maxwellian molecules, Sec. V offers some useful but elementary comments regarding the general structure of the symmetric traceless collision integrals. We conclude our work with final remarks of Sec. VI.

Whether the direct notation for tensors is preferable over the Cartesian one must be decided individually for each problem. In the present paper, except for Appendices A, B, and C, we make no use of the latter.

II. PROLEGOMENA

The symmetrizer Π, the trace operator with respect to the pair (β,ν), denoted by $\mathrm{Tr}_{(\beta,\nu)}$, the inner product $M^{\alpha}\circ M^{\beta}$ of the tensors M^{α} and M^{β}, the so-called contravariant metric (unit) tensor of a three-dimensional Euclidean vector space $\mathbb{E}$, denoted by I, and the ∇ operator are precisely defined in Appendix A.

A. Symmetric traceless tensors

(1) Let $\mathbb{E}$ be a three-dimensional Euclidean vector space and consider for each $\alpha\geqslant 2$ the αth tensorial power $\mathbb{E}^{\alpha} := \otimes^{\alpha}\mathbb{E}$ of $\mathbb{E}$. We extend the definition of $\mathbb{E}^{\alpha}$ to the cases $\alpha=1$ and $\alpha=0$ by setting $\mathbb{E}^1 := \mathbb{E}$ and $\mathbb{E}^0 := \mathbb{R}$, where $\mathbb{R}$ stands for the set of real numbers. The image space $\Pi\mathbb{E}^{\alpha}$ $(\alpha\geqslant 0)$ of the symmetrizer Π in $\mathbb{E}^{\alpha}$ will be denoted by $\mathbb{E}_s^{\alpha}$. Elementary calculus shows that[1,2]

$$\dim \mathbb{E}^{\alpha} = 3^{\alpha}, \quad \dim \mathbb{E}_s^{\alpha} = \tfrac{1}{2}(\alpha+1)(\alpha+2), \quad \alpha\geqslant 2. \tag{2.1}$$

(2) Suppose that M^{α} and M^{β} are the tensors of degrees α and β, respectively ($M^{\alpha}\in\mathbb{E}^{\alpha}$, $M^{\beta}\in\mathbb{E}^{\beta}$). Then the equality

$$M^{\alpha}\vee M^{\beta} := \Pi(M^{\alpha}\otimes M^{\beta})\in\mathbb{E}_s^{\alpha+\beta}, \quad \alpha\geqslant 0, \quad \beta\geqslant 0, \tag{2.2}$$

defines the *symmetric tensor product* of M^{α} and M^{β} ($M^0\otimes M^{\alpha} := M^{\alpha}\otimes M^0 := M^0 M^{\alpha}$, $M^0\in\mathbb{R}$).

(3) The trace operator with respect to the pair (β,ν), denoted by $\mathrm{Tr}_{(\beta,\nu)}$, determines a linear map:

$$\mathop{\mathrm{Tr}}_{(\beta,\nu)} : \mathbb{E}^{\alpha}\Rightarrow\mathbb{E}^{\alpha-2}, \quad \alpha\geqslant 2, \quad \alpha\geqslant\beta\geqslant 1,$$

$$\alpha\geqslant\nu\geqslant 1, \quad \beta\neq\nu.$$

The trace operator Tr is the restriction of a linear map $\mathrm{Tr}_{(\beta,\nu)}$ with the arbitrarily chosen pair (β,ν) to $\mathbb{E}_s^{\alpha}\subset\mathbb{E}^{\alpha}$ $(\alpha\geqslant 2)$. Moreover, we define Tr to be the identity on $\mathbb{E}_s^1=\mathbb{E}$ and $\mathbb{E}_s^0=\mathbb{R}$:

$$\mathrm{Tr}\, M^1 := M^1, \quad \mathrm{Tr}\, M^0 = M^0.$$

(4) The kernel of Tr in $\mathbb{E}_s^{\alpha}$, denoted by $\mathrm{Ker}_{\alpha}\,\mathrm{Tr}$ $(\alpha\geqslant 2)$, is the subset of tensors $M^{\alpha}\in\mathbb{E}_s^{\alpha}$ such that $\mathrm{Tr}\,M^{\alpha}=0$. We extend the definition of $\mathrm{Ker}_{\alpha}\,\mathrm{Tr}$ to the cases $\alpha=1$ and $\alpha=0$ by setting $\mathrm{Ker}_1\,\mathrm{Tr} := \mathbb{E}_s^1=\mathbb{E}$ and $\mathrm{Ker}_0\,\mathrm{Tr} := \mathbb{E}_s^0=\mathbb{R}$. The elements of $\mathrm{Ker}_{\alpha}\,\mathrm{Tr}$ will be called *symmetric traceless tensors* of degree α. The image space of Tr in $\mathbb{E}_s^{\alpha}$, denoted by $\mathrm{Im}_{\alpha}\,\mathrm{Tr}$ $(\alpha\geqslant 2)$, is the set of tensors $M^{\alpha-2}\in\mathbb{E}_s^{\alpha-2}$ of the form $M^{\alpha-2}=\mathrm{Tr}\,M^{\alpha}$ for some $M^{\alpha}\in\mathbb{E}_s^{\alpha}$. Adopting the standard theorem associated with the notion of the rank of a linear mapping of finite-dimensional vector spaces, we find that

$$\dim \mathrm{Ker}_{\alpha}\,\mathrm{Tr} + \dim \mathrm{Im}_{\alpha}\,\mathrm{Tr} = \dim \mathbb{E}_s^{\alpha}, \quad \alpha\geqslant 2. \tag{2.3}$$

(5) Given $M^{\alpha}\in\mathbb{E}_s^{\alpha}$ and $M^{\alpha+2\beta}\in\mathbb{E}_s^{\alpha+2\beta}$, it is useful to introduce the following tensors:

$$\hat{M}^{\alpha[\beta]} := \mathrm{Tr}^{\beta}\, M^{\alpha}, \quad \hat{M}^{\alpha|\beta} := \mathrm{Tr}^{\beta}\, M^{\alpha+2\beta}, \tag{2.4}$$

where $\mathrm{Tr}^{\beta}\,M^{\alpha}\in\mathbb{E}_s^{\alpha-2\beta}$ and $\mathrm{Tr}^{\beta}\,M^{\alpha+2\beta}\in\mathbb{E}_s^{\alpha}$ are the results of the β-fold successive application of the Tr operator to the tensors M^{α} and $M^{\alpha+2\beta}$, respectively ($\mathrm{Tr}^0\,M^{\alpha} := M^{\alpha}$).

(6) The action of $\cup$ on $M^{\alpha}\in\mathbb{E}^{\alpha}$ and $M^{\beta}\in\mathbb{E}^{\beta}$ $(\alpha+\beta\geqslant 2)$ is characterized by

$$M^{\alpha}\cup M^{\beta} := \Pi \mathop{\mathrm{Tr}}_{(1,\alpha+\beta)} (M^{\alpha}\otimes M^{\beta})\in\mathbb{E}_s^{\alpha+\beta-2}. \tag{2.5}$$

We call $M^{\alpha}\cup M^{\beta}$ the *contracted symmetric tensor product* of M^{α} and M^{β}.

(7) Let us suppose that $\nu := \min(\alpha,\beta)$. Then in contracting $M^{\alpha}\in\mathbb{E}^{\alpha}$ with $M^{\beta}\in\mathbb{E}^{\beta}$ the ν-fold contraction[2] is denoted by $\circ$:

$$M^\alpha \circ M^\beta = M^\beta \circ M^\alpha \in \mathbb{E}^{\alpha+\beta-2\nu}, \qquad (2.6a)$$

$$M^0 \circ M^\alpha = M^\alpha \circ M^0 = M^0 M^\alpha. \qquad (2.6b)$$

However, some convention, as to which of the 2ν indices are to be contracted, must be followed when doing the contraction; for more details, see Appendix A. The tensor $M^\alpha \circ M^\beta$ will be termed the *inner product* of M^α and M^β.

(8) The norm $|M^\alpha|$ of the tensor $M^\alpha \in \mathbb{E}^\alpha$ is defined as the positive square root of $M^\alpha \circ M^\alpha$:

$$|M^\alpha| := (M^\alpha \circ M^\alpha)^{1/2}. \qquad (2.7)$$

(9) The abbreviated symbols I and I^α $(\alpha \geqslant 2)$ stand for the so-called contravariant metric (unit) tensor of $\mathbb{E}$ $(I \in \mathbb{E}_s^2)$ and the tensor product $\otimes^\alpha I \in \mathbb{E}^{2\alpha}$, respectively. We extend the definition of I^α to the cases $\alpha = 1$ and $\alpha = 0$ by setting $I^1 := I$ and $I^0 := 1 \in \mathbb{E}^0 = \mathbb{R}$.

(10) Assume that

$$M^\alpha * I^\beta := \mathbf{l}(\alpha + 2\beta, \beta)\, M^\alpha \vee I^\beta \in \mathbb{E}_s^{\alpha+2\beta}, \qquad (2.8a)$$

where

$$\mathbf{l}(\alpha,\beta) := \alpha!/2^\beta \beta!(\alpha - 2\beta)!, \quad \alpha \geqslant 2\beta. \qquad (2.8b)$$

Then certain useful properties of the Tr operator are included in the statement of the following theorem.

Theorem: (a) Suppose that $\alpha \geqslant 2$, $\beta \geqslant 1$, and $M^\alpha \in \mathbb{E}_s^\alpha$. Then we have

$$\mathrm{Tr}(M^\alpha * I^\beta) = \widehat{M}^{\alpha[1]} * I^\beta + (2\alpha + 2\beta + 1) M^\alpha * I^{\beta - 1}. \qquad (2.9a)$$

(b) Let $\alpha = 0,1$ and $\beta \geqslant 1$. Then we obtain

$$\mathrm{Tr}(M^\alpha * I^\beta) = (2\alpha + 2\beta + 1) M^\alpha * I^{\beta-1}. \qquad (2.9b)$$

(c) Assume that $M^\alpha \in \mathrm{Ker}_\alpha \mathrm{Tr}$ $(\alpha \geqslant 2)$ and consider the quantity

$$\mathbf{k}(\alpha,\beta) := \frac{(2\alpha + 1)!!}{(2\alpha + 2\beta + 1)!!}. \qquad (2.10a)$$

Then the equality

$$\mathrm{Tr}^\beta(M^\alpha * I^\omega) = [\mathbf{k}(\alpha + \omega - \beta, \beta)]^{-1} M^\alpha * I^{\omega - \beta}, \quad \omega \geqslant \beta \qquad (2.10b)$$

holds; Eq. (2.10b) holds also for $\alpha = 0,1$.

The equality (2.10b) can easily be derived from (2.9). Insofar as Eqs. (2.9) are concerned, the main ingredients involved in the algebraically difficult and very tedious calculation of $\mathrm{Tr}(M^\alpha * I^\beta)$ have been satisfactorily presented by Grad on pp. 393–395 in Ref. 3, although for the special case in which $\alpha = 0$. Thus the uninteresting proof of (2.9) will not be given. ∎

(11) The linear operator Appr: $\mathbb{E}_s^\alpha \Rightarrow \mathbb{E}_s^{\alpha+2}$ $(\alpha \geqslant 0)$ is determined by

$$\mathrm{Appr}\, M^\alpha := \sum_{\beta=0}^{[\alpha/2]} (-1)^\beta \mathbf{k}(\alpha - \beta, \beta + 1) \times \widehat{M}^{\alpha[\beta]} * I^{\beta+1}, \qquad (2.11)$$

where

$[u]:=$ the greatest integer $\leqslant u$.

In view of the theorem just formulated, the foregoing definition yields Tr Appr $M^\alpha = M^\alpha$. With this observation in mind, and by no more than a direct reasoning, we arrive at

$$\mathrm{Im}_\alpha\, \mathrm{Tr} = \mathbb{E}_s^{\alpha-2}, \quad \alpha \geqslant 2. \qquad (2.12)$$

Carrying our argumentation a step further, we see from (2.1), (2.3), and (2.12) that $\dim \mathrm{Ker}_\alpha \mathrm{Tr} = 2\alpha + 1$.

(12) Let $M^\alpha \in \mathbb{E}_s^\alpha$ $(\alpha \geqslant 2)$. Then the symmetric traceless tensor $\langle M^\alpha \rangle$ is defined by

$$\langle M^\alpha \rangle := M^\alpha - \mathrm{Appr}\, \widehat{M}^{\alpha[1]} \in \mathrm{Ker}_\alpha \mathrm{Tr}. \qquad (2.13a)$$

Moreover, we postulate that

$$\langle M^1 \rangle := M^1, \quad \langle M^0 \rangle := M^0. \qquad (2.13b)$$

We will show in Appendix B that

$$\langle M^\alpha * I^\beta \rangle := \mathbf{l}(\alpha + 2\beta, \beta) \langle M^\alpha \vee I^\beta \rangle = 0 \text{ for } \beta > 0. \qquad (2.14)$$

(13) Construct $M^{\alpha[\beta]}$ and $M^{\alpha|\beta}$:

$$M^{\alpha[\beta]} := \langle \widehat{M}^{\alpha[\beta]} \rangle, \quad M^{\alpha|\beta} := \langle \widehat{M}^{\alpha|\beta} \rangle. \qquad (2.15)$$

Then, as the *important* equality[10,12]

$$M^\alpha = \sum_{\beta=0}^{[\alpha/2]} \mathbf{k}(\alpha - 2\beta, \beta)\, M^{\alpha[\beta]} * I^\beta, \quad \alpha \geqslant 0, \qquad (2.16)$$

reveals, each tensor $M^\alpha \in \mathbb{E}_s^\alpha$ $(\alpha \geqslant 0)$ can directly be decomposed in terms of $M^{\alpha[\beta]} \in \mathrm{Ker}_{\alpha - 2\beta} \mathrm{Tr}$.

(14) Given $M^\alpha \in \mathbb{E}^\alpha$ and $M^\beta \in \mathbb{E}^\beta$, we introduce the *symmetric traceless tensor product*:

$$M^\alpha \wedge M^\beta := \langle M^\alpha \vee M^\beta \rangle \in \mathrm{Ker}_{\alpha+\beta} \mathrm{Tr}, \qquad (2.17)$$

and the *contracted symmetric traceless tensor product*:

$$M^\alpha \cap M^\beta := \langle M^\alpha \cup M^\beta \rangle \in \mathrm{Ker}_{\alpha+\beta-2} \mathrm{Tr}. \qquad (2.18)$$

(15) The symmetric traceless tensors constructed from a decomposable tensor and, in particular, those generated by a *single unit* vector g ($\otimes^\alpha g$; $\otimes^1 g := g$, $\otimes^0 g := 1 \in \mathbb{E}^0 = \mathbb{R}$),

$$Y^\alpha(g) := \langle \otimes^\alpha g \rangle \in \mathrm{Ker}_\alpha \mathrm{Tr}, \quad \alpha \geqslant 0, \quad |g| = 1, \qquad (2.19)$$

are used quite widely and have been discussed by several authors.[2,5–12,15] We call $Y^\alpha(g)$ *Ikenberry's tensorial harmonics*. Weinert[15] was able to prove that $Y^\alpha(g)$ are nothing else than Maxwell's multipole representations of spherical harmonics.[25]

(16) The projection of $\mathbb{E}^\alpha$ onto the irreducible subspace Ker_α Tr of symmetric traceless tensors of degree α will be denoted by $E(\alpha|\alpha)$. The natural projection $E(\alpha|\alpha) \in \mathbb{E}^{2\alpha}$, for which we have

$$E(\alpha|\alpha) \circ M^\alpha := \langle \Pi M^\alpha \rangle, \qquad (2.20)$$

is separately symmetric traceless in both sets of α indices, because, by definition, it is symmetric traceless in one set; in this context, see the work of Coope and Snider.[2]

B. The ∇ operator and its various modifications

Let M^α be a differentiable tensor field in $\mathbb{E}$ and ∇ be a gradient with respect to $x \in \mathbb{E}$. Given the ∇ operator, for the most part we shall deal with a variety of other tensorial operators associated with ∇.

(1) The action of $\nabla \circ$ on a tensor field M^α of degree $\alpha \geqslant 1$ is defined by

$$\nabla \circ M^\alpha := \underset{(1,\alpha+1)}{\mathrm{Tr}} \nabla M^\alpha \in \mathbb{E}^{\alpha - 1}. \qquad (2.21)$$

(2) In accord with the Meyer–Schröter work,[18–20] the action of the symmetric operator $\nabla\vee$ on a tensor field M^α is defined by

$$\nabla\vee M^\alpha := \Pi\nabla M^\alpha \in \mathbb{E}_s^{\alpha+1}. \tag{2.22}$$

(3) Finally, the effect of the symmetric traceless operator $\nabla\wedge$ on a tensor field M^α is given by

$$\nabla\wedge M^\alpha := \langle\nabla\vee M^\alpha\rangle \in \mathrm{Ker}_{\alpha+1}\,\mathrm{Tr}. \tag{2.23}$$

III. REDUCIBLE AND IRREDUCIBLE EQUATIONS OF TRANSFER FOR THE MOMENTS

Given a three-dimensional, classical, moderately rarefied, simple, monatomic gas for which each molecule of unit mass is subject to an external body force K_1 [K_1 being a function of position x and time t but not of momentum (velocity) λ], we start our discussion with the following equation of change for the one-particle density $f(\lambda,x,t)$:

$$\partial_t f + \lambda\circ\nabla_x f + a\circ\nabla_\lambda f = J(f), \tag{3.1}$$

where a is the molecular acceleration and J denotes the Boltzmann collision operator.[3,4] The molecular acceleration a is tied to the external body force K_1 through Newton's second law which takes the form[27,28]

$$a = K_1 + K_2 - \lambda\circ L_2 \tag{3.2}$$

in any arbitrary noninertial frame. Here K_2 and L_2 are, respectively, the velocity-independent part of the inertial force and the skew-symmetric tensor associated with the angular velocity of the reference frame relative to an inertial framing.

For immediate use we introduce the *relative reducible* moment $M^\alpha \in \mathbb{E}_s^\alpha$ of the form

$$M^\alpha := \int_{\mathbb{E}} (\otimes^\alpha \bar\lambda) f\, d\lambda, \quad \alpha\geqslant 0, \tag{3.3}$$

where

$$\bar\lambda := \lambda - u, \quad u := \int_{\mathbb{E}} \lambda f\, d\lambda \left(\int_{\mathbb{E}} f\, d\lambda\right)^{-1}. \tag{3.4}$$

Of course, the quantity u can be interpreted as a macroscopic velocity field.

By (3.1)–(3.4) we easily show that

$$\partial_t M^\alpha + \nabla\circ(M^\alpha\otimes u + M^{\alpha+1}) + \alpha L\cup M^\alpha + \alpha(\partial_t u + u\circ L - K)\vee M^{\alpha-1} = P^\alpha, \quad \alpha\geqslant 0, \tag{3.5a}$$

$$K := K_1 + K_2, \quad L := L_1 + L_2, \tag{3.5b}$$

$$L_1 := \nabla u, \quad M^{-1} := 0, \tag{3.5c}$$

where the collision integrals P^α are defined by

$$P^\alpha := \int_{\mathbb{E}} (\otimes^\alpha\bar\lambda) J(f)\, d\lambda \in \mathbb{E}_s^\alpha, \quad \alpha\geqslant 0. \tag{3.6}$$

[To simplify our notation, in Eq. (3.5a) the ∇_x operator is denoted by ∇.] From the balance of linear momentum, we see at once that the evolution of M^α is unaffected by both external body forces and translational accelerations of the noninertial frame and is affected by rotations of the reference frame only through the molecular Coriolis force represented by $-\lambda\circ L_2$. Indeed, combining

$$\partial_t u + u\circ L + R = K, \quad R := (1/M^0)\nabla\circ M^2 \tag{3.7}$$

with (3.5a), we arrive at

$$\partial_t M^\alpha + \nabla\circ(M^\alpha\otimes u + M^{\alpha+1}) + \alpha(L\cup M^\alpha - R\vee M^{\alpha-1}) = P^\alpha, \quad \alpha\geqslant 0. \tag{3.8}$$

We now wish to obtain the balance equations for $\hat M^{\alpha[\beta]}$. To this end we appeal to the following lemma.

Lemma: Let $M^\alpha\in\mathbb{E}_s^\alpha$, $L\in\mathbb{E}^2$, and $R\in\mathbb{E}$ be three arbitrary tensors. Then the action of the operator $\alpha\,\mathrm{Tr}^\beta$ on $L\cup M^\alpha$ and $R\vee M^{\alpha-1}$ is given by

$$\alpha\,\mathrm{Tr}^\beta(L\cup M^\alpha) = (\alpha-2\beta)L\cup\hat M^{\alpha[\beta]} + 2\beta L\circ\hat M^{\alpha[\beta-1]}, \tag{3.9a}$$

$$\alpha\,\mathrm{Tr}^\beta(R\vee M^{\alpha-1}) = (\alpha-2\beta)R\vee\hat M^{\alpha-1[\beta]} + 2\beta R\circ\hat M^{\alpha-1[\beta-1]}, \quad \alpha\geqslant 2\beta\geqslant 0. \tag{3.9b}$$

Proof: If $\beta=1$, the lemma follows from the argumentation in Appendix C. Now assume that (3.9a) is true for $\beta-1$ ($\beta\geqslant 2$). Then we have

$$\alpha\mathrm{Tr}^\beta(L\cup M^\alpha) = \mathrm{Tr}[\alpha\,\mathrm{Tr}^{\beta-1}(L\cup M^\alpha)] = (\alpha-2\beta+2)\mathrm{Tr}(L\cup\hat M^{\alpha[\beta-1]}) + 2(\beta-1)L\circ\hat M^{\alpha[\beta-1]}.$$

Applying (3.9a) for $\beta=1$ and $\alpha=\alpha'-2\beta+2$ ($\alpha'\Rightarrow\alpha$), we obtain

$$\alpha\mathrm{Tr}^\beta(L\cup M^\alpha) = (\alpha-2\beta)L\cup\hat M^{\alpha[\beta]} + 2L\circ\hat M^{\alpha[\beta-1]} + 2(\beta-1)L\circ\hat M^{\alpha[\beta-1]},$$

and so the induction is closed. In exactly the same way one can show that Eq. (3.9b) holds also. This completes the proof of (3.9). ∎

Due to the fact that for certain admissible choices of the pair (α,β) the contracted moments

$$\mathbb{M}_\beta^\alpha := \{\hat M^{\alpha[\beta-1]}, \hat M^{\alpha-1[\beta]}, \hat M^{\alpha-1[\beta-1]}\}$$

appearing on the rhs of (3.9a) and (3.9b) are not *well defined*, we extend the range of validity of Eqs. (3.9) to these cases by setting then $\mathbb{M}_\beta^\alpha = 0$.

In view of (3.8) and (3.9), an equation of change for $\hat M^{\alpha[\beta]}$ ($\alpha\geqslant 2\beta\geqslant 0$) is easily seen to be

$$\partial_t\hat M^{\alpha[\beta]} + \nabla\circ(\hat M^{\alpha[\beta]}\otimes u + \hat M^{\alpha+1[\beta]}) + (\alpha-2\beta)(L\cup\hat M^{\alpha[\beta]} - R\vee\hat M^{\alpha-1[\beta]}) + 2\beta(L\circ\hat M^{\alpha[\beta-1]} - R\circ\hat M^{\alpha-1[\beta-1]}) = \hat P^{\alpha[\beta]}, \tag{3.10}$$

where, according to our notation [see Sec. II A, Eqs. (2.4)],

$$\hat P^{\alpha[\beta]} := \mathrm{Tr}^\beta P^\alpha. \tag{3.11}$$

To proceed further in the calculation of the irreducible equations of transfer, we have to evaluate the action of the natural projection $E(\alpha-2\beta\,|\,\alpha-2\beta)$ upon each tensorial expression in Eq. (3.10) directly in terms of $M^{\alpha[\beta]}$ or of $M^{\alpha|\beta}$; we call $M^{\alpha[\beta]}\in\mathrm{Ker}_{\alpha-2\beta}\,\mathrm{Tr}$ and $M^{\alpha|\beta}\in\mathrm{Ker}_\alpha\,\mathrm{Tr}$ the *relative symmetric traceless moments* of the one-point distribution function f. This decisive step is the only one which is algebraically difficult, and so for the moment we simply assert that

$$\langle\partial_t\hat M^{\alpha[\beta]}\rangle = \partial_t M^{\alpha[\beta]}, \tag{3.12}$$

$$\langle\nabla\circ(\hat M^{\alpha[\beta]}\otimes u)\rangle = \nabla\circ(M^{\alpha[\beta]}\otimes u), \tag{3.13}$$

$$\langle \nabla \circ (\hat{M}^{\alpha+1[\beta]}) \rangle = \frac{\alpha - 2\beta}{2\alpha - 4\beta + 1} \nabla \wedge M^{\alpha+1[\beta+1]} + \nabla \circ M^{\alpha+1[\beta]}, \tag{3.14}$$

$$(\alpha - 2\beta)\langle L \cup \hat{M}^{\alpha[\beta]} \rangle = (\alpha - 2\beta) \frac{\alpha - 2\beta - 1}{2\alpha - 4\beta - 1} L \wedge M^{\alpha[\beta+1]} + (\alpha - 2\beta) L \cap M^{\alpha[\beta]}, \tag{3.15}$$

$$(\alpha - 2\beta)\langle R \vee \hat{M}^{\alpha-1[\beta]} \rangle = (\alpha - 2\beta) R \wedge M^{\alpha-1[\beta]}, \tag{3.16}$$

$$2\beta \langle L \circ \hat{M}^{\alpha[\beta-1]} \rangle = \frac{2\beta}{2\alpha - 4\beta + 3} S M^{\alpha[\beta]} + \frac{2\beta(\alpha - 2\beta)(\alpha - 2\beta - 1)}{(2\alpha - 4\beta - 1)(2\alpha - 4\beta + 1)} L \wedge M^{\alpha[\beta+1]} + \frac{2\beta(\alpha - 2\beta)}{2\alpha - 4\beta + 3} L \cap M^{\alpha[\beta]} + \frac{2\beta(\alpha - 2\beta)}{2\alpha - 4\beta + 3} M^{\alpha[\beta]} \cap L + 2\beta L \circ M^{\alpha[\beta-1]}, \tag{3.17}$$

$$2\beta \langle R \circ \hat{M}^{\alpha-1[\beta-1]} \rangle = \frac{2\beta(\alpha - 2\beta)}{2\alpha - 4\beta + 1} R \wedge M^{\alpha-1[\beta]} + 2\beta R \circ M^{\alpha-1[\beta-1]}, \tag{3.18}$$

$$\langle \hat{P}^{\alpha[\beta]} \rangle = P^{\alpha[\beta]}, \tag{3.19}$$

where

$$S := \mathrm{Tr}\, \Pi L = \mathrm{Tr}\, \Pi L_1. \tag{3.20}$$

For completeness, however, we shall present in Appendix D a sketch of the proof of the equality (3.17). Then the derivation of Eqs. (3.12)–(3.16), (3.18), and (3.19) can be repeated essentially word for word with only slight technical changes in the method.

A glance at Eqs. (3.10), (3.12)–(3.19), and the transformation rule $M^{\alpha+2\beta[\beta]} = M^{\alpha|\beta}$ shows that

$$\partial_t M^{\alpha|\beta} + \nabla \circ (M^{\alpha|\beta} \otimes u + M^{\alpha+1|\beta}) + \frac{\alpha}{2\alpha+1} \nabla \wedge M^{\alpha-1|\beta+1} + \frac{1}{2\alpha+3}(2\beta S M^{\alpha|\beta} + \alpha(2\alpha + 2\beta + 3) L \cap M^{\alpha|\beta} + 2\alpha\beta M^{\alpha|\beta} \cap L) + \frac{\alpha(2\alpha + 2\beta + 1)}{2\alpha + 1}\left(\frac{\alpha - 1}{2\alpha - 1} L \wedge M^{\alpha-2|\beta+1} - R \wedge M^{\alpha-1|\beta}\right) + 2\beta(L \circ M^{\alpha+2|\beta-1} - R \circ M^{\alpha+1|\beta-1}) = P^{\alpha|\beta}, \tag{3.21}$$

for $\alpha \geqslant 0$ and $\beta \geqslant 0$, where, by definition,

$$M^{\alpha|-1} := 0, \quad M^{-1|\beta} := 0, \quad M^{-2|\beta} := 0 \tag{3.22}$$

and

$$P^{\alpha|\beta} = P^{\alpha+2\beta[\beta]}. \tag{3.23}$$

As elementary inspection reveals, the differential equations just obtained do not seperate; in general, not only $M^{\nu|\gamma}$ with $\nu + 2\gamma = \alpha + 2\beta$ and $\gamma = \beta - 1 < \beta$ occurs in Eq. (3.21), but also the symmetric traceless moments $M^{\nu|\gamma}$ with $\nu + 2\gamma = \alpha + 2\beta + 1 > \alpha + 2\beta$ are found in the irreducible equation of transfer for $M^{\alpha|\beta}$, at least insofar as the lhs of that equation is concerned. Even if one ignores a very difficult problem concerning the form of the explicit and exact dependence of $P^{\alpha|\beta}$ upon $M^{\nu|\gamma}$, these observations demonstrate the so-called forward coupling of the equations of moments.

In conclusion, we rest content to point out that the special cases of Eq. (3.21) were derived and extensively applied, as one of many segments of the whole procedure called Maxwellian iteration, by Truesdell and Ikenberry in the middle 1950's, but, to the best of our knowledge, they have not to this day published the *single* and *compact* expression (3.21). Johnston[10] was the first to solve some of the nasty tensorial problems of this work and to invent the general method of deriving the irreducible equations of transfer for Ikenberry's tensorial expansion coefficients $f_\alpha \in \mathrm{Ker}_\alpha$ Tr of f. On the other side, since the irreducible "moments" f_α are not fields in the ordinary sense of continuum mechanics (aside from the dependence upon position x and time t, they are functions of $|\hat{\lambda}| \in \mathbb{R}$), it is hoped that the present theory of the equations of moments constitutes partly an extension and partly a supplement to that in Refs. 7, 10, and 11.

IV. ILLUSTRATION OF THE IDEA OF IRREDUCIBLE TENSOR DESCRIPTION, APPLIED TO THE MOMENT PROCEDURE OF GRAD'S TYPE

A. Reducible and irreducible representations of the distribution function

In the paper in which Grad[3] developed his 13 moment approximation to the velocity distribution function f for a rarefied gas he utilized certain three-dimensional Hermite polynomials, whose properties he considered in a companion paper.[23] In terms of a dimensionless vector $\hat{\lambda}$, the Hermite polynomials, which are components of tensors $B^\alpha(\hat{\lambda}) \in \mathbb{E}_s^\alpha$ defined by

$$B^\alpha(\hat{\lambda}) := (-1)^\alpha \mathbb{W}_0^{-1}\left(\frac{\partial}{\partial \hat{\lambda}}\right)^\alpha \mathbb{W}_0, \tag{4.1a}$$

$$\mathbb{W}_0(\hat{\lambda}) := (2\pi)^{-3/2} \exp(-\tfrac{1}{2}|\hat{\lambda}|^2), \tag{4.1b}$$

are of total degree α in the three components of $\hat{\lambda} \in \mathbb{E}$, as is readily seen from (4.1). [In Eq. (4.1a) $(\partial/\partial\hat{\lambda})^\alpha \mathbb{W}_0$ is the result of the α-fold successive application of the differential operator $\partial/\partial\hat{\lambda} := \nabla_{\lambda|\lambda=\hat{\lambda}}$ to the scalar function $\mathbb{W}_0$.] Inspection shows that the Hermite polynomials are orthogonal with respect to $\mathbb{W}_0$.[23] In addition, in the space of functions which are square integrable over $\mathbb{E}$ with weight $\mathbb{W}_0$, the Hermite polynomials are complete.[23]

Grad considers only such molecular densities f as can be expanded in a series of Hermite (tensorial) polynomials $B^\alpha(\hat{\lambda})$. In order to set down the expansion of the distribution function f around a local Maxwellian f_0 given by

$$f_0(\lambda,x,t) := \rho(x,t)\,[k_{\mathrm{B}}T(x,t)]^{-3/2}\mathbb{W}_0(\hat{\lambda}), \tag{4.2a}$$

$$\rho := M^0, \quad k_{\mathrm{B}}T := \tfrac{1}{3}\operatorname{Tr} M^2/M^0, \tag{4.2b}$$

$$\hat{\lambda} := (k_{\mathrm{B}}T)^{-1/2}\overline{\lambda} = (k_{\mathrm{B}}T)^{-1/2}(\lambda - u), \tag{4.2c}$$

we let ρ, u, and $\frac{3}{2}k_{\mathrm{B}}T$ be the mass density, the macroscopic velocity, and the internal energy per unit mass, respectively, that correspond to f.[29] In the result we obtain

$$f(\lambda,x,t) = f_0(\lambda,x,t)\sum_{\alpha=0}^{\infty}\frac{1}{\alpha!}\,b^\alpha(x,t)\circ B^\alpha(\hat{\lambda}). \tag{4.3}$$

If we multiply each side of (4.3) by $B^\alpha(\hat{\lambda})$, integrate with respect to λ, and make use of the orthogonality properties of $B^\alpha(\hat{\lambda})$, then we find that the so-called Hermite expansion coefficients $b^\alpha \in \mathbb{E}_s^\alpha$ of the distribution function are functionally related to f:

$$b^\alpha = \frac{1}{\rho}\int_{\mathbb{E}} B^\alpha(\hat{\lambda})\,f\,d\lambda. \tag{4.4}$$

Both by placing (4.1) into (4.4) and by appealing to the definitions of ρ, u, and $k_{\mathrm{B}}T$, we arrive at

$$b^0 = 1, \quad b^1 = \operatorname{Tr} b^2 = 0. \tag{4.5}$$

Of course, Grad's expansion is not ideally suited to the irreducible equations of transfer. However, once Eqs. (4.1)–(4.5) have been set down, we may derive from them, through the elementary operations of multilinear algebra, the alternative representation of the distribution function f that fits in together and with the system (3.21), omitting but not disregarding other independent, important approaches fashioned by mathematical apparatus such as spherical tensors and harmonics. [In fact, spherical harmonics have been extremely useful for the calculation of collision integrals for all interaction potentials, not only the Maxwellian one; see, e.g., Suchy[13,14] and Weinert.[15,16]] Insofar as we are aware, in the literature on the kinetic theory the nasty problems generated by, and associated with, the *rigorous* transformation of Grad's expansion of the distribution function f into its irreducible analog have been outflanked rather than logically solved.

We begin by eliminating $B^\alpha(\hat{\lambda})$ in (4.3) through use of

$$B^\alpha = \sum_{\beta=0}^{[\alpha/2]}\mathbb{l}(\alpha,\beta)\mathbb{k}(\alpha-2\beta,\beta)\,B^{\alpha[\beta]}\vee I^\beta. \tag{4.6}$$

Hence

$$f = f_0\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{[\alpha/2]}\frac{1}{\alpha!}\mathbb{l}(\alpha,\beta)\,\mathbb{k}(\alpha-2\beta,\beta) \times b^{\alpha[\beta]}\circ B^{\alpha[\beta]}. \tag{4.7}$$

To say more about f than Eq. (4.7) asserts, we must have precise knowledge of $B^{\alpha[\beta]}$. Recalling now[23]

$$B^\alpha(\hat{\lambda}) = \sum_{\omega=0}^{[\alpha/2]}(-1)^\omega\mathbb{l}(\alpha,\omega)\,[(\otimes^{\alpha-2\omega}\hat{\lambda})\vee I^\omega] \tag{4.8}$$

as well as (2.8a), (2.9a), and (2.14), we obtain for $B^{\alpha[\beta]}$

$$B^{\alpha[\beta]} = \langle \operatorname{Tr}^\beta B^\alpha\rangle = \sum_{\omega=0}^{\beta}(-1)^\omega\mathbb{l}(\alpha,\omega)\,|\hat{\lambda}|^{\alpha-2\omega} \times\langle \operatorname{Tr}^\beta[(\otimes^{\alpha-2\omega}g)\vee I^\omega], \tag{4.9}$$

where

$$g := |\hat{\lambda}|^{-1}\hat{\lambda}. \tag{4.10}$$

Using for $M^{\alpha-2\omega}$ here the tensor $\otimes^{\alpha-2\omega}g$, and observing that $M^{\alpha-2\omega[\gamma]} = Y^{\alpha-2\omega-2\gamma}(g)$, we obtain from the decomposition (2.16) in which $\alpha\Rightarrow\alpha-2\omega$ the equality

$$\otimes^{\alpha-2\omega}g = \sum_{\gamma=0}^{[(\alpha-2\omega)/2]}\mathbb{l}(\alpha-2\omega,\gamma)\mathbb{k}(\alpha-2\omega-2\gamma,\gamma) \times Y^{\alpha-2\omega-2\gamma}(g)\vee I^\gamma. \tag{4.11}$$

Clearly, in virtue of (4.11), Eq. (4.9) may be written as

$$\begin{aligned} B^{\alpha[\beta]} &= \sum_{\omega=0}^{\beta}\sum_{\gamma=0}^{[(\alpha-2\omega)/2]}(-1)^\omega\mathbb{l}(\alpha,\omega)\mathbb{l}(\alpha-2\omega,\gamma)\mathbb{k}(\alpha-2\omega-2\gamma,\gamma)\,|\hat{\lambda}|^{\alpha-2\omega}\langle\operatorname{Tr}^\beta[Y^{\alpha-2\omega-2\gamma}(g)\vee I^{\omega+\gamma}]\rangle \\ &= \sum_{\omega=0}^{\beta}(-1)^\omega\mathbb{l}(\alpha,\omega)\mathbb{l}(\alpha-2\omega,\beta-\omega)\,\mathbb{k}(\alpha-2\beta,\beta-\omega)\,|\hat{\lambda}|^{\alpha-2\omega}\operatorname{Tr}^\beta[Y^{\alpha-2\beta}(g)\vee I^\beta]. \end{aligned} \tag{4.12}$$

Just as before, in (4.12) we have made use of (2.14). Taking

$$Y^{\alpha-2\beta}(g)\vee I^\beta = [\mathbb{l}(\alpha,\beta)]^{-1}Y^{\alpha-2\beta}(g)*I^\beta \tag{4.13}$$

and appealing to

$$\operatorname{Tr}^\beta[Y^{\alpha-2\beta}(g)*I^\beta] = [\mathbb{k}(\alpha-2\beta,\beta)]^{-1}Y^{\alpha-2\beta}(g), \tag{4.14}$$

we conclude that

$$\begin{aligned} B^{\alpha[\beta]} &= \sum_{\omega=0}^{\beta}(-1)^\omega\mathbb{l}(\alpha,\omega)\mathbb{l}(\alpha-2\omega,\beta-\omega)\,\mathbb{k}(\alpha-2\beta,\beta-\omega)\,[\mathbb{l}(\alpha,\beta)\,\mathbb{k}(\alpha-2\beta,\beta)]^{-1}|\hat{\lambda}|^{\alpha-2\omega}Y^{\alpha-2\beta}(g) \\ &= (-1)^\beta 2^\beta\beta!\,Y^{\alpha-2\beta}(\hat{\lambda})\sum_{\omega=0}^{\beta}(-1)^{\beta+\omega}\frac{[2(\alpha-2\beta)+2\beta+1]!!}{2^\omega\omega!(\beta-\omega)![2(\alpha-2\beta)+2(\beta-\omega)+1]!!}\,z^{\beta-\omega}, \end{aligned} \tag{4.15}$$

where

$$Y^\alpha(\hat{\lambda}) := \langle \otimes^\alpha \hat{\lambda} \rangle, \quad z := \tfrac{1}{2}|\hat{\lambda}|^2. \tag{4.16}$$

If, as usual, we identify[24,25]

$$\sum_{\omega=0}^{\beta} (-1)^{\beta+\omega} \frac{(2\alpha+2\beta+1)!!}{2^\omega \omega!(\beta-\omega)![2\alpha+2(\beta-\omega)+1]!!} z^{\beta-\omega} \tag{4.17}$$

with the so-called Laguerre polynomials $L_\beta^{(\alpha+1/2)}(z)$, then we obtain

$$B^{\alpha[\beta]}(\hat{\lambda}) = (-1)^\beta 2^\beta \beta! Y^{\alpha-2\beta}(\hat{\lambda}) L_\beta^{(\alpha-2\beta+1/2)}(z), \tag{4.18}$$

and hence by placing (4.18) into (4.7) we find that the irreducible counterpart of (4.3), which fits in *together* and *with* Eqs. (3.21), is

$$f = f_0 \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} (-1)^\beta \mathbf{k}(\alpha,\beta) \frac{1}{\alpha!} \times b^{\alpha|\beta}(x,t) \circ Y^\alpha(\hat{\lambda}) L_\beta^{(\alpha+1/2)}(z). \tag{4.19}$$

Denoting by $\mathbb{R}^+$ the real interval $(0, +\infty)$ and by $\mathbb{K}$ the set of unit vectors, $\mathbb{K} := \{g \in \mathbb{E}: |g| = 1\}$, we see from[24,25]

$$\int_{\mathbb{R}^+} z^{\alpha+1/2} e^{-z} L_\beta^{(\alpha+1/2)}(z) L_\gamma^{(\alpha+1/2)}(z)\, dz = \frac{\sqrt{\pi}(2\alpha+2\beta+1)!!}{2^{\alpha+\beta+1}\beta!} \delta_{\beta\gamma} \tag{4.20}$$

and[2]

$$\int_{\mathbb{K}} Y^\alpha(g) \otimes Y^\nu(g)\, dg = \frac{4\pi\alpha!}{(2\alpha+1)!!} \delta_{\alpha\nu} E(\alpha|\alpha) \tag{4.21}$$

that the Laguerre–Ikenberry expansion coefficients $b^{\alpha|\beta} \in \mathrm{Ker}_\alpha$ Tr of the one-particle density f are the expectations of the corresponding Laguerre–Ikenberry functions $Y^\alpha(\hat{\lambda}) L_\beta^{(\alpha+1/2)}(z)$:

$$b^{\alpha|\beta} = (-1)^\beta \frac{(2\beta)!!}{\rho} \int_{\mathbb{E}} Y^\alpha(\hat{\lambda}) L_\beta^{(\alpha+1/2)}(z) f\, d\lambda. \tag{4.22}$$

The quantity $\delta_{\alpha\beta}$ appearing in Eqs. (4.20) and (4.21) denotes the Kronecker delta and $E(\alpha|\alpha)$ is, as we know, the natural projection defined by (2.20).

In order to facilitate conversions between $M^{\alpha|\beta}$ and $b^{\alpha|\beta}$, we record here the algebraic equation[24,25] for z^β:

$$z^\beta = \frac{1}{2^\beta} \sum_{\omega=0}^{\beta} (-1)^\omega \binom{\beta}{\omega} (2\omega)!! \times [\mathbf{k}(\alpha,\beta)]^{-1} \mathbf{k}(\alpha,\omega) L_\omega^{(\alpha+1/2)}(z). \tag{4.23}$$

This equality and the decomposition (4.17) for $L_\beta^{(\alpha+1/2)}(z)$ allow us to write

$$M^{\alpha|\beta} = \rho (k_B T)^{\beta+\alpha/2} \sum_{\omega=0}^{\beta} \binom{\beta}{\omega} [\mathbf{k}(\alpha,\beta)]^{-1} \mathbf{k}(\alpha,\omega) b^{\alpha|\omega}, \tag{4.24a}$$

$$(-1)^\beta b^{\alpha|\beta} = \frac{1}{\rho} (k_B T)^{-\alpha/2} \sum_{\omega=0}^{\beta} (-1)^\omega \binom{\beta}{\omega} \times (k_B T)^{-\omega} [\mathbf{k}(\alpha,\beta)]^{-1} \mathbf{k}(\alpha,\omega) M^{\alpha|\omega}. \tag{4.24b}$$

In summary, the expansion (4.19) confirms the importance of Laguerre polynomials in the kinetic theory and, at the same time, shows that spherical harmonics may be replaced by those of Ikenberry. According to Coope and Snider,[2] the relation of $Y^\alpha(\hat{\lambda}) = |\hat{\lambda}|^\alpha Y^\alpha(g)$ to "the properties of three-dimensional space is more apparent than with spherical tensors, where one axis is arbitrarily distinguished, and, furthermore, they can often be more easily handled, i.e., without the necessity of tables of numerical coefficients." However, in regard to the Coope–Snider statement to demonstrate that one axis has to be arbitrarily distinguished, the reader should note our comment in Ref. 30.

Johnston[10] has described a general procedure for expressing f in terms of its tensorial irreducible expansion coefficients $f_\alpha \in \mathrm{Ker}_\alpha$ Tr,

$$f(\lambda,x,t) = \sum_{\alpha=0}^{\infty} f_\alpha(|\overline{\lambda}|,x,t) \circ Y^\alpha(g), \tag{4.25}$$

and also derived the system of balance equations for them. It is important to observe that Eq. (12) in Ref. 10 and Eq. (3.21) in the present work are not completely equivalent. First, from the very beginning Johnston considers only such molecular densities f as may be represented in the *mean* by a series of Ikenberry's coefficients f_α, and second, he makes no use of the expansion

$$f_\alpha = f_0 \sum_{\beta=0}^{\infty} (-1)^\beta \mathbf{k}(\alpha,\beta) \frac{1}{\alpha!} L_\beta^{(\alpha+1/2)} b^{\alpha|\beta} \tag{4.26}$$

and the transformation rule (4.24b). Had we chosen the suitable adjustment of Johnston's method, substituting (4.26) into Eq. (12) in Ref. 10, we would have been able to see clearly how to obtain the equations of transfer for $b^{\alpha|\beta}$. Of course, because of (4.25) and (4.26), in the procedure just outlined the one-particle density f must possess the Laguerre–Ikenberry moments $b^{\alpha|\beta}$ of all orders. In contrast, Eq. (3.21) is valid, as its derivation in Sec. III shows, for a much broader class of functions f, indeed for ones which need possess only a finite number of the direct moments $M^{\alpha|\beta}$.

B. Irreducible variant of Grad's method of truncation

Before proposing and exploiting the irreducible variant of Grad's truncation scheme, we shall complete our survey of results that conform with those on one-dimensional classical gases.[31] To this end let us consider the following standard expression for the entropy density h:

$$h = -k_B \int_{\mathbb{E}} f \ln(\mathbb{C} f)\, d\lambda, \tag{4.27}$$

$\mathbb{C}$ being a constant directly associated with that of Planck; the exact form of this constant is of no importance in our further investigations. We can arrive at an approximate formula for an h corresponding to

$$f = f_0(1+\epsilon), \tag{4.28a}$$

$$\epsilon := \sum_{\alpha+2\beta \geqslant 2} (-1)^\beta \mathbf{k}(\alpha,\beta) \times \frac{1}{\alpha!} b^{\alpha|\beta}(x,t) \circ Y^\alpha(\hat{\lambda}) L_\beta^{(\alpha+1/2)}(z), \tag{4.28b}$$

which amounts to determining h from $b^{\alpha|\beta}$. In the neighborhood of local equilibrium, if instead of the logarithm $\ln(1+\epsilon)$ we use the first two terms in its Taylor expansion $\epsilon - \frac{1}{2}\epsilon^2 + \cdots$, we obtain for $h - h_0$

$$h - h_0 \cong -\frac{1}{2}\rho k_B \sum_{\alpha+2\beta\geqslant 2} \frac{1}{(\alpha+2\beta)!} \times \mathbf{l}(\alpha+2\beta,\beta)\,\mathbf{k}(\alpha,\beta)\,b^{\alpha|\beta}\circ b^{\alpha|\beta} \leqslant 0, \tag{4.29}$$

where

$$h_0 := -k_B \int_{\mathbb{E}} f_0 \ln(\mathbb{C} f_0)\,d\lambda$$
$$= \frac{3}{2}\rho k_B - \rho k_B \ln\left[\mathbb{C}\rho\left(\frac{1}{2\pi k_B T}\right)^{3/2}\right], \tag{4.30}$$

and the series (4.29) converges, because we suppose that $\exp(\frac{1}{4}|\hat{\lambda}|^2) f$ be square integrable over $\mathbb{E}$. Interpreting (4.29), the choice of the Laguerre–Ikenberry expansion coefficients $b^{\alpha|\beta}$ of f always *diagonalizes* the largest contribution to $h - h_0$ and the right-hand inequality in (4.29) suggests that f_0 gives h the greatest value it can attain for all molecular densities f corresponding to the same gross condition ($b^0 = 1$, $b^1 = \mathrm{Tr}\, b^2 = 0$). Obvious, trivial, and well-known as these observations are,[3,4] they were[32] and will be[33,34] of great, almost decisive import in our extensions of the range of validity of Grad's ideas[3,4] to quasiparticle gaseous systems.

Given Boltzmann's operator J for any spherically symmetrical model, the balance equation (3.21) in which $\partial_t M^{\alpha|\beta}$ occurs contains a contribution from the collision integral

$$P^{\alpha|\beta} = \int_{\mathbb{E}} |\bar{\lambda}|^{2\beta} Y^\alpha(\bar{\lambda}) J(f)\,d\lambda. \tag{4.31}$$

We insert the expansion (4.19) into (4.31) and make use of (4.24b). The result is an infinite sequence of differential equations in the symmetric traceless moments $M^{\alpha|\beta}$ of f, which in general cannot be solved without some truncation procedure. If we are willing to accept Grad's method, in the system of equations up to $\partial_t M^{\alpha|\beta}$, $\alpha + 2\beta = r$ ($r = 2,3,\ldots$) we may set all Laguerre–Ikenberry coefficients $b^{\alpha/\beta}$ for which $\alpha + 2\beta > r$ equal to zero. Then, due to (4.24b), we obtain

$$\sum_{\omega=0}^{\beta} (-1)^\omega \binom{\beta}{\omega} (k_B T)^{-\omega} [\mathbf{k}(\alpha,\beta)]^{-1} \mathbf{k}(\alpha,\omega) M^{\alpha|\omega} = 0, \tag{4.32a}$$

$$\alpha + 2\beta > r. \tag{4.32b}$$

On the left differential side, only the final equations for $\partial_t M^{\alpha|\beta}$, $\alpha + 2\beta = r$ themselves are altered; in particular, using (4.32) in which $\alpha + 2\beta = r + 1$,

$$M^{\alpha+1|\beta} = \sum_{\omega=0}^{\beta-1} (-1)^{\beta-\omega+1} \binom{\beta}{\omega} (k_B T)^{\beta-\omega} \times [\mathbf{k}(\alpha+1,\beta)]^{-1} \mathbf{k}(\alpha+1,\omega) M^{\alpha+1|\omega}, \tag{4.33a}$$

$$M^{\alpha-1|\beta+1} = \sum_{\omega=0}^{\beta} (-1)^{\beta-\omega} \binom{\beta+1}{\omega} (k_B T)^{\beta-\omega+1} \times [\mathbf{k}(\alpha-1,\beta+1)]^{-1} \mathbf{k}(\alpha-1,\omega) M^{\alpha-1|\omega}, \tag{4.33b}$$

$$\alpha + 2\beta = r. \tag{4.33c}$$

On the right collision side, we must replace (4.19) by

$$f = f_0 \sum_{\alpha+2\beta\leqslant r} (-1)^\beta \mathbf{k}(\alpha,\beta) \frac{1}{\alpha!} b^{\alpha|\beta}(x,t) \circ Y^\alpha(\hat{\lambda})\, L_\beta^{(\alpha+1/2)}(z) \tag{4.34}$$

in (4.31) except in the case of a gas of Maxwellian molecules where the exact expression (4.31) for $P^{\alpha|\beta}$, $\alpha + 2\beta \leqslant r$ already has the "proper" form depending on $M^{\nu|\gamma}$, $\nu + 2\gamma \leqslant \alpha + 2\beta \leqslant r$ alone.

The method of Müller and Liu, described in detail in Ref. 35 and based upon the effect of diagonalization of an approximate formula for an h,[31,32] allows one to establish the precise sense of the statement that the system of differential equations for $M^{\alpha|\beta}$, $\alpha + 2\beta \leqslant r$ just obtained, if supplemented by the principle of conservation of momentum (3.7), describes a certain thermomechanical process.[37]

V. THE STRUCTURE OF COLLISION INTEGRALS FOR A GAS OF MAXWELLIAN MOLECULES

From[23]

$$\otimes^\alpha \hat{\lambda} = \sum_{\omega=0}^{[\alpha/2]} \mathbf{l}(\alpha,\omega)\,[B^{\alpha-2\omega}(\hat{\lambda}) \vee \mathbf{I}^\omega] \tag{5.1}$$

and the definitions (4.2c) and (3.6) it follows that

$$P^\alpha = \rho(k_B T)^{\alpha/2} \sum_{\omega=0}^{[\alpha/2]} \mathbf{l}(\alpha,\omega)\,(Q^{\alpha-2\omega} \vee I^\omega), \tag{5.2}$$

where

$$Q^\alpha := \frac{1}{\rho}\int_{\mathbb{E}} B^\alpha(\hat{\lambda}) J(f)\,d\lambda. \tag{5.3}$$

In a gas of Maxwellian molecules, the collision integral Q^α, $\alpha \geqslant 2$, is a linear combination of the Hermite moments of degree α plus a bilinear combination of the Hermite moments of lower degree, the sum of the degrees in each term being α:

$$Q^\alpha = Q_L^\alpha + Q_N^\alpha, \quad \alpha \geqslant 2, \tag{5.4a}$$

$$Q_L^\alpha = \rho \sum_{\omega=0}^{[\alpha/2]} \mathbb{A}_\omega^\alpha\, \hat{b}^{\alpha[\omega]} \vee I^\omega, \tag{5.4b}$$

$$Q_N^\alpha = \cdots, \tag{5.4c}$$

$$\hat{b}^{\alpha[\omega]} = \mathrm{Tr}^\omega b^\alpha. \tag{5.4d}$$

Here the scalar coefficients $\mathbb{A}_\omega^\alpha$ are functions of the molecular parameters (constants) alone and are independent of f. In addition, the response of a gas of Maxwellian molecules is exactly the same for P^α. That is, if we replace b^β by M^β in $Q^\alpha = \tilde{Q}^\alpha(b^\beta)$, we obtain $P^\alpha = \tilde{Q}^\alpha(M^\beta)$. Bearing this[22] in mind, the major breakthrough came from Truesdell and Muncaster,[11] who in their theorem, henceforth referred to as the Truesdell–Muncaster theorem, not only gave the explicit formula for $\mathbb{A}_\omega^\alpha$ but also actually succeeded in calculating Q_N^α. However, since their result is extremely complicated, the missing right-hand-side of (5.4c) will not be given here; for more details, see Eqs. (XVI.20) and (XVI.22) on p. 242 in Ref. 11.

Now, regardless of (5.4), the evaluation of $Q^{\alpha|\beta} = \langle \mathrm{Tr}^\beta Q^{\alpha+2\beta} \rangle$, or of $P^{\alpha|\beta}$, was originally proposed by

Ikenberry (and Truesdell).[7] One of the most essential aspects of Ikenberry's theorem comes when we recall his result for $Q_L^{\alpha|\beta} = \langle \mathrm{Tr}^\beta Q_L^{\alpha+2\beta} \rangle$ of the form

$$Q_L^{\alpha|\beta} = \rho\, \mathbb{B}_\beta^\alpha b^{\alpha|\beta}, \tag{5.5}$$

where $\mathbb{B}_\beta^\alpha$ stands for the scalar coefficient first determined in Ref. 7. According to Truesdell and Muncaster,[11] the fact that $Q_L^{\alpha|\beta}$ is a scalar multiple of the symmetric traceless "moment" $b^{\alpha|\beta}$ alone leads to results which would be difficult to extract from (5.4b), although that formula certainly implies them. Since the hope for setting up an undeniable correspondence between both theorems has surprisingly not materialized as yet, here we shall examine this question more closely.

Thus let us consider (5.4b) as a starting point. Then, due to the decomposition

$$\hat{b}^{\alpha[\omega]} = \sum_{\gamma=0}^{[(\alpha-2\omega)/2]} \mathbb{l}(\alpha-2\omega,\gamma)\mathbb{k}(\alpha-2\omega-2\gamma,\gamma) \times b^{\alpha[\omega+\gamma]} \vee I^\gamma \tag{5.6}$$

directly resulting from (2.16) and (2.8a), as well as in virtue of the properties (2.14) and (2.10b), we are justified in doing the following sequence of transformations:

$$\begin{aligned} Q_L^{\alpha[\beta]} = \langle \mathrm{Tr}^\beta Q_L^\alpha \rangle &= \rho \sum_{\omega=0}^{[\alpha/2]} \sum_{\gamma=0}^{[(\alpha-2\omega)/2]} \mathbb{A}_\omega^\alpha\, \mathbb{l}(\alpha-2\omega,\gamma)\mathbb{k}(\alpha-2\omega-2\gamma,\gamma) \langle \mathrm{Tr}^\beta (b^{\alpha[\omega+\gamma]} \vee I^{\omega+\gamma}) \rangle \\ &= \rho \sum_{\omega=0}^{\beta} \mathbb{A}_\omega^\alpha \mathbb{l}(\alpha-2\omega,\beta-\omega)\mathbb{k}(\alpha-2\beta,\beta-\omega) \langle \mathrm{Tr}^\beta (b^{\alpha[\beta]} \vee I^\beta) \rangle \\ &= \rho [\mathbb{l}(\alpha,\beta)\mathbb{k}(\alpha-2\beta,\beta)]^{-1} \sum_{\omega=0}^{\beta} \mathbb{A}_\omega^\alpha\, \mathbb{l}(\alpha-2\omega),\beta-\omega)\mathbb{k}(\alpha-2\beta,\beta-\omega) b^{\alpha[\beta]}. \end{aligned} \tag{5.7}$$

By making use of $Q_L^{\alpha|\beta} = Q_L^{\alpha+2\beta[\beta]}$, we get a relation for $Q_L^{\alpha|\beta}$ which may easily be identified with Ikenberry's formula (5.5), provided that the coefficient $\mathbb{B}_\beta^\alpha$ is given by

$$\mathbb{B}_\beta^\alpha = [\mathbb{l}(\alpha+2\beta,\beta)\mathbb{k}(\alpha,\beta)]^{-1} \sum_{\omega=0}^{\beta} \mathbb{A}_\omega^{\alpha+2\beta} \times \mathbb{l}(\alpha+2\beta-2\omega,\beta-\omega)\mathbb{k}(\alpha,\beta-\omega). \tag{5.8}$$

Hence the derivation of (5.5) from (5.4b) is complete. Curiously enough, the expression (5.8) relating $\mathbb{B}_\beta^\alpha$ to $\mathbb{A}_\omega^{\alpha+2\beta}$ has never been established before.

Following Ikenberry and Truesdell,[7] we may prove that $\mathbb{B}_\beta^\alpha < 0$. Finally, making use of (5.2), (5.5), and (4.24b), and appealing to the universal method of Sec. IV A, we are able to arrive at

$$P^{\alpha|\beta} = \rho \sum_{\omega=0}^{\beta} \sum_{\nu=\omega}^{\beta} (k_B T)^{\beta-\omega} (-1)^{\omega+\nu} \binom{\beta}{\nu}\binom{\nu}{\omega} \times \mathbb{B}_\nu^\alpha [\mathbb{k}(\alpha,\beta)]^{-1} \mathbb{k}(\alpha,\omega) M^{\alpha|\omega} + \cdots, \tag{5.9}$$

where the ellipses represents that part of the collision integral $P^{\alpha|\beta}$ which in the linear approximation is considered to be negligible.

VI. FINAL REMARKS

A. General observations

It may seem odd that we have not yet confronted the routine question, for what reason we should be appreciative of having a completely irreducible tensor description of three-dimensional, classical, moderately rarefied, simple, monatomic gases. Here, quite intentionally, we shall not evaluate our postulational basis by a mere listing of the types of predictions of the theory in describing physically meaningful facts. Since the present approach covers very traditional ground, almost all of them have already been convincingly and excellently summarized.[11] Quite apart from the embarrassing observation, whether or not the irreducible tensor description of classical gases can culminate in real physical results, i.e., in numbers, from the very beginning we offer this work as a partial and certainly incomplete reply to the following Truesdell–Muncaster statement on p. XIX in their monograph[11]: "..., many of the analyses contain important gaps, ... Our first purpose is to uncover these gaps and to illuminate them as challenges to future research by mathematicians."

Thus, for the time being, this paper should be considered as an end in itself. However, part of the usefulness of the formalism is that it enables us in the next two papers,[33,34] henceforth referred to as Parts II and III, to extend almost "effortlessly" to quasiparticle gaseous systems a vast quantity of conclusions originally thought limited to a gas of classical molecules.

B. Comments regarding quasiparticle gases

The essential point we wish to stress here once more is that the severe restrictions on the types of gaseous systems considered in this work are not basic limitations on the generality of our ideas, but are adopted merely for pedagogical purposes.

Let us suppose that a nonclassical gas composed of phonons, magnons, rotons, etc. is adequately described by the Boltzmann-Peierls equation.[36] Then, by the introduction of the purely mathematical set of functions f satisfying Condition I of Sec. VI A in Part II, a more subtle but very similar theory can be proposed. In consequence of applying Condition I, this generalized theory rests primarily on the effect of diagonalization of an approximate formula for an h and provides in essence the only real key in establishing a strict quasiparticle analog of the irreducible variant of Grad's expansion of f in terms of Hermite polynomials.

As useful as the characterization of the nonequilibrium occupation probability f of quasiparticle states by its reducible moments has proved to be (in Gurevich's monograph,[36] for instance), it must clearly be conceded that Condition I is capable of yielding an irreducible series representation of f which, *in contrast with classical gases*, may later demonstrate to have no reducible counterpart at all. Thus the attempt to invent from the outset in Parts II and III the irreducible tensor description is *not* only an academic game, intellectually challenging but of no import.

ACKNOWLEDGMENT

We are deeply indebted to the referee for both drawing our attention to the work of T. W. Johnston and providing very useful comments concerning the relationship between spherical harmonics and those of Ikenberry.

APPENDIX A: DEFINITIONS OF Π, $\mathrm{Tr}_{(\beta,\nu)}$, $M^\alpha \circ M^\beta$, I, and ∇

Let $\mathbb{E}$ be a three-dimensional Euclidean vector space. Choose an orthonormal basis $\{e_1,e_2,e_3\}$ in $\mathbb{E}$ and set

$$e_{r_1\cdots r_\alpha} := e_{r_1} \otimes \cdots \otimes e_{r_\alpha}. \tag{A1}$$

(1) The action of the symmetrizer Π on a tensor M^α of degree α ($\alpha \geqslant 2$) is given by

$$\Pi M^\alpha := \sum_{r_1\cdots r_\alpha = 1}^{3} M^\alpha_{(r_1\cdots r_\alpha)} e_{r_1\cdots r_\alpha}, \tag{A2}$$

where the coefficients $M^\alpha_{r_1\cdots r_\alpha}$ are components of M^α with respect to the basis $\{e_1,e_2,e_3\}$ and parentheses enclosing a set of α indices represent symmetrization of these indices, i.e., the sum over $\alpha!$ permutations of the indices, divided by $\alpha!$ Moreover, we define Π to be the identity on $\mathbb{E}$ and $\mathbb{R}$,

$$\Pi M^1 := M^1, \quad \Pi M^0 := M^0. \tag{A3}$$

Obviously, $\Pi: \mathbb{E}^\alpha \Rightarrow \mathbb{E}^\alpha_s$ is a projection operator:

$$\Pi^2 = \Pi. \tag{A4}$$

(2) Since the trace operator with respect to the pair (β,ν), denoted by $\mathrm{Tr}_{(\beta,\nu)}$, determines a linear map,

$$\underset{(\beta,\nu)}{\mathrm{Tr}} : \mathbb{E}^\alpha \Rightarrow \mathbb{E}^{\alpha-2}, \quad \alpha \geqslant 2, \quad \alpha \geqslant \beta \geqslant 1, \quad \alpha \geqslant \nu \geqslant 1, \quad \beta \neq \nu,$$

we only need to consider how it operates on the generator $e_{r_1\cdots r_\alpha}$,

$$\underset{(\beta,\nu)}{\mathrm{Tr}}\, e_{r_1\cdots r_\alpha} := \delta_{r_\beta r_\nu} e_{r_1\cdots \hat{r}_\beta \cdots \hat{r}_\nu \cdots r_\alpha}, \tag{A5}$$

where $\delta_{r_\beta r_\nu}$ denotes the Kronecker delta and the hat over r_β and r_ν tells us that e_{r_β} and e_{r_ν} do not appear in $e_{r_1\cdots r_\alpha}$.

(3) Suppose that $\alpha \geqslant 1$, $\beta \geqslant 1$, and $\nu := \min(\alpha,\beta)$. Then the bilinear operator

$$\hat{\Gamma}: \mathbb{E}^\alpha \times \mathbb{E}^\beta \Rightarrow \mathbb{E}^{\alpha+\beta-2\nu},$$

for which we write

$$M^\alpha \circ M^\beta := \hat{\Gamma}(M^\alpha, M^\beta), \tag{A6}$$

is uniquely determined by its action upon the generators $e_{r_1\cdots r_\alpha}$ and $e_{s_1\cdots s_\beta}$:

$$e_{r_1\cdots r_\alpha} \circ e_{s_1\cdots s_\beta} := \begin{cases} \delta_{r_1 s_1} \cdots \delta_{r_\beta s_\beta} e_{r_{\beta+1}\cdots r_\alpha} & (\alpha > \beta), \\ \delta_{r_1 s_1} \cdots \delta_{r_\beta s_\beta} & (\alpha = \beta), \\ \delta_{r_1 s_1} \cdots \delta_{r_\alpha s_\alpha} e_{s_{\alpha+1}\cdots s_\beta} & (\beta > \alpha). \end{cases} \tag{A7}$$

In addition, we set

$$M^\alpha \circ M^0 := M^0 \circ M^\alpha := M^0 M^\alpha, \quad \alpha \geqslant 0. \tag{A8}$$

It is easily checked that

$$M^\alpha \circ M^\beta = M^\beta \circ M^\alpha. \tag{A9}$$

We call the tensor $M^\alpha \circ M^\beta$ the inner product of M^α and M^β.

(4) The contravariant metric (unit) tensor I of the Euclidean vector space $\mathbb{E}$ is given by

$$I := \sum_{r=1}^{3} e_r \otimes e_r \in \mathbb{E}^2_s. \tag{A10}$$

(5) Let M^α be a differentiable tensor field in $\mathbb{E}$. Then the action of ∇ on a tensor field M^α of degree $\alpha \geqslant 0$ is defined by

$$\nabla M^\alpha|_x := \sum_{j r_1 \cdots r_\alpha = 1}^{3} \nabla_j M^\alpha_{r_1\cdots r_\alpha} e_{j r_1 \cdots r_\alpha} \in \mathbb{E}^{\alpha+1}, \tag{A11}$$

where

$$\nabla_j := \frac{\partial}{\partial x_j}, \quad x = \sum_{j=1}^{3} x_j e_j. \tag{A12}$$

APPENDIX B: PROOF OF THE IMPLICATION $(\beta>0) \Rightarrow (\langle M^\alpha * I^\beta \rangle = \mathrm{l}(\alpha+2\beta,\beta)\langle M^\alpha \vee I^\beta \rangle = 0)$

Let $\{\hat{e}_\alpha(r);\ r = 1,2,\ldots,2\alpha+1\}$ be a basis of $\mathrm{Ker}_\alpha\,\mathrm{Tr}$. Then the natural projection $E(\alpha|\alpha) \in \mathbb{E}^{2\alpha}$ of $\mathbb{E}^\alpha$ onto the irreducible subspace $\mathrm{Ker}_\alpha\,\mathrm{Tr}$ of symmetric traceless tensors can uniquely be written as a sum

$$E(\alpha|\alpha) = \sum_{r=1}^{2\alpha+1} \hat{e}_\alpha(r) \otimes \hat{e}_\alpha(r). \tag{B1}$$

Hence, on recalling (2.20), we get

$$\langle \Pi M^\alpha \rangle = E(\alpha|\alpha) \circ M^\alpha = \sum_{r=1}^{2\alpha+1} [\hat{e}_\alpha(r) \circ M^\alpha] \hat{e}_\alpha(r). \tag{B2}$$

From (B2) we conclude that, for $\beta > 0$,

$$\begin{aligned}
&[\mathrm{l}(\alpha+2\beta,\beta)]^{-1} \langle M^\alpha * I^\beta \rangle \\
&= \langle M^\alpha \vee I^\beta \rangle \\
&= \langle \Pi(M^\alpha \otimes I^\beta) \rangle \\
&= \sum_{r=1}^{2\alpha+4\beta+1} [\hat{e}_{\alpha+2\beta}(r) \circ (M^\alpha \otimes I^\beta)] \hat{e}_{\alpha+2\beta(r)} \\
&= \sum_{r=1}^{2\alpha+4\beta+1} \{[\mathrm{Tr}^\beta \hat{e}_{\alpha+2\beta}(r)] \circ M^\alpha\} \hat{e}_{\alpha+2\beta}(r) = 0.
\end{aligned} \tag{B3}$$

Thus the implication holds.

APPENDIX C: PROOF OF (3.9a) AND (3.9b) FOR $\beta = 1$

Recalling the notation of Appendix A and using Einstein's summation convention, we obtain

$$
\begin{aligned}
\alpha \operatorname{Tr}(L \cup M^{\alpha}) &= \alpha \operatorname*{Tr}_{(1,2)} (L \cup M^{\alpha}) = \sum_{r_1 \cdots r_\alpha = 1}^{3} \alpha L_{k(r_1} M^{\alpha}_{r_2 \cdots r_\alpha)k} \operatorname*{Tr}_{(1,2)} e_{r_1 \cdots r_\alpha} \\
&= \sum_{r_1 \cdots r_\alpha = 1}^{3} \alpha L_{k(r_1} M^{\alpha}_{r_2 \cdots r_\alpha)k} \delta_{r_1 r_2} e_{r_3 \cdots r_\alpha} \\
&= \sum_{r_1 \cdots r_\alpha = 1}^{3} (\alpha - 2) L_{k(r_3} M^{\alpha}_{r_4 \cdots r_\alpha) k r_1 r_2} \delta_{r_1 r_2} e_{r_3 \cdots r_\alpha} + \sum_{r_1 \cdots r_\alpha = 1}^{3} 2 L_{k r_1} M^{\alpha}_{k r_2 r_3 \cdots r_\alpha} \delta_{r_1 r_2} e_{r_3 \cdots r_\alpha} \\
&= \sum_{r_3 \cdots r_\alpha = 1}^{3} (\alpha - 2) L_{k(r_3} M^{\alpha}_{r_4 \cdots r_\alpha) k p p} e_{r_3 \cdots r_\alpha} + \sum_{r_3 \cdots r_\alpha = 1}^{3} 2 L_{kp} M^{\alpha}_{k p r_3 \cdots r_\alpha} e_{r_3 \cdots r_\alpha} \\
&= (\alpha - 2) L \cup \widehat{M}^{\alpha[1]} + 2 L \circ \widehat{M}^{\alpha[0]}.
\end{aligned}
\tag{C1}
$$

Hence the proof of (3.9a) for $\beta = 1$ is complete. Of course, exactly the same view applies to (3.9b) for $\beta = 1$.

APPENDIX D: PROOF OF THE EQUALITY (3.17)

Our immediate purpose here is to sketch the proof of Eq. (3.17) for the special case in which $\alpha - 2\beta - 4 \geqslant 0$ and $\beta \geqslant 1$. It is readily seen then that (3.17) holds also for $\alpha - 2\beta \geqslant 0$.

Before we can proceed, we must begin with the two auxiliary equalities of the form

$$M^2 \circ (M^\alpha * I) = (\operatorname{Tr} \Pi M^2) M^\alpha + \alpha M^2 \cup M^\alpha + \alpha M^\alpha \cup M^2 + (M^2 \circ M^\alpha) * I, \tag{D1}$$

$$M^2 \circ (M^\alpha * I^2) = (\alpha + 1)(\alpha + 2) M^2 \vee M^\alpha + (\operatorname{Tr} \Pi M^2) M^\alpha * I + \alpha (M^2 \cup M^\alpha) * I + \alpha (M^\alpha \cup M^2) * I + (M^2 \circ M^\alpha) * I^2, \tag{D2}$$

where $M^2 \in \mathbb{E}^2$ and $M^\alpha \in \mathbb{E}_s^\alpha$ $(\alpha \geqslant 2)$.

By making use of the decomposition

$$\widehat{M}^{\alpha[\beta - 1]} = \sum_{\omega = 0}^{[(\alpha - 2\beta + 2)/2]} \mathbf{k}(\alpha - 2\beta - 2\omega + 2, \omega) M^{\alpha[\beta + \omega - 1]} * I^\omega \tag{D3}$$

and Eq. (2.14), it follows from (D1), (D2), (2.10a), (3.20), (2.17), and (2.18) that

$$
\begin{aligned}
\langle L \circ \widehat{M}^{\alpha[\beta - 1]} \rangle &= \sum_{\omega = 0}^{[(\alpha - 2\beta + 2)/2]} \mathbf{k}(\alpha - 2\beta - 2\omega + 2, \omega) \langle L \circ (M^{\alpha[\beta + \omega - 1]} * I^\omega) \rangle \\
&= \mathbf{k}(\alpha - 2\beta + 2, 0) L \circ M^{\alpha[\beta - 1]} + \mathbf{k}(\alpha - 2\beta, 1) \langle L \circ (M^{\alpha[\beta]} * I) \rangle + \mathbf{k}(\alpha - 2\beta - 2, 2) \langle L \circ (M^{\alpha[\beta + 1]} * I^2) \rangle \\
&= L \circ M^{\alpha[\beta - 1]} + \frac{1}{2\alpha - 4\beta + 3} S M^{\alpha[\beta]} + \frac{\alpha - 2\beta}{2\alpha - 4\beta + 3} L \cap M^{\alpha[\beta]} + \frac{\alpha - 2\beta}{2\alpha - 4\beta + 3} M^{\alpha[\beta]} \cap L \\
&\quad + \frac{(\alpha - 2\beta)(\alpha - 2\beta - 1)}{(2\alpha - 4\beta - 1)(2\alpha - 4\beta + 1)} L \wedge M^{\alpha[\beta + 1]}.
\end{aligned}
\tag{D4}
$$

Hence the proof of Eq. (3.17) for the $\alpha - 2\beta - 4 \geqslant 0$ $(\beta \geqslant 1)$ case is complete.

[1]W. Greub, *Multilinear Algebra* (Springer, New York, 1978).
[2]J. A. R. Coope and R. F. Snider, J. Math. Phys. **11**, 1003 (1970).
[3]H. Grad, Commun. Pure Appl. Math. **2**, 331 (1949).
[4]H. Grad, *Principles of the Kinetic Theory of Gases, Handbuch der Physik*, edited by S. Flugge (Springer, Berlin, 1958), Vol. XII, p. 205.
[5]E. Ikenberry, Am. Math. Monthly **62**, 719 (1955).
[6]E. Ikenberry, J. Math Anal. Appl. **3**, 355 (1961).
[7]E. Ikenberry and C. Truesdell, J. Rational Mech. Anal. **5**, 1 (1956).
[8]M. Hamermesh, *Group Theory* (Pergamon, London, 1962).
[9]C. Truesdell, J. Rational Mech. Anal. **5**, 55 (1956).
[10]T. W. Johnston, J. Math. Phys. **7**, 1453 (1966).
[11]C. Truesdell and R. G. Muncaster, *Fundamentals of Maxwell's Kinetic Theory of a Simple Monatomic Gas* (Academic, New York, 1980).
[12]K. Kumar, Ann. Phys. (NY) **37**, 113 (1966).
[13]K. Suchy, Ergeb. Exakten Naturwiss. **35**, 103 (1964).
[14]K. Suchy, *Transport Coefficients and Collision Frequencies for Aeronomic Plasmas, Handbuch der Physik*, edited by S. Flugge (Springer, Berlin, 1984), Vol. 49/7, p. 57.
[15]U. Weinert, Arch. Rat. Mech. Anal. **74**, 165 (1980).
[16]U. Weinert, Phys. Rep. **91**, 297 (1982).
[17]U. Weinert, S. L. Lin, and E. A. Mason, Phys. Rev. A **22**, 2262 (1980).
[18]J. Meyer and J. Schröter, Arch. Rat. Mech. Anal. **76**, 193 (1981).
[19]J. Meyer and J. Schröter, J. Stat. Phys. **32**, 53 (1983).
[20]J. Schröter, J. Non-Equilib. Thermodyn. **11**, 315 (1986).
[21]M. H. Ernst, Phys. Lett. A **69**, 390 (1979).
[22]E. M. Hendriks and T. M. Nieuwenhuizen, J. Stat. Phys. **29**, 591 (1982).
[23]H. Grad, Commun. Pure. Appl. Math. **2**, 325 (1949).
[24]G. Szegö, *Orthogonal Polynomials* (Am. Math. Soc., Providence, RI, 1939).
[25]G. Sansone, *Orthogonal Functions* (Interscience, New York, 1959).
[26]W. Loose and S. Hess, "The nonlinear flow behavior of gases: Kinetic theory and molecular dynamics," in *Proceedings of the ISIMM Sympo-*

sium on Kinetic Theory and Extended Thermodynamics, edited by I. Müller and T. Ruggeri (Pitagora, Bologna, 1987), p. 217; see also Phys. Rev. A **37**, 2099 (1988).
[27]Z. Banach, Physica A **129**, 95 (1984).
[28]C. G. Speziale, Int. J. Engng. Sci. **26**, 211 (1988).
[29]The quantities k_B and T, which we introduce only formally in this section, can be termed Boltzmann's constant and the absolute temperature, respectively.
[30]Insofar as the collision process rather than the system of irreducible equations of transfer is concerned, this statement is somewhat dubious, because the axis under consideration points in the direction of the (relative) velocity after a binary encounter; see, e.g., Suchy, Refs. 13 and 14, and Weinert, Ref. 16.
[31]Z. Banach, J. Stat. Phys. **48**, 813 (1987).
[32]Z. Banach, to be published in Arch. Mech.
[33]Z. Banach and S. Piekarski, J. Math. Phys. **30**, 1816 (1989).
[34]Z. Banach and S. Piekarski, J. Math. Phys. **30**, 1826 (1989).
[35]I.-S. Liu and I. Müller, Arch. Rat. Mech. Anal. **83**, 285 (1983).
[36]V. L. Gurevich, *Kinetika Fononnykh Sistem* (Nauka, Moscow, 1980).
[37]Z. Banach, Physica A **145**, 105 (1987).

Irreducible tensor description. II. A quasiparticle gas

Zbigniew Banach
Institute of Fundamental Technological Research, Department of Fluid Mechanics, Polish Academy of Sciences, Swietokrzyska 21, 00-049 Warsaw, Poland
Slawomir Piekarski
Institute of Fundamental Technological Research, Department of Acoustoelectronics, Polish Academy of Sciences, Swietokryska 21, 00-049 Warsaw, Poland

(Received 7 June 1988; accepted for publication 15 March 1989)

Let E be a three-dimensional Euclidean vector space and assume that $\hbar\Omega(k)$ is a quasiparticle energy in the mode $k\in\mathrm{E}$; thus k is a wave vector. Within the framework of the Boltzmann-Peierls equation and a broad class of isotropic dispersion relations $[\Omega(k)\Rightarrow\Omega(\not k),\not k:=|k|]$, the exact system of irreducible equations of transfer for the symmetric traceless moments of the distribution function f is derived and the range of validity of Grad's moment procedure is extended to the case of quasiparticle gases. Thus not without reason, an expansion with respect to k of the one-particle density f around the local Bose–Einstein occupation probability f_0 in terms of the appropriately chosen Tchebychef functions $A_\beta(z;\Theta)$ and Ikenberry's harmonics $Y^\alpha(g)$ is carefully recognized. Also, the importance of the Tchebychef basis $\{A_\beta;\beta=0,1,...\}$, both in any serious analysis of equilibrium fluctuations and in exploiting the Chapman–Enskog procedure, is clearly established.

I. INTRODUCTION

Given a gas consisting of quasiparticles (phonons,[1] magnons,[2] rotons,[1] etc.), the main object of this paper is both to solve some of the basic tensorial problems generated by the Boltzmann–Peierls equation[3–5] and to invent the quasiparticle counterpart to Grad's moment procedure.[6] The method parallels that of Banach and Piekarski.[7] We first began to study quasiparticle gases, within the framework of the one-dimensional Boltzmann–Peierls equation, in Ref. 8. The equations of transfer for the moments of the distribution function f and the quasiparticle analogue of Grad's expansion of f in terms of Hermite polynomials were some of the new results we found there. Explicit in our derivation of them was the specific assumption concerning the dispersion relation $\Omega(k)$: $\Omega(k)\sim|k|^r, 1<r\leqslant 2$, k being a wavenumber. In extending the range of validity of Grad's ideas to the case of three-dimensional quasiparticle gases, here, too, we adopt the method of Ref. 8. However, in contrast with our previous analysis we wish to make use of somewhat deeper properties of the function $\Omega(k)$ than those we have presented so far, and for this reason we shall simply define, and deal from the very outset with, a broad class of *isotropic* dispersion relations.

Here we proceed as follows. In outline, Sec. II begins with the introduction of the Boltzmann–Peierls equation, which is a necessary prerequisite for the kinetic theory, and ends with the precise specification of a certain class of isotropic dispersion relations. The model collision operator J is the subject of Sec. III. In Sec. IV we define the *symmetric traceless moments* of the distribution function f and derive the exact system of *irreducible* equations of transfer for them. Yet, in order to render the infinite hierarchy of equations of transfer just mentioned definite, we must necessarily face the difficult problem of evaluating or estimating the collision integrals (which appear on the rhs of those equations) as an exhibited series in the symmetric traceless moments of f. Of course, since the collision integrals $P^{\alpha,\beta}$ are functionally related to f, we cannot calculate them without first having to determine the irreducible moment representation of the distribution function, except in the case of a specific choice of the transition rate $\mathscr{R}(k,k_1)$ where the exact expression for $P^{\alpha,\beta}$ already has the "proper" form (Sec. V). With these statements in mind, Secs. VI and VII may turn out to be the most important. Indeed, in consequence of the introduction of a suitably chosen Hilbert space, these sections provide in essence the only real key for establishing a strict qusiparticle analogue of Grad's expansion of f in terms of Hermite polynomials and Hermite coefficients. Some auxiliary material, being in fact a supplement to Sec. VI, is included as an Appendix. The theoretical method, as developed, is directly applicable to the description of fluctuation phenomena that originate in the molecular activity but manifest themselves in the random variations of the Tchebychef–Ikenberry expansion coefficients $a^{\alpha,\beta}$ of f.[9,10] We discuss this interesting topic in Sec. VIII. In addition, in Sec. VIII we try to demonstrate that our framework can be viewed as an important building block, not only in constructing the theory that parallels that of Grad for classical gases, but also in exploiting the Chapman–Enskog procedure.[3,11]

In the present paper we use the notation of Part I. Thus the reader interested in learning more about it is referred to Sec. II and Appendix A of our previous work.[7]

II. THE BOLTZMANN-PEIERLS EQUATION AND THE DEFINITION OF A CERTAIN CLASS OF DISPERSION RELATIONS

For the purposes of this paper, the kinetic theory of a three-dimensional gas of quasiparticles should be regarded as a strongly simplified mathematical model in which the nonequilibrium occupation probability f, aside from its obvious dependence on the wave vector $k\in\mathrm{E}$, is a function of position x and time t, satisfying the Boltzmann–Peierls equation of the form[3–5]

 0022-2488/89/081816-10$02.50

$$\partial_t f + \nabla_k \Omega \circ \nabla_x f - \nabla_x \Omega \circ \nabla_k f = J(f). \quad (2.1)$$

Here J stands for the collision operator, whereas $\hbar\Omega(k,x)$ represents the energy of a single quasiparticle in the mode k; the quantity $2\pi\hbar$ denotes, as usual, Planck's constant.

We now proceed to work out the details for three simple special classes of isotropic dispersion relations.

Example 1. The isotropic (phonon) model with dispersion: We consider the dispersion relation given by

$$\Omega(k,x) = a\mu(y), \quad y:= b\kappa, \quad \kappa:= |k|, \quad (2.2)$$

where a and b are certain positive constants having the dimensions of frequency and length, respectively. By definition, $\mu(y)$ is a function in $\mathbb{R}^+ := (0, +\infty)$ with the following properties[12]:

(A) $\mu \in C^2(\mathbb{R}^+), \quad \mu'(y) > 0, \quad \mu''(y) > 0, \quad y \in \mathbb{R}^+.$

(B) $\lim_{y \Rightarrow 0} \mu(y) = 0.$

(C) $z_0 := \lim_{y \Rightarrow 0} \mu'(y) > 0.$

(D) Let $\lambda(z), z \in \mathbb{R}_0^+ := (z_0, +\infty)$ be a solution for y of the equation $\mu'(y) = z$.[13] Then as $z \Rightarrow +\infty$ the rate of growth of z is no greater than that of $\mu[\lambda(z)]$.

(E) There are positive constants $\mathbb{C}_1$ and $\mathbb{C}_2$ and a number $p, p \geqslant 0$, such that

$$\lambda'(z)[\lambda(z)]^2 \leqslant \mathbb{C}_1 + \mathbb{C}_2(z - z_0)^p, \quad \text{for all } z \in \mathbb{R}_0^+. \quad (2.3)$$

In view of Properties A, B, and C we immediately conclude that

$$y < (1/z_0)\mu(y), \quad \text{when } y \in \mathbb{R}^+. \quad (2.4)$$

An important example of $\mu(y)$ is provided by the following expression, which originates in the theory of phonon excitations in Helium II[3]:

$$\mu(y) = y(1 + \vartheta y^m), \quad (2.5a)$$

where

$$0 < \vartheta = \text{const}, \quad 0 < m = \text{const} \leqslant 3. \quad (2.5b)$$

Example 2. The power (magnon) model: We introduce now the dispersion relation of the form

$$\Omega(k,x) = c(x)\kappa^r, \quad 1 < r \leqslant 2, \quad (2.6)$$

where the dimension-bearing coefficient c in front of κ^r is a differentiable function of position x. The usefulness of the $r = 2$ case arises from every attempt to study magnon excitations in a continuum.[2] The "unrealistic" proposition $\Omega \sim \kappa^r$, $1 < r < 2$ still retains the crude qualitative physical properties of acoustic excitations in a continuum that might be important ($\Omega \Rightarrow 0$ for $\kappa \Rightarrow 0$, $\Omega \Rightarrow +\infty$ for $\kappa \Rightarrow +\infty$).

Example 3. The isotropic and dispersionless (phonon) model: Moreover, we shall deal with

$$\Omega(k,x) = c(x)\kappa, \quad (2.7)$$

where again c is a differentiable function of x. The unquestionable importance of (2.7) follows from the fact that the thermal properties of an insulating crystal at low temperature can be discussed assuming a validity of the isotropic and dispersionless approximation to the generally unknown function $\Omega(k,x)$.[3-5]

III. THE MODEL COLLISION OPERATOR J

A class of collision operators, which we shall find to include, in a sense, a counterpart of results that follow from the kinetic theory of a gas of Maxwellian molecules,[14] may be defined by the integral equality

$$J(f) = -\frac{1}{\tau}(f - f_0) + \int_{\mathrm{E}} [\mathscr{R}(k_1,k) f_1(1+f) - \mathscr{R}(k,k_1) f(1+f_1)] dk_1, \quad (3.1a)$$

in which

$$f_1 := f(k_1,x,t), \quad (3.1b)$$

$$\mathscr{R}(k,k_1) := G(\kappa)H(g \circ g_1)\delta(\kappa - \kappa_1), \quad (3.1c)$$

$$\kappa_1 := |k_1|, \quad g := \kappa^{-1}k, \quad g_1 := \kappa_1^{-1}k_1. \quad (3.1d)$$

Here $G(\kappa)$ and $H(\xi)$, $|\xi| \leqslant 1$ are certain positive functions of their arguments and $\delta(\kappa - \kappa_1)$ is Dirac's symbol. The quantity τ can be interpreted as a mean (k-independent) relaxation time; in contrast with $G(\kappa)$ and $H(\xi)$ the precise specification of τ is of no greater import in our further investigations. The f_0 denotes the Bose–Einstein distribution function that corresponds to f:

$$f_0 := (e^\omega - 1)^{-1}, \quad \omega := \frac{\hbar\Omega}{k_B T}, \quad (3.2a)$$

$$\int_{\mathrm{E}} \Omega(f - f_0) dk = 0, \quad (3.2b)$$

k_B and T being Boltzmann's constant and the kinetic theory temperature, respectively.

Insofar as the physical content of (3.1) is concerned, the integral on the rhs of (3.1a), which appears here as a mere definition, was systematically exploited by Gurevich[3] in his studies on the elastic scattering of phonons by different defects (imperfections) of a dielectric crystal. Of course, for such a scattering process the energy of a single phonon remains unchanged, i.e., the transition rate $\mathscr{R}(k,k_1)$ vanishes unless $\Omega(k,x) = \Omega(k_1,x)$, and the simplest possible assumption that

$$J(f) = \int_{\mathrm{E}} [\mathscr{R}(k_1,k) f_1(1+f) - \mathscr{R}(k,k_1) f(1+f_1)] dk_1 \quad (3.3)$$

evidently prevents the quasiparticle system from reaching local or absolute equilibrium. Thus if we wish to model the real collision operator adequately, we should slightly modify the rhs of (3.3) by adding a contribution of Callaway's type as well.

Elementary inspection shows that the proposition (3.1) for $J(f)$ correctly reproduces the three basic features of the Boltzmann–Peierls equation.

(i) A one-particle density is unaffected by collisions if and only if it is a Bose–Einstein distribution:

$$J(f) = 0 \Leftrightarrow f = f_0. \quad (3.4a)$$

(ii) The fundamental conservation equation of quasiparticle hydrodynamics, which expresses the time variation of the local density of energy as the divergence of the corresponding flow, is fulfilled because of the property

$$\int_{\mathbb{E}} \Omega J(f)dk=0. \tag{3.4b}$$

(iii) The Boltzmann–Peierls equation satisfies an H theorem:

$$\int_{\mathbb{E}} J(f)\ln[(1+f)/f]dk \geqslant 0. \tag{3.4c}$$

The ansatz (3.1c) concerns all positive functions $H(\xi)$, $|\xi|\leqslant 1$, so as to make the integral on the rhs of (3.1a) exist, not merely those $H(\xi)$ that can be expanded in a series of the Legendre polynomials[15,16] $P_\nu(\xi)$, $\nu=0,1,\ldots$, as follows:

$$H(\xi)=\sum_{\nu=0}^{\infty} d_\nu P_\nu(\xi). \tag{3.5}$$

Let $\mathscr{N}$ be any function of $\mathscr{k}$. Then the expansion (3.5) serves as basis from which to evaluate the collision integral of the form

$$\mathscr{P}[Y^\alpha(g)\mathscr{N}(\mathscr{k})]:=\int_{\mathbb{E}} Y^\alpha(g)\mathscr{N}(\mathscr{k})J(f)dk, \tag{3.6}$$

$Y^\alpha(g)$, $\alpha=0,1,\ldots$, being Ikenberry's harmonics.[14,7] As a step toward calculating $\mathscr{P}[Y^\alpha(g)\mathscr{N}(\mathscr{k})]$ from (3.1), (3.5), and (3.6), we recall the two remarkable formulas of Coope and Snider,[17]

$$P_\nu(g\circ g_1)=\frac{(2\nu)!}{2^\nu(\nu!)^2}\,Y^\nu(g)\circ Y^\nu(g_1), \tag{3.7}$$

$$\int_{\mathbb{K}} Y^\alpha(g)\otimes Y^\nu(g)dg=\frac{4\pi\alpha!}{(2\alpha+1)!!}\,\delta_{\alpha\nu}E(\alpha|\alpha), \tag{3.8}$$

where

$$\mathbb{K}:=\{g\in\mathbb{E}:|g|=1\}. \tag{3.9}$$

Here $\delta_{\alpha\nu}$ and $E(\alpha|\alpha)$ stand for the Kronecker delta and the natural projection,[7] respectively. If we adopt (3.7) and (3.8), we obtain

$$\mathscr{P}[Y^\alpha(g)\mathscr{N}(\mathscr{k})] =-\frac{1}{\tau}\int_{\mathbb{E}} Y^\alpha(g)\mathscr{N}(\mathscr{k})(f-f_0)dk +4\pi c_\alpha\int_{\mathbb{E}} Y^\alpha(g)\mathscr{k}^2\mathscr{N}(\mathscr{k})G(\mathscr{k})(f-f_0)dk, \tag{3.10a}$$

where

$$c_\alpha:=[1/(2\alpha+1)]d_\alpha\text{-}d_0 \quad (\Rightarrow c_0=0). \tag{3.10b}$$

IV. IRREDUCIBLE EQUATIONS OF TRANSFER

A. The isotropic (phonon) model with dispersion

After substituting for Ω from (2.2), the Boltzmann–Peierls equation (2.1) becomes

$$\partial_t f+ab\mu'(y)g\circ\nabla f=J(f),\quad \nabla:=\nabla_x. \tag{4.1}$$

We take as the relations determining the symmetric traceless moments $N^{\alpha,\beta}\in\mathrm{Ker}_\alpha\,\mathrm{Tr}$, $\alpha,\beta=0,1,\ldots$, of f those that are obtained if $Y^\alpha(g)\mu(y)f$ is multiplied by

$$S_\beta:=\begin{cases}[\mu'(y)]^\nu & \text{for } \beta=0,2,\ldots,\\ [y/\mu(y)][\mu'(y)]^\nu & \text{for } \beta=1,3,\ldots,\end{cases} \tag{4.2}$$

ν being the greatest integer $\leqslant\beta/2$[18] and integrated over all values of k:

$$N^{\alpha,\beta}:=\int_{\mathbb{E}} Y^\alpha(g)\mu(y)S_\beta f\,dk. \tag{4.3}$$

Looking at the definitions of the fields of physical interest, we see that all of them are proportional to certain moments of f, regarded as a function of k:

Name	Symbol
Energy per unit volume	$\epsilon:=(\hbar a/8\pi^3)N^{0,0}$
Quasimomentum per unit volume	$Q:=(\hbar/8\pi^3 b)N^{1,1}$
Heat flux	$q:=(\hbar a^2 b/8\pi^3)N^{1,2}$

With the aid of (4.1)–(4.3), the explicit calculation of the irreducible equations of transfer for $N^{\alpha,\beta}$ presents no difficulties in principle. Techniques for treating this problem were developed in Sec. III of Ref. 7. Although individual methods differ in detail, depending on whether Boltzmann's equation or that of Peierls is under consideration, all are based on exactly the same structural ingredients. In the result we obtain

$$\partial_t N^{\alpha,\beta}+ab\nabla\circ N^{\alpha+1,\beta+2} +[\alpha/(2\alpha+1)]ab\nabla\wedge N^{\alpha-1,\beta+2}=P^{\alpha,\beta}, \tag{4.4a}$$

$$\alpha,\beta=0,1,\ldots,\quad N^{-1,\beta}:=0, \tag{4.4b}$$

where

$$P^{\alpha,\beta}:=\int_{\mathbb{E}} Y^\alpha(g)\mu(y)S_\beta J(f)dk. \tag{4.5}$$

By choosing $\mathscr{N}(\mathscr{k})$ to be $\mu(y)S_\beta$ in Eq. (3.10a), we arrive, in virtue of $P^{\alpha,\beta}=\mathscr{P}[Y^\alpha(g)\mu(y)S_\beta]$, at

$$P^{\alpha,\beta}=-\frac{1}{\tau}(N^{\alpha,\beta}-N^{\alpha,\beta}_{\mathrm{eq}})+4\pi c_\alpha \times\int_{\mathbb{E}} Y^\alpha(g)\mathscr{k}^2 G(\mathscr{k})\mu(y)S_\beta(f-f_0)dk, \tag{4.6}$$

where

$$N^{\alpha,\beta}_{\mathrm{eq}}:=\int_{\mathbb{E}} Y^\alpha(g)\mu(y)S_\beta f_0\,dk. \tag{4.7}$$

It is easily verified that $N^{\alpha,\beta}_{\mathrm{eq}}=0$ unless $\alpha=0$.

B. The power (magnon) model

If we substitute (2.6) into (2.1), we obtain

$$\partial_t f+rcv(\mathscr{k})g\circ\nabla_x f-\mathscr{k}^r W\circ\nabla_k f=J(f), \tag{4.8a}$$

where

$$v(\mathscr{k}):=\mathscr{k}^{r-1},\quad W:=\nabla_x c. \tag{4.8b}$$

The symmetric traceless moments of f are defined as follows:

$$N^{\alpha,\beta}:=\int_{\mathbb{E}} Y^\alpha(g)\mathscr{k}^r[v(\mathscr{k})]^{\beta-1}f\,dk, \tag{4.9a}$$

$$\alpha,\beta=0,1,\ldots. \tag{4.9b}$$

Multiplying the Boltzmann–Peierls equation (4.8a) by $Y^\alpha(g)\mathscr{k}^r[v(\mathscr{k})]^{\beta-1}$ and integrating the result over $\mathbb{E}$ yield

$$\partial_t N^{\alpha,\beta}+rc\nabla\circ N^{\alpha+1,\beta+1}+[\alpha/(2\alpha+1)]rc\nabla\wedge N^{\alpha-1,\beta+1} +[2r+(\beta-1)(r-1)-\alpha]W\circ N^{\alpha+1,\beta+1}$$

$$+ [\alpha/(2\alpha+1)][2r+(\beta-1)(r-1) + \alpha + 1] W\wedge N^{\alpha-1,\beta+1} = P^{\alpha,\beta}; \quad (4.10a)$$

$$\alpha,\beta = 0,1,\ldots, \quad N^{-1,\beta} := 0, \quad (4.10b)$$

where

$$P^{\alpha,\beta} := \int_{\mathrm{E}} Y^{\alpha}(g) k^{r}[v(k)]^{\beta-1} J(f) dk = -\frac{1}{\tau}(N^{\alpha,\beta} - N^{\alpha,\beta}_{\mathrm{eq}}) + 4\pi c_{\alpha} \times \int_{\mathrm{E}} Y^{\alpha}(g) k^{r+2} G(k)[v(k)]^{\beta-1}(f-f_0) dk. \quad (4.11)$$

Here the $N^{\alpha,\beta}_{\mathrm{eq}}$ is given by

$$N^{\alpha,\beta}_{\mathrm{eq}} := \int_{\mathrm{E}} Y^{\alpha}(g) k^{r}[v(k)]^{\beta-1} f_0\, dk. \quad (4.12)$$

As to the physical interpretation of some of the irreducible moments $N^{\alpha,\beta}$ of f, we have

$$\epsilon = \frac{\hbar c}{8\pi^3} N^{0,1}, \quad Q = \frac{\hbar}{8\pi^3} N^{1,0}, \quad q = \frac{r\hbar c^2}{8\pi^3} N^{1,2}. \quad (4.13)$$

C. The isotropic and dispersionless (phonon) model

Using (2.1) and (2.7), it is a straightforward matter to show that

$$\partial_t f + c g \circ \nabla_x f - k W \circ \nabla_k f = J(f). \quad (4.14)$$

Within the framework set up here, the symmetric traceless moments of f are given by

$$N^{\alpha,\beta} := \int_{\mathrm{E}} Y^{\alpha}(g) k^{\beta+1} f\, dk, \quad (4.15a)$$

$$\alpha,\beta = 0,1,\ldots. \quad (4.15b)$$

It follows directly from (4.14) and (4.15) that

$$\partial_t N^{\alpha,\beta} + c\nabla \circ N^{\alpha+1,\beta} + [\alpha/(2\alpha+1)] c \nabla \wedge N^{\alpha-1,\beta} + (\beta-\alpha+2) W \circ N^{\alpha+1,\beta} + [\alpha/(2\alpha+1)](\alpha+\beta+3) W \wedge N^{\alpha-1,\beta} = P^{\alpha,\beta}, \quad (4.16a)$$

$$\alpha,\beta = 0.1,\ldots, \quad N^{-1,\beta} := 0, \quad (4.16b)$$

$P^{\alpha,\beta}$ being the collision integral,

$$P^{\alpha,\beta} := \int_{\mathrm{E}} Y^{\alpha}(g) k^{\beta+1} J(f) dk = -\frac{1}{\tau}(N^{\alpha,\beta} - N^{\alpha,\beta}_{\mathrm{eq}}) + 4\pi c_{\alpha} \int_{\mathrm{E}} Y^{\alpha}(g) k^{\beta+3} G(k)(f-f_0) dk, \quad (4.17a)$$

$$N^{\alpha,\beta}_{\mathrm{eq}} := \int_{\mathrm{E}} Y^{\alpha}(g) k^{\beta+1} f_0\, dk. \quad (4.17b)$$

Regardless of the rhs of (4.16a), the differential equations just obtained are *unsymmetrically* coupled (uncoupled), in that a single branch $\mathscr{M}_{\gamma}: = \{N^{0,\gamma}, N^{1,\gamma},\ldots\}$ of moments with $\gamma \neq \beta$ is not allowed to affect the space-time behavior of $\mathscr{M}_{\beta}$, while $N^{\alpha+1,\beta}$ influences only $N^{\alpha-1,\beta}$ and $N^{\alpha,\beta}$ and none of the other variables.

Chief among $N^{\alpha,\beta}$ are the following fields:

$$\epsilon = \frac{\hbar c}{8\pi^3} N^{0,0}, \quad Q = \frac{\hbar}{8\pi^3} N^{1,0}, \quad q = \frac{\hbar c^2}{8\pi^3} N^{1,0}. \quad (4.18)$$

V. EXPLICIT COLLISION INTEGRALS FOR A SPECIAL CHOICE OF THE TRANSITION RATE $\mathscr{R}(k,k_1)$: $\Omega(k,x) = a\mu(y)$

In Sec. III we defined and explained the model collision operator J in general terms. One possible way of making further progress is to descend to special cases of that operator. When the transition rate $\mathscr{R}(k,k_1)$ in (3.1c) is specialized by letting $G(k)$ have the form

$$G(k) = k^{-2} \sum_{\nu=0}^{p} \mathbb{B}_{\nu}[\mu'(y)]^{\nu}, \quad p = 0,1,\ldots, \quad (5.1)$$

$\mathbb{B}_{\nu}, 0 \leqslant \nu \leqslant p$, being certain constants, then placing (5.1) into (4.6) and using the definition (4.3) we obtain

$$P^{\alpha,\beta} = -\frac{1}{\tau}(N^{\alpha,\beta} - N^{\alpha,\beta}_{\mathrm{eq}}) + 4\pi c_{\alpha} \sum_{\nu=0}^{p} \mathbb{B}_{\nu}(N^{\alpha,\beta+2\nu} - N^{\alpha,\beta+2\nu}_{\mathrm{eq}}). \quad (5.2)$$

Thus we see now that the ansatz (5.1) delivers the collision integral (4.5) as an *exhibited* linear function of finitely many moments of f and the predictions of the kinetic theory of quasiparticle gases for (3.1), (3.5), and (5.1) have in a sense the same quality as those for a classical gas composed of Maxwellian molecules.[14] [However, as we have demonstrated in Sec. V of Ref. 7, in a gas of Maxwellian molecules the rhs of the equations of transfer does not introduce, *in contrast with that to* (4.4a), the so-called forward coupling of the equations of moments.] Of course, our calculations are rather simple. But whether there be another quasiparticle model, so as to satisfy the extra requirement of evaluating $P^{\alpha,\beta}$ in terms of $N^{\alpha,\beta}$ alone without first having to determine the irreducible moment representation of f, is not presently known!

The functions $G(k)$, as defined by (5.1), are of physical interest whenever $p = 0,1,\ldots$, and $\mathbb{B}_{\nu}$, $0 \leqslant \nu \leqslant p$ are such that $G(k) > 0$ for any $k \in \mathbb{R}^{+}$, and also each acceptable $G(k)$ cannot increase without bound with decreasing k. Thus, considering the proposition (2.5) as a useful illustration, we easily arrive, denoting by s the greatest integer $\leqslant 2/m$, at the following restrictions on the admissible values of p and $\mathbb{B}_{\nu}$, $0 \leqslant \nu \leqslant p$: (i) $p > s$; (ii) $\mathbb{B}_{\nu}$, $s < \nu \leqslant p$ are positive; and (iii) $\mathbb{B}_{\nu}$, $0 \leqslant \nu \leqslant s$ are calculated from $\mathbb{B}_{\nu}$, $s < \nu \leqslant p$:

$$\sum_{\nu=\gamma}^{s} \binom{\nu}{\gamma} \mathbb{B}_{\nu} = -\sum_{\nu=1+s}^{p} \binom{\nu}{\gamma} \mathbb{B}_{\nu}, \quad (5.3a)$$

$$0 \leqslant \gamma \leqslant s. \quad (5.3b)$$

After perusing the foregoing calculations for $\Omega(k,x) = a\mu(y)$, the reader will conclude, almost effortlessly, that there are reasons in support of the universality of the method when other dispersion relations, namely, (2.6) and (2.7), are taken. This is then a good point to leave simplifying assumptions, such as, for instance, (5.1), and move to the general problem of finding the irreducible moment representation of f.

VI. A QUASIPARTICLE ANALOG OF GRAD'S METHOD FOR $\Omega(k,x)=a\mu(y)$

A. The Tchebychef–Ikenberry expansion of the distribution function

The choice (5.1) of $G(\not{k})$ ensures the satisfaction of (5.2): it also implies that we can outflank the original Boltzmann–Peierls equation itself by working directly with the equations of transfer for the expectations $N^{\alpha,\beta}$ of $Y^\alpha(g)\mu(y)S_\beta$. As, however, the function $G(\not{k})$ other than (5.1) causes difficulty in the rigorous determination of $P^{\alpha,\beta}$ from finitely many moments of f, the general problem of formally evaluating $P^{\alpha,\beta}$ as an *infinite* series in the moments may be solved in a manner analogous to that employed for classical gases.[6,19–22,7] Equation (4.6) indicates in turn that in order to calculate $P^{\alpha,\beta}$, it is necessary first to express f in terms of $N^{\alpha,\beta}$ and therefore to propose the irreducible moment representation of f. The only object of this section is both to invent such a representation and to investigate its elementary properties.

(1) To this end, let us introduce the following definitions:

$$\mathbb{W}(z;\Theta):=\lambda'\lambda^2\omega^2e^\omega/(e^\omega-1)^2,\quad z\in\mathbb{R}_0^+,\tag{6.1a}$$

$$\{\varphi_1|\varphi_2\}_1:=\int_{\mathbb{R}_0^+}\varphi_1(z)\varphi_2(z)\mathbb{W}(z;\Theta)dz,\tag{6.1b}$$

$$\|\varphi\|_1:=(\{\varphi|\varphi\}_1)^{1/2},\tag{6.1c}$$

$$\mathbb{L}^2(\mathbb{R}_0^+;\mathbb{W}(z;\Theta)dz):=\{\varphi:\|\varphi\|_1<+\infty\},\tag{6.1d}$$

in which

$$\Theta:=ha/k_BT,\quad \omega=\Theta\mu[\lambda(z)].\tag{6.1e}$$

It is clear from (6.1d) that $\mathbb{L}^2(\mathbb{R}_0^+;\mathbb{W}(z;\Theta)dz)$ is a real Hilbert space with the scalar product $\{\ |\ \}_1$ given by (6.1b). Referring back to Sec. II, we can utilize Properties A,...,E of $\mu(y)$ so as to demonstrate that, for certain functions $\mathbb{M}$ and $\mathbb{C}$ of Θ,

$$\mathbb{W}(z;\Theta)\leqslant\mathbb{M}e^{-\mathbb{C}z},\quad z\in\mathbb{R}_0^+\tag{6.2}$$

and to prove, by appealing to the lemma of Dijkstra and van Leeuwen as formulated in Ref. 23 on p. 468, that the collection of polynomials defined on $\mathbb{R}_0^+$ is a dense subset of $\mathbb{L}^2(\mathbb{R}_0^+;\mathbb{W}(z;\Theta)dz)$. Now, writing S_β, $\beta=0,1,\ldots$, from (4.2), in the form

$$S_{2\nu}(z)=z^\nu,\quad\text{when }\beta=0,2,\ldots,\tag{6.3a}$$

$$S_{2\nu+1}(z)=(\lambda(z)/\mu[\lambda(z)])z^\nu,\quad\text{when }\beta=1,3,\ldots,\tag{6.3b}$$

and applying the inequality $\lambda(z)<z_0^{-1}\mu[\lambda(z)]$ which parallels that in (2.4), we see at a glance not only that $S_\beta\in\mathbb{L}^2(\mathbb{R}_0^+;\mathbb{W}(z;\Theta)dz)$ but also that the functions S_β, $\beta=0,1,\ldots$, are dense in our Hilbert space. Let $\{S_\beta;\beta=0,1,\ldots\}$ be a linearly independent set. Then, as we know from the monographs of Szegö[15] and Sansone,[16] it is possible to construct, via a standard procedure called orthogonalization, an *orthonormal* set of Tchebychef functions $A_\beta(z;\Theta)$, $\beta=0,1,\ldots$, such that

$$A_\beta(z;\Theta)=\sum_{\nu=0}^{\beta}c_{\beta\nu}(\Theta)S_\nu(z),\tag{6.4a}$$

$$S_\beta(z)=\sum_{\nu=0}^{\beta}d_{\beta\nu}(\Theta)A_\nu(z;\Theta),\tag{6.4b}$$

$$c_{\beta\beta}>0,\quad d_{\beta\beta}=1/c_{\beta\beta},\tag{6.4c}$$

$c_{\beta\nu}$ and $d_{\beta\nu}$ being certain uniquely determined functions of the parameter Θ.[24] The Tchebychef functions $A_\beta(z;\Theta)$, $\beta=0,1,\ldots$, are complete, i.e., they form a basis in $\mathbb{L}^2(\mathbb{R}_0^+;\mathbb{W}(z;\Theta)dz)$.

(2) Of course, a sequence of Ikenberry's harmonics $Y^\alpha(g)$, $\alpha=0,1,\ldots$, can be treated as a complete orthogonal[17] [but, according to Eq. (3.8), not normalized] set of "angular" functions of the "point" g of the unit sphere $\mathbb{K}$ for immediate use in expanding an arbitrary element of the real Hilbert space $\mathbb{L}^2(\mathbb{K};dg)$ generated by the scalar product

$$\{\varphi_1|\varphi_2\}_2:=\textstyle\int_{\mathbb{K}}\varphi_1(g)\varphi_2(g)dg.\tag{6.5}$$

(3) Let $\mathbb{R}_0^+\times\mathbb{K}$ be a set of all ordered pairs (z,g), $z\in\mathbb{R}_0^+$, $g\in\mathbb{K}$, and consider the induced [by $\mathbb{L}^2(\mathbb{R}_0^+;\mathbb{W}(z;\Theta)dz)$ and $\mathbb{L}^2(\mathbb{K};dg)$] Hilbert space $\mathbb{L}^2(\mathbb{R}_0^+\times\mathbb{K};\mathbb{W}(z;\Theta)dz\,dg)$ with the scalar product $\{\ |\ \}_0$ given by

$$\{\varphi_1|\varphi_2\}_0:=\textstyle\int_{\mathbb{R}_0^+}\int_{\mathbb{K}}\varphi_1(z,g)\varphi_2(z,g)\mathbb{W}(z;\Theta)dz\,dg.\tag{6.6}$$

Inspection shows that the system $\{Y^\alpha(g)A_\beta(z;\Theta);\alpha,\beta=0,1,\ldots\}$ is a basis in $\mathbb{L}^2(\mathbb{R}_0^+\times\mathbb{K};\mathbb{W}(z,\Theta)dz\,dg)$.

Remembering that $y=b|k|$, to each wave vector $k\in\mathbb{E}\setminus\{0\}$ corresponds a uniquely determined element $(z,g)\in\mathbb{R}_0^+\times\mathbb{K}$ such that $z=\mu'(y)$ and $g=|k|^{-1}k$. Since $k=b^{-1}\lambda(z)g$, the opposite statement is also true. Whenever there is no possibility of confusion, it may be convenient not to distinguish between $\mathbb{E}\setminus\{0\}$ and $\mathbb{R}_0^+\times\mathbb{K}$, but to regard them as the same set.

Condition I: Let φ be a certain function defined on $\mathbb{R}_0^+\times\mathbb{K}$ and suppose that

$$\hat{\varphi}(z,g):=\frac{(e^\omega-1)}{\omega e^\omega}\frac{[\varphi(z,g)-f_0(\omega)]}{f_0(\omega)},\tag{6.7}$$

f_0 being equal to $(e^\omega-1)^{-1}$. Then φ satisfies Condition I if $\hat{\varphi}\in\mathbb{L}^2(\mathbb{R}_0^+\times\mathbb{K};\mathbb{W}(z;\Theta)dz\,dg)$.

Now, if the distribution function satisfies Condition I, then f has a unique expansion

$$f=f_0\Big[1+\frac{\omega e^\omega}{e^\omega-1}\times\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}a^{\alpha,\beta}(x,t)\circ Y^\alpha(g)A_\beta(z;\Theta)\Big]\tag{6.8}$$

and this expansion converges in the *mean* to the function f. The Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}\in\mathrm{Ker}_\alpha$ Tr have a simple representation in terms of $f-f_0$.

Because of the orthogonality properties of $A_\beta(z;\Theta)$ and $Y^\alpha(g)$, we find that

$$a^{\alpha,\beta}=\frac{(2\alpha+1)!!}{4\pi\alpha!}b^3\int_{\mathbb{E}}\omega Y^\alpha(g)A_\beta(z;\Theta)(f-f_0)dk,\tag{6.9}$$

and if we assume that f_0 is a Bose–Einstein distribution function which corresponds to f, we get

$$a^{0,0}=0.\tag{6.10}$$

For every fixed α, $\alpha=0,1,\ldots$, each member of one of the two sets $\{N^{\alpha,\beta}-N_{\mathrm{eq}}^{\alpha,\beta};\beta=0,\ldots,\gamma\}$ and $\{a^{\alpha,\beta};\beta=0,\ldots,\gamma\}$ is a linear combination of the members of the other. Indeed, using (4.3), (4.7), (6.9), and (6.4), we arrive at

$$a^{\alpha,\beta} = \frac{(2\alpha+1)!!}{4\pi\alpha!}\Theta b^3 \sum_{\nu=0}^{\beta} c_{\beta\nu}(N^{\alpha,\nu} - N_{\rm eq}^{\alpha,\nu}), \tag{6.11a}$$

$$N^{\alpha,\beta} - N_{\rm eq}^{\alpha,\beta} = \frac{4\pi\alpha!}{(2\alpha+1)!!}\frac{1}{\Theta b^3}\sum_{\nu=0}^{\beta} d_{\beta\nu}a^{\alpha,\nu}. \tag{6.11b}$$

In this sense the $a^{\alpha,\beta}$ are simply an alternative set of moments of f, just as were the Laguerre–Ikenberry coefficients $b^{\alpha|\beta}$ introduced in Sec. IV A of Ref. 7.

One final point to note is as follows. Our interest in studying the irreducible moment representation of f rests altogether on the fact that, in contrast with the irreducible expansion of f for classical gases which results from Eq. (4.3) of Part I, there is *no* reducible counterpart to (6.8) at all.[25] Although, instead of (4.3) and (4.4), we could equally well have defined in Sec. IV A the reducible moments of f:

$$\widehat{N}^{\alpha,\beta} := \int_{\mathbb{E}} (\otimes^\alpha g)\mu(y)S_\beta f\,dk \tag{6.12}$$

and also derived there the balance equations for them; the benefit which arises if we deal with (4.3) rather than (6.12) is manifestly strengthened by the *nonexistence* of the reducible variant of (6.11) as well.[25]

The procedure of this section can be repeated, essentially word for word with only slight technical changes in the method, to take into account (2.6) as well as (2.7). Some discussion of these cases is offered in the Appendix.

B. The scheme of constructing Tchebychef functions

In connection with the expansion (6.8) we introduced the set of real-valued continuous functions $A_\beta(z;\Theta)$ on $\mathbb{R}_0^+$ which are orthonormal with respect to weight $\mathbb{W}(z;\Theta)$: $\{A_\beta|A_\nu\}_1 = \delta_{\beta\nu}$, $\delta_{\beta\nu}$ being Kronecker's delta and $\{\ |\ \}_1$ being a scalar product defined by (6.1b). The importance of Tchebychef functions $A_\beta(z;\Theta)$ is obvious, for it is only they that permit us even to frame such questions as whether Grad's ideas[6,26] are of interest in the kinetic theory of quasi-particle gases. Of course, it is one thing to state that there are certain special Tchebychef functions of $z\in\mathbb{R}_0^+$ and Θ, but it is quite another to determine them. Thus we wish to conclude Sec. VI A with a brief mention of the method of constructing $A_\beta(z;\Theta)$.

We begin by defining the class of auxiliary quantities to be considered:

$$k_\beta := \{zA_{\beta-2}|A_{\beta-2}\}_1, \tag{6.13a}$$

$$\ell_\beta := \{z^2A_{\beta-2}|A_{\beta-2}\}_1, \tag{6.13b}$$

$$n_\beta := \{zA_{\beta-2}|A_{\beta-1}\}_1. \tag{6.13c}$$

We are now ready to establish the following useful theorem.

Theorem: The following relation holds for any five consecutive orthonormal Tchebychef functions:

$$A_\beta(z;\Theta) = (e_\beta z - f_\beta)A_{\beta-2}(z;\Theta) - g_\beta A_{\beta-1}(z;\Theta) - h_\beta A_{\beta-3}(z;\Theta) - i_\beta A_{\beta-4}(z;\Theta),\ \beta = 4,5,\dots . \tag{6.14a}$$

Here e_β, f_β, g_β, h_β, and i_β are functions of Θ alone; $e_\beta > 0$, $f_\beta > 0$, and $i_\beta > 0$. If the highest coefficient $c_{\beta\beta}(\Theta)$ of $A_\beta(z;\Theta)$ is denoted by τ_β in Eqs. (6.4), then we have

$$e_\beta = \tau_\beta/\tau_{\beta-2},\quad i_\beta = e_\beta/e_{\beta-2}, \tag{6.14b}$$

$$g_\beta = e_\beta n_\beta,\quad h_\beta = e_\beta n_{\beta-1}, \tag{6.14c}$$

$$f_\beta = e_\beta k_\beta, \tag{6.14d}$$

$$\tau_\beta = \tau_{\beta-2}^2[(\ell_\beta - k_\beta^2 - n_\beta^2 - n_{\beta-1}^2)\tau_{\beta-2}^2 - \tau_{\beta-4}^2]^{-1/2}. \tag{6.14e}$$

For the proof, we first observe that $zS_{\beta-2} = S_\beta$ and then determine e_β so that $A_\beta - e_\beta zA_{\beta-2}$ is a linear combination $\delta_0A_0 + \cdots + \delta_{\beta-1}A_{\beta-1}$. Because of the orthogonality properties of Tchebychef functions it is readily seen that $\delta_\nu = 0$ if $\nu < \beta - 4$. Therefore (6.14a) follows. The first part of (6.14b) is a consequence of (6.14a); the second part follows from

$$\begin{aligned}\int_{\mathbb{R}_0^+} &A_{\beta-4}(z;\Theta)A_\beta(z;\Theta)\mathbb{W}(z;\Theta)dz = 0\\ &= -i_\beta + e_\beta\int_{\mathbb{R}_0^+} zA_{\beta-4}(z;\Theta)A_{\beta-2}\\ &\quad\times(z;\Theta)\mathbb{W}(z;\Theta)dz,\end{aligned}$$

since the integral of the right-hand member is equal to

$$\tau_{\beta-4}\{S_{\beta-2}|A_{\beta-2}\}_1 = 1/e_{\beta-2}.$$

The proof of (6.14c)–(6.14e) may be obtained by combining (6.14a) with $\{A_\beta|A_\nu\}_1 = \delta_{\beta\nu}$ where $\nu = \beta-4,\dots,\beta$. ∎

The recurrence formulas (6.14), which tell us that A_β is *uniquely* determined from $A_{\beta-4}$, $A_{\beta-3}$, $A_{\beta-2}$, and $A_{\beta-1}$, are valid also for $\beta = 2,3$ if we write $A_{-2} = A_{-1} = \tau_{-2} = \tau_{-1} = 0$.

In summary, Eqs. (6.14) deliver through routine algebra as many of the orthonormal Tchebychef functions A_β as be desired, provided only that A_0 and A_1 shall have been calculated first. The theorem we present here has not been published before, so far as we know. We were led to it in an attempt to generalize some results of Szegö on Tchebychef polynomials (Ref. 15, p. 42).

C. Formal evaluation of collision integrals

So as to be able to arrive at the collision integrals $P^{\alpha,\beta}$ on the rhs of (4.4a) that depend upon $N^{\alpha,\beta} - N_{\rm eq}^{\alpha,\beta}$, we follow, step by step, Grad's line of thought.[6] Thus we formally substitute the representation (6.8) of f into (4.6) and then, applying $k^2\,dk = b^{-5}\lambda'\lambda^4\,dz\,dg$, (3.8), (3.10b), (3.2a), (6.1e), (6.1a), and (6.11a), expand the result,

$$\begin{aligned}P^{\alpha,\beta} &= -\frac{1}{\tau}(N^{\alpha,\beta} - N_{\rm eq}^{\alpha,\beta})\\ &\quad + 4\pi c_\alpha\int_{\mathbb{E}} dk\, Y^\alpha(g)k^2G(k)\mu(y)S_\beta\\ &\quad\times f_0\left[1 + \frac{\omega e^\omega}{e^\omega - 1}\sum_{\nu=0}^{\infty}\sum_{\gamma=0}^{\infty} a^{\nu,\gamma}\circ Y^\nu(g)A_\gamma(z;\Theta)\right]\\ &= -\frac{1}{\tau}(N^{\alpha,\beta} - N_{\rm eq}^{\alpha,\beta}) + \frac{4\pi c_\alpha}{b^2}\sum_{\nu=0}^{\infty}\\ &\quad\times\sum_{\gamma=\nu}^{\infty} p_{\beta\gamma}c_{\gamma\nu}(N^{\alpha,\nu} - N_{\rm eq}^{\alpha,\nu}),\end{aligned} \tag{6.15a}$$

in which

$$p_{\beta\gamma} := \int_{\mathbb{R}_0^+}(\lambda(z))^2G(b^{-1}\lambda(z))S_\beta(z)A_\gamma(z;\Theta)\mathbb{W}(z;\Theta)dz. \tag{6.15b}$$

The effect of using (6.15) to determine the rhs of (4.4a) is not only the system of equations of transfer for the expectations $N^{\alpha,\beta}$ of the Tchebychef–Ikenberry functions $Y^{\alpha}(g)\mu(y)S_{\beta}$ but also the precise algorithm (6.15b) for expressing the collision elements $p_{\beta\gamma}$ in (6.15a) in terms of certain functions of Θ alone. Of course, just as in Grad's approach, our results are only formal since the right-hand side in (6.15a) involves the product of two series, neither of which need converge pointwise. If we set G given by (5.1) in (6.15b), it follows from

$$(\lambda(z))^2 G(b^{-1}\lambda(z))S_{\beta}(z) = b^2 \sum_{\nu=0}^{p} \mathbb{B}_{\nu} S_{\beta+2\nu}(z), \qquad (6.16)$$

(6.4), and the orthogonality properties of $A_{\beta}(z;\Theta)$ that $p_{\beta\gamma}$ vanishes if $\gamma > \beta + 2p$ and that the series

$$\sum_{\gamma=\nu}^{\infty} p_{\beta\gamma} c_{\gamma\nu} = \sum_{\gamma=\nu}^{\beta+2p} p_{\beta\gamma} c_{\gamma\nu} \qquad (6.17)$$

is equal to 0 unless $\nu = \beta + 2s$ where $s = 0,...,p$. For functions $G(\kappa)$ other than (5.1) all the collision elements are usually nonzero, but presumably in many cases of interest the dominant ones appear to be those that are already nonzero if $G(\kappa)$ is to be approximated fairly closely by the sum

$$\sum_{\nu=0}^{p} \mathbb{B}_{\nu} [\mu'(y)]^{\nu},$$

in which the choice of $p = 0,1,...,$ and $\mathbb{B}_{\nu}$, $0 \leqslant \nu \leqslant p$ depends on the shape of $G(\kappa)$. By this means it is easy to think of at least one reasonable way[21] of simplifying the right-hand side in (6.15a),

$$\frac{4\pi c_{\alpha}}{b^2} \sum_{\nu=0}^{\infty} \sum_{\gamma=\nu}^{\infty} p_{\beta\gamma} c_{\gamma\nu} (N^{\alpha,\nu} - N^{\alpha,\nu}_{\mathrm{eq}})$$
$$\Rightarrow \frac{4\pi c_{\alpha}}{b^2} \sum_{\nu=\beta}^{\beta+2p} \sum_{\gamma=\nu}^{\beta+2p} p_{\beta\gamma} c_{\gamma\nu} (N^{\alpha,\nu} - N^{\alpha,\nu}_{\mathrm{eq}}). \qquad (6.18)$$

Now, if we replace $P^{\alpha,\beta}$ by the last two terms on the rhs of (6.15a) in Eq. (4.4a) and then adopt (6.18), we see that the ansatz (6.18) causes the so-called forward coupling of the equations of moments, as first indicated in Sec. V in a slightly different context.

On the understanding that either $G(\kappa)$ is given by (5.1) or the transformation (6.18) may occur approximately (in a sense not yet made precise) when a more general function $G(\kappa)$ is used, the method of solving the infinite system of equations of transfer for $N^{\alpha,\beta}$ will be discussed elsewhere.

VII. THE APPROXIMATE FORMULAS FOR THE ENTROPY DENSITY h AND ITS FLUX Φ: $\Omega(k,x) = a\mu(y)$

Insofar as classical gases are concerned, Grad (Ref. 6, pp. 267 and 284) wrote that "by the choice of a Maxwellian, local or absolute, as a starting point, one has almost no alternative to the use of a Hermite (Laguerre–Ikenberry) polynomial expansion" and that "the choice of Hermite (Laguerre–Ikenberry) coordinates [coefficients (moments)] always diagonalizes the entropy." Although in actual fact there are other expansions of comparable if not greater mathematical complexity, Grad's framework (Ref. 6, p. 284) "has been seen to be particularly appropriate in generalizing fluid type equations (near a locally Maxwellian state) and for the linearized Boltzmann equation."

In Sec. VI our discussion has centered around the elementary recurrence and conversion formulas, but *first of all* we need a plausible argument in favor of the usefulness of the very specific postulate (6.1a). However, under just this proposition for the weight function, the expansion (6.8) of f constitutes the underlying theoretical structure of what one could recognize as a kinetic background of nonequilibrium thermodynamics of a gas of quasiparticles. Indeed, in taking stock of the connection between kinetics and thermodynamics,[27] one of the most important features of (6.8) is the unique way the specific choice $\mathbb{W}(z;\Theta)$ affects the approximate dependence of both the entropy density h and the entropy flux Φ upon the Tchebychef–Ikenberry expansion coefficients $a^{\alpha,\beta}$.

The tradition of the kinetic theory of quasiparticle gases chooses the following definitions of h and Φ[3]:

$$h := \frac{k_{\mathrm{B}}}{8\pi^3} \int_{\mathbb{E}} \mathbb{F}\, dk, \qquad (7.1a)$$

$$\Phi := \frac{k_{\mathrm{B}}}{8\pi^3} ab \int_{\mathbb{E}} g\mu'(y) \mathbb{F}\, dk, \qquad (7.1b)$$

where

$$\mathbb{F} := (1+f)\ln(1+f) - f\ln f. \qquad (7.1c)$$

Of course, when we wish to compare propositions of the kinetic theory with phenomenological statements about entropy and its flux, we must multiply (4.1) by a function $(k_{\mathrm{B}}/8\pi^3)[\ln(1+f) - \ln f]$ and integrate the result over $\mathbb{E}$. Then, using (3.4c), we obtain the entropy principle of the form

$$\partial_t h + \nabla \circ \Phi \geqslant 0. \qquad (7.2)$$

The crucial role of the Bose–Einstein distribution function f_0, especially in relation to the concept of local equilibrium, provides a persuasive reason for estimating the deviation of the entropy density h from its equilibrium counterpart h_0 for f_0,

$$h_0 := \frac{k_{\mathrm{B}}}{8\pi^3} \int_{\mathbb{E}} \mathbb{F}_0\, dk. \qquad (7.3a)$$

Here we have set

$$\mathbb{F}_0 := (1+f_0)\ln(1+f_0) - f_0 \ln f_0. \qquad (7.3b)$$

If f satisfies Condition I of Sec. VI A, i.e., if

$$\hat{f}(z,g) = \frac{(e^{\omega}-1)}{\omega e^{\omega}} \frac{[f(z,g) - f_0(\omega)]}{f_0(\omega)}$$

belongs to the space of functions which are square integrable over $\mathbb{R}_0^+ \times \mathbb{K}$ with weight $\mathbb{W}(z;\Theta)$, then we are justified in using for $f_0^{-1}(f - f_0)$ here the function ϵ_0 as defined by

$$\epsilon_0 := \frac{\omega e^{\omega}}{e^{\omega}-1} \vartheta, \qquad (7.4a)$$

$$\vartheta := \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} a^{\alpha,\beta}(x,t) \circ Y^{\alpha}(g) A_{\beta}(z;\Theta). \qquad (7.4b)$$

In order to arrive at an approximate formula for an h corresponding to $f = f_0(1+\epsilon_0)$ which amounts to expressing $h - h_0$ in terms of the Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}$ of f, we replace the logarithm $\ln(1+X)$ ($|X| < 1$) by the first two terms in its Taylor expansion $X - \frac{1}{2}X^2 + \ldots$, and so we obtain

$$
\begin{aligned}
\mathbb{F} &= [1 + f_0(1+\epsilon_0)]\ln[(1+f_0)(1+e^{-\omega}\epsilon_0)] \\
&\quad - f_0(1+\epsilon_0)\ln[f_0(1+\epsilon_0)] \\
&= \mathbb{F}_0 + f_0\omega\epsilon_0 + (1+f_0)(1+e^{-\omega}\epsilon_0)\ln(1+e^{-\omega}\epsilon_0) \\
&\quad - f_0(1+\epsilon_0)\ln(1+\epsilon_0) \\
&= \mathbb{F}_0 + f_0\omega\epsilon_0 - \tfrac{1}{2}e^{-\omega}\epsilon_0^2 + \ldots \\
&\cong \mathbb{F}_0 + (\lambda'\lambda^2)^{-1}\mathbb{W}(z;\Theta)\vartheta - \tfrac{1}{2}(\lambda'\lambda^2)^{-1}\mathbb{W}(z;\Theta)\vartheta^2.
\end{aligned} \tag{7.5}
$$

Substituting (7.5) into (7.1a) and referring to the orthogonality properties of $A_\beta(z;\Theta)$ and $Y^\alpha(g)$, we easily show that

$$h - h_0 \cong -\frac{k_B}{4\pi^2 b^3}\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\frac{\alpha!}{(2\alpha+1)!!}a^{\alpha,\beta}\circ a^{\alpha,\beta}. \tag{7.6}$$

In deriving (7.6) from (7.1a) and (7.5) we have used (6.10). Routine calculation shows that

$$\{\hat{f}|\hat{f}\}_0 = 4\pi\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\frac{\alpha!}{(2\alpha+1)!!}a^{\alpha,\beta}\circ a^{\alpha,\beta}, \tag{7.7}$$

where $\{\ |\ \}_0$ is a scalar product given by (6.6). Therefore if we assume that f satisfies Condition I, the series on the rhs of (7.6) converges. Clearly, our study of the dependence of $h - h_0$ upon $a^{\alpha,\beta}$ applies only to small values of $f_0^{-1}(f - f_0)$, they being the only ones for which a plausible argument, however limited, in favor of the practical usefulness of the specific expansion (6.8) could be given (see also Sec. VIII). In other words, provided that the expansion terms in $\mathbb{F}$ of third (fourth, etc.) order with respect to ϑ can be considered negligible, it is worthwhile applying the weight function (6.1a) as well as the transformation (6.11) to diagonalize the largest contribution to $h - h_0$—that of great interest in linear extended thermodynamics.[9,10] [By the way, Eq. (7.6) suggests that among all quasiparticle occupation probabilities f having the same principal moment $N^{0,0}$ ($a^{0,0} = 0$) the Bose–Einstein distribution function f_0 gives h its maximal value.]

We are now ready to calculate Φ. By use of (7.1b), (7.5), (7.4b), $\mu'(y) = z$, $dk = b^{-3}\lambda'\lambda^2\,dz\,dg$, (6.14), $\{A_\beta|A_\nu\}_1 = \delta_{\beta\nu}$, and (3.8), it is a simple but very tedious matter to do so. The result is

$$\Phi - \frac{1}{T}q \cong -\frac{k_B a}{2\pi^2 b^2}\sum_{\alpha=1}^{\infty}\sum_{\beta=0}^{\infty}\frac{\alpha!}{(2\alpha+1)!!}a^{\alpha,\beta}\circ c^{\alpha-1,\beta}, \tag{7.8a}$$

where

$$
\begin{aligned}
c^{\alpha,\beta} :&= \frac{\tau_\beta}{\tau_{\beta+2}}a^{\alpha,\beta+2} + \frac{\tau_{\beta-2}}{\tau_\beta}a^{\alpha,\beta-2} + n_{\beta+2}a^{\alpha,\beta+1} \\
&\quad + n_{\beta+1}a^{\alpha,\beta-1} + k_{\beta+2}a^{\alpha,\beta},
\end{aligned} \tag{7.8b}
$$

$$c^{\alpha,\beta} \in \mathrm{Ker}_\alpha\,\mathrm{Tr}, \quad a^{\alpha,-2} = a^{\alpha,-1} = 0. \tag{7.8c}$$

In going from (7.1b) to (7.8) we have utilized

$$\frac{k_B a}{6\pi^2 b^2}\sum_{\beta=0}^{2}\{z|A_\beta\}_1 a^{1,\beta} = \frac{1}{T}q, \tag{7.9}$$

q being the heat flux and T being the kinetic theory temperature.

As a step toward obtaining an estimate of the rhs of (7.8a), let us observe that by (6.13), (6.14e), and the Schwarz inequality we arrive at

$$k_\beta \leqslant \sqrt{\ell_\beta}, \tag{7.10a}$$

$$|n_\beta| \leqslant \min(\sqrt{\ell_\beta}, \sqrt{\ell_{\beta+1}}), \tag{7.10b}$$

$$\frac{\tau_{\beta-2}}{\tau_\beta} \leqslant \min(\sqrt{\ell_\beta}, \sqrt{\ell_{\beta+2}}). \tag{7.10c}$$

Then combining (7.10) with

$$\frac{\alpha!}{(2\alpha+1)!!} < \frac{1}{2}\,\frac{(\alpha-1)!}{(2\alpha-1)!!}, \quad \alpha = 1,2,\ldots, \tag{7.11a}$$

and

$$
\begin{aligned}
|a^{\alpha,\beta}\circ a^{\alpha-1,\nu}| &\leqslant 3|a^{\alpha,\beta}|\,|a^{\alpha-1,\nu}| \\
&\leqslant \tfrac{3}{2}(a^{\alpha,\beta}\circ a^{\alpha,\beta} + a^{\alpha-1,\nu}\circ a^{\alpha-1,\nu}),
\end{aligned} \tag{7.11b}
$$

we may easily find that

$$
\begin{aligned}
&\left|\sum_{\alpha=1}^{\infty}\sum_{\beta=0}^{\infty}\frac{\alpha!}{(2\alpha+1)!!}a^{\alpha,\beta}\circ c^{\alpha-1,\beta}\right| \\
&\quad < \hat{0} := 10\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\frac{\alpha!}{(2\alpha+1)!!}\mathscr{H}_\beta a^{\alpha,\beta}\circ a^{\alpha,\beta},
\end{aligned} \tag{7.12a}
$$

where

$$\mathscr{H}_\beta := \max(1, \sqrt{\ell_{\beta+2}}). \tag{7.12b}$$

It follows from the definition of $\hat{0}$ and the inequality (7.12a) that if $\hat{0} < +\infty$, then f satisfies Condition I and the series on the rhs of (7.8a) converges in the sense of the norm in $\mathbb{E}$.

In summary, the important properties of (6.8) just established should cast light on the far reaching similarity between Grad's method and the one we are presently developing so as to derive and generalize all results that his approach,[6] if it could be specialized by him to the case of a gas of quasiparticles, would deliver. No doubt, we may regard (6.8) as a strict quasiparticle analog of the irreducible variant of Grad's expansion of f in terms of Hermite polynomials and call the Tchebychef–Ikenberry "moments" $a^{\alpha,\beta}$ analogs of the Laguerre–Ikenberry coefficients $b^{\alpha|\beta}$

VIII. FINAL REMARKS

In closing, it is pertinent to summarize the main feature of the analysis.

(1) The derivation of a set of equations of transfer and of the irreducible moment representation of f is only the first step in constructing a kinetic theory of flow phenomena for quasiparticle gases. There still remains the problem of obtaining a solution (or solutions) to Eqs. (4.4a) and the subsequent utilization of these solutions so as to be able to demonstrate that the sum of the series (6.8) of which the Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}$ are gained by (6.11a) is a solution of the original Boltzmann–Peierls equation (4.1). Insofar as the model proposition (3.1) and the simplifying assumptions (3.5) and (5.1) are concerned, research in this direction can be outlined and is under way.

(2) Of particular, even decisive importance in our analysis is the method of defining the weight (6.1a) and of introducing the Tchebychef functions $A_\beta(z;\Theta)$ that culminates

in, and leads to, the approximate formula (7.6) for $h - h_0$ with no cross terms coupling different Tchebychef–Ikenberry coefficients. At this point we wish to stress once more that the old Grad's ideas regarding classical gases have served as useful guides to new methods; conversely, the new theories have helped to clarify the earlier points of view. Physically speaking, the unique advantage of having (7.6) is certain to be elaborate.[27]

(3) Notice that Eq. (7.6) contains complete information about the so-called *second moments* of the equilibrium fluctuations of the Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}$ of f; the fluctuations are supposed to occur around $a^{\alpha,\beta} = 0$ provided the $a^{0,0}$ is being kept equal to zero. Since $a^{\alpha,\beta} = 0$ gives $h - h_0$ the greatest value it can attain for one-particle densities f corresponding to the same gross condition $a^{0,0} = 0$, we may follow in a sense Einstein and Smoluchowski and assume that the (relative) probability density of fluctuations $\mathbb{P}$ is given by

$$\mathbb{P} = \exp[k_B^{-1} V(h - h_0)],$$

V being the volume of the gaseous system. In this way the approximate expression (7.6) for $h - h_0$ defines the simplest acceptable potential for the random variations of $a^{\alpha,\beta}$, or of $N^{\alpha,\beta}$, near the *absolute* equilibrium state. Of course, this becomes a statement with meaning only if we first tell what is meant by probability. There are many ways of assigning mathematically a probability, i.e., a normalized measure in the space of sequences $\{a^{\alpha,\beta}\}$ such that

$$\sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} \frac{\alpha!}{(2\alpha+1)!!} a^{\alpha,\beta} \circ a^{\alpha,\beta} < +\infty,$$

but if expectations according to this probability are to correspond to the interesting ideas of Jou *et al.*,[9,10] we demand a physical explanation of it. Since $\mathbb{P}$ is, roughly speaking, a joint probability density of infinitely many random variables, these problems deserve some more attention in future investigations and will be considered elsewhere.

(4) We should note one more thing. Our decision to study the Tchebychef–Ikenberry expansion (6.8) is considerably strengthened by the fact that the abstract space $\mathbb{L}^2(\mathbb{R}_0^+ \times \mathbb{K}; \mathbb{W}(z;\Theta)dz\,dg)$ invented in Sec. VI A and the Hilbert space $\mathbb{L}^2(\mathbb{E}; f_0(1+f_0)dk)$ in which the linearized Boltzmann–Peierls operator $\mathbb{C}(f_0,\circ)$ is symmetric and nonpositive are *canonically isomorphic*. Now, let $\mathbb{L}_1^2$ be a set consisting of either (i) functions $\psi \in \mathbb{L}^2(\mathbb{E}; f_0(1+f_0)dk)$ such that

$$\int_{\mathbb{E}} \Omega \psi f_0(1+f_0)dk = 0, \text{ or}$$

(ii) functions $\psi \in \mathbb{L}^2(\mathbb{E}; f_0(1+f_0)dk)$ such that

$$\int_{\mathbb{E}} \Omega \psi f_0(1+f_0)dk = 0,$$

$$\int_{\mathbb{E}} k \psi f_0(1+f_0)dk = 0,$$

and denote by $\mathbb{C}_L$ the restriction of $\mathbb{C}(f_0,\circ)$ to $\mathbb{L}_1^2$. Then, recalling both the Chapman–Enskog method[11] and Gurevich's monograph,[3] we know that the problem of calculating a variety of transport coefficients is simply one of solving a Fredholm equation for ψ of the form

$$\mathbb{C}_L \psi = \mathscr{H},$$

$\mathscr{H}$ being given in terms of k, Ω, $\nabla_k \Omega$, T, and the first space derivatives of the conserved moments of f. That this equation has a solution for many common quasiparticle models, is ensured by the Riesz–Schauder theory of completely continuous operators, as presented, for example, in Chap. X of Yosida's book.[28] What is more important for the present purpose, however, is the observation as follows. The specific and well-founded assumption of the expressibility of $\Omega^{-1}\psi$ by the sum of a convergent series of Tchebychef functions $A_\beta(z;\Theta)$ allows one to estimate the expansion coefficients, which are quantities independent of k, through a sequence of approximations usually made when dealing with the integral equation of Fredholm's type (see, e.g., Ref. 11, pp. 124–129). Clearly this is a formulation of the basic step in the argument—that the Tchebychef basis $\{A_\beta;\ \beta = 0,1,...\}$ is very useful in exploiting the Chapman–Enskog procedure as well.

ACKNOWLEDGMENT

We are indebted to Prof. D. Jou for his valuable comments regarding the generalized entropy on the one hand and the Einstein–Smoluchowski fluctuation theory on the other.

APPENDIX: THE TCHEBYCHEF–IKENBERRY EXPANSION FOR THE DISPERSION RELATIONS (2.6) AND (2.7)

A. The power (magnon) model: $\Omega(k,x) = c(x)k^r$, $1 < r \leqslant 2$

We begin with the definitions

$$i := r/(r-1), \quad d := (r+4)/r, \quad \Theta := \hbar c/k_B T, \tag{A1}$$

$$z := \Theta^{i^{-1}} v(k) \quad (\Rightarrow \omega = \hbar\Omega/k_B T = z^i), \tag{A2}$$

$$\mathbb{W}_r(z) := \omega^d e^\omega/(e^\omega - 1)^2 \leqslant \mathbb{M}e^{-Cz}. \tag{A3}$$

If in addition we replace $\mathbb{L}^2(\mathbb{R}_0^+ \times \mathbb{K}; \mathbb{W}(z;\Theta)dz\,dg)$ by $\mathbb{L}^2(\mathbb{R}^+ \times \mathbb{K}; \mathbb{W}_r(z)dz\,dg)$ in Sec. VI A, then by a mere application of the transformations

$$\{S_\beta(z);\ \beta = 0,1,...\} \Rightarrow \{1, z^{-1}, z^\beta;\ \beta = 1,2,...\}, \tag{A4}$$

$$A_\beta(z;\Theta) \Rightarrow A_\beta(z), \tag{A5}$$

$$\{c_{\beta\nu}(\Theta), d_{\beta\nu}(\Theta)\} \to \{c_{\beta\nu}, d_{\beta\nu}\}, \tag{A6}$$

$$b \Rightarrow (r-1)^{1/3}\Theta^{1/r} \tag{A7}$$

we conclude that (6.4) and (6.8)–(6.10) are still valid.

B. The isotropic and dispersionless (phonon) model: $\Omega(k,x) = c(x)k$

In this case we have

$$z := \Theta k, \quad \Theta := \hbar c/k_B T, \tag{A8}$$

$$\mathbb{W}(z) := z^4 e^z/(e^z - 1)^2 \leqslant \mathbb{M}e^{-Cz}, \tag{A9}$$

and

$$\mathbb{L}^2(\mathbb{R}_0^+ \times \mathbb{K};\mathbb{W}(z;\Theta)dz\,dg) \Rightarrow \mathbb{L}^2(\mathbb{R}^+ \times \mathbb{K};\mathbb{W}(z)dz\,dg), \tag{A10}$$

$$\{S_\beta(z);\quad \beta=0,1,...\} \Rightarrow \{z^\beta;\quad \beta=0,1,...\}, \tag{A11}$$

$$A_\beta(z;\Theta) \Rightarrow A_\beta(z), \tag{A12}$$

$$\{c_{\beta\nu}(\Theta),d_{\beta\nu}(\Theta)\} \Rightarrow \{c_{\beta\nu},d_{\beta\nu}\}, \tag{A13}$$

$$b \Rightarrow \Theta. \tag{A14}$$

[1] J. A. Reissland, *The Physics of Phonons* (Wiley, New York, 1973).
[2] J. W. Tucker and V. W. Rampton, *Microwave Ultrasonics in Solid State Physics* (North-Holland, Amsterdam, 1972).
[3] V. L. Gurevich, *Kinetika Fononnykh Sistem* (Nauka, Moscow, 1980).
[4] R. A. Guyer and J. A. Krumhansl, Phys. Rev. **148**, 766 (1966).
[5] H. Beck, P. F. Meier, and A. Thellung, Phys. Status Solidi A **24**, 11 (1974).
[6] H. Grad, *Principles of the Kinetic Theory of Gases, Handbuch der Physik*, edited by S. Flugge (Springer, Berlin, 1958), Vol. XII, p. 205.
[7] Z. Banach and S. Piekarski, J. Math. Phys. **30**, 1804 (1989).
[8] Z. Banach, to be published in Arch. Mech.
[9] L. S. García-Colín, M. López de Haro, R. F. Rodríguez, J. Casas-Vázquez, and D. Jou, J. Stat. Phys. **37**, 465 (1984).
[10] D. Jou, J. Casas-Vázquez, and G. Lebon, Rep. Prog. Phys. **51**, 1105 (1988).
[11] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge U. P., Cambridge, 1953).
[12] $C^p(\mathscr{C})$ is a set of real-valued functions on an open interval $\mathscr{C} \subset \mathbb{R} := (-\infty, +\infty)$ that are p times continuously differentiable. The symbols $\mu'(y)$, $\mu''(y)$, and $\lambda'(z)$ denote $d\mu/dy$, $d^2\mu/dy^2$, and $d\lambda/dz$, respectively.
[13] It is not difficult to see that this solution is unique and that $\lambda \in C^1(\mathbb{R}_0^+)$ and $\lambda'(z) > 0$.
[14] C. Truesdell and R. G. Muncaster, *Maxwell's Kinetic Theory of a Simple Monatomic Gas* (Academic, New York, 1980).
[15] G. Szegö, *Orthogonal Polynomials* (American Mathematical Society, Providence, RI, 1939).
[16] G. Sansone, *Orthogonal Functions* (Interscience, New York, 1959).
[17] J. A. R. Coope and R. F. Snider, J. Math. Phys. **11**, 1003 (1970).
[18] Whether or not a set of functions $\{S_\beta;\ \beta=0,1,...\}$ consists of linearly independent elements, cannot be deduced without first having to determine $\mu(y)$. As an example, for a class of dispersion relations defined by (2.5), we see at a glance that $mS_{2\beta+1} + S_{2\beta+3} = (m+1)S_{2\beta}$, $\beta = 0,1,...$. Hence only the subset $\{S_0,S_1,S_{2\beta};\beta=1,2,...\}$ of the set $\{S_\beta;\ \beta=0,1,...\}$ is linearly independent.
[19] K. Suchy, Ergeb. Exakten Naturwiss. **35**, 103 (1964).
[20] K. Suchy, *Transport Coefficients and Collision Frequencies for Aeronomic Plasmas, Handbuch der Physik*, edited by S. Flugge (Springer, Berlin, 1984), Vol. 49, p. 57.
[21] U. Weinert, S. L. Lin, and E. A. Mason, Phys. Rev. A **22**, 2262 (1980).
[22] U. Weinert, Phys. Rep. **91**, 297 (1982).
[23] J. J. Dijkstra and W. A. van Leeuwen, Physica A **90**, 450 (1978).
[24] If some of the functions S_β are linearly dependent, then we must reject all unnecessary elements, leaving the procedure of orthogonalization otherwise unchanged. As an example, in the case of (2.5) the set $\{S_\beta;\ \beta=0,1,...\}$ should be replaced by its subset $\{S_0, S_1, S_{2\beta}; \beta=1,2,...\}$.
[25] It is well known that a complete set of orthonormal polynomials in three variables x_1, x_2, and x_3 can be obtained by using products of such polynomials in a single variable. Such a procedure lacks symmetry, and there is sometimes an advantage to be gained by expressing the polynomials in a tensor invariant (reducible or irreducible) notation. Thus, insofar as classical gases are concerned, Grad [Commun. Pure. Appl. Math. **2**, 325 (1949)] defined the Hermite function $B^\alpha(x)$, $x \in \mathbb{E}$ which is a reducible symmetric tensor of order α whose components are polynomials of degree α in the coordinates x_1, x_2, x_3 of x. Due to the unique fact that $\exp(|x|^2) = \exp(x_1^2 + x_2^2 + x_3^2) = \exp(x_1^2)\exp(x_2^2)\exp(x_3^2)$, the Hermite "polynomials" $B^\alpha(x) \in \mathbb{E}_s^\alpha$ are orthogonal with respect to the weight $(2\pi)^{-3/2}\exp(-\frac{1}{2}|x|^2)$. Since the same reasoning cannot be repeated for other weight functions, our description of quasiparticle gases is inherently in tensor irreducible notation.
[26] Z. Banach, J. Stat. Phys. **48**, 813 (1987).
[27] Z. Banach and S. Piekarski, J. Math. Phys. **30**, 1826 (1989).
[28] K. Yosida, *Functional Analysis* (Springer, Berlin, 1974).

Irreducible tensor description. III. Thermodynamics of a low-temperature phonon gas

Zbigniew Banach
Institute of Fundamental Technological Research, Department of Fluid Mechanics, Polish Academy of Sciences, Swietokrzyska 21, 00-049 Warsaw, Poland
Slawomir Piekarski
Institute of Fundamental Technological Research, Department of Acoustoelectronics, Polish Academy of Sciences, Swietokrzyska 21, 00-049 Warsaw, Poland

(Received 7 June 1988; accepted for publication 15 March 1989)

Let one assume that the interacting phonon gas, whose behavior is governed by the Boltzmann–Peierls equation, inhabits an insulating crystal at sufficiently low temperature. Then, within the framework of a single acoustic phonon branch and of an isotropic long-wavelength approximation to the dispersion relation, the simplest acceptable version of extended irreversible thermodynamics, based upon the nine-moment system of differential equations for the slow and fast gas-state variables, is carefully investigated. It is clearly demonstrated that, in virtue of the structure simplifications just mentioned, the conceptually different (macroscopic, kinetic, and variational) procedures, which are discussed in this paper, appear complementary to each other. Finally, with the help of a suitable contraction of the nine-moment system of field equations, for Callaway's relaxation model a slightly generalized nonlinear variant of ordinary low-temperature phonon hydrodynamics is explicitly derived.

I. INTRODUCTION

The straightforward way to discuss some of the important aspects of a hydrodynamic behavior of the aggregate of phonons which inhabit an insulating crystal at low temperature is to consider the Boltzmann–Peierls equation[1–5] as a starting point. Of course, the Boltzmann–Peierls equation has somewhat limited validity [see, e.g., Gurevich's monograph[5]], and a more microscopic procedure may be developed in terms of the powerful Green's function techniques.[4,6,7] However, since we are primarily interested in recognizing the nature and origin of fundamental ingredients of continuum thermomechanics, we can easily find on closer examination that this procedure is of no greater generality than the Boltzmann–Peierls method, as employed for long-wavelength collective phonon behavior to be discussed here.[8] Explicitly, unlike all other approaches, the Boltzmann–Peierls description provides almost automatically the physically relevant conservation principles of local hydrodynamics and establishes, due to the unique H theorem, the irreversibility of the evolution of the distribution function f toward an equilibrium state of the Bose–Einstein type.

Following Guyer and Krumhansl,[3] we introduce a number of simplifications, not present in the kinetic equation itself but allowing for more physical insight. Thus we consider an isotropic one-branch Debye model of the form $\hbar\Omega \sim |k|$, where k and $\hbar\Omega$ are the wave vector and the energy of a single phonon in the mode k, respectively. We further appeal to the generally accepted properties of the collision operator J appearing on the rhs of the Boltzmann–Peierls equation; precisely speaking, we recall the standard decomposition[3–5] $J = J_{\mathcal{N}} + J_{\mathcal{R}}$, $|J_{\mathcal{N}}| \gg |J_{\mathcal{R}}|$ in which $J_{\mathcal{N}}$ represents the normal-process collision operator and $J_{\mathcal{R}}$ refers to the momentum-destroying scattering events.

Although in performing formal manipulations in (phonon) thermodynamics it is rather important to make a definite commitment to one or the other of the possible approaches, sometimes a great deal of confusion results from a vacillation between a variety of approaches within a single problem. Having this in mind, our fundamental purpose here is to prove that (i) the phenomenological theory formulated in the spirit of Liu and Müller,[9] (ii) the moment truncation procedure obtained by extending the range of validity of Grad's ideas[10,11] to the case of quasiparticle gases,[12] and (iii) the variational method based upon the entropy maximum principle in the form of Dreyer[13,14] and Banach,[15,16] not only concord perfectly but also bring to light different aspects of the same basic structure which we conveniently call extended irreversible thermodynamics.[17–19]

In the paper we proceed as follows. The nine-moment system of balance equations for the principal moments of f, i.e., for the irreducible fields $N^\alpha \in \mathrm{Ker}_\alpha\,\mathrm{Tr}$, $\alpha = 0, 1, 2$, is the subject of Sec. II. To see its importance, we need only to restate in terms of the physical functions of position x and time t the integral equalities defining N^α. Indeed, inspection shows that the scalar moment N^0 is proportional to the energy density ϵ, whereas the slow vector moment N^1 can be determined either from the quasimomentum density Q or from the heat flux q. The same is not true of the fast irreducible moment N^2. However, regardless of whether or not the N^2 has a readily visible physical significance, its crucial role in both ordinary and extended thermodynamics will be disclosed as well.

Of course, the nine-moment system constitutes only one of the needful axiomatic segments in the purely phenomenological investigations of Sec. III. Nevertheless, the simplest acceptable formulation of extended irreversible thermodynamics of quasiparticle gases, which we introduce by following the Liu–Müller line of thought,[9,17] rests substantially upon it. The main purpose of Sec. IV is to harmonize the Tchebychef–Ikenberry representation of the one-particle density f, as invented in Ref. 12, with the contents of Sec. III. Section V discusses some of the consequences of utilizing for phonon gases the entropy maximum principle and, what is equally important, establishes the contrast and analogy

 0022-2488/89/081826-11$02.50

between the variational method[13–16] and that of the preceding section. Taking Dreyer's theorem,[13,14] in Sec. V a short response to two concluding statements of Eu[20] is also provided. Given the Callaway's model expressions[21] for both $J_{\mathcal{I}}$ and $J_{\mathcal{N}}$, Sec. VI achieves the object of deriving, with the help of a suitable contraction of the nine-moment system of field equations for N^α, $\alpha = 0, 1, 2$, a slightly generalized nonlinear variant of ordinary low-temperature phonon hydrodynamics. We end the work with final remarks in Sec. VII.

The advantages of direct notation for tensors are very well known, so that we use it throughout the paper. Concerning more details, see Ref. 22 and the Appendix.

II. THE NINE-MOMENT SYSTEM OF BALANCE EQUATIONS

A. Prolegomena

We begin with the Boltzmann–Peierls equation that governs the evolution of the one-particle density f:

$$\partial_t f + \nabla_k \Omega \circ \nabla_x f - \nabla_x \Omega \circ \nabla_k f = J(f). \tag{2.1}$$

Here J denotes the collision operator. In Eq. (2.1) the nonequilibrium occupation probability f of the phonon states, irrespective of its obvious dependence upon the wave vector k, is a function of position x and time t. Denoting by $2\pi\hbar$ Planck's constant, the quantity $\hbar\Omega(k,x,t)$ represents the energy of a single quasiparticle in the mode $k\in\mathbb{E}$, $\mathbb{E}$ being a three-dimensional Euclidean vector space.

For the purposes of this paper, we admit the appropriateness of an isotropic, long-wavelength approximation to the generally unknown dispersion relation $\Omega(k,x,t)$ and simply postulate that

$$\Omega(k,x,t) = c(x,t)\kappa, \quad \kappa := |k|, \tag{2.2}$$

where c can be interpreted as the sound velocity. Substituting (2.2) into (2.1) yields

$$\partial_t f + cg\circ\nabla_x f - \kappa W\circ\nabla_k f = J(f), \tag{2.3a}$$

where

$$g := \kappa^{-1}k, \quad W := \nabla c, \quad \nabla := \nabla_x. \tag{2.3b}$$

Within the framework set up here, we define the "principal" moments of f as follows:

$$N^\alpha := \int_{\mathbb{E}} Y^\alpha(g)\kappa f\,dk, \quad \alpha = 0,\ldots,3, \tag{2.4}$$

$Y^\alpha(g)$ being Ikenberry's harmonics.[22] Hence, by no more than a direct reasoning,[12] we arrive at the nine-moment system of balance equations for N^α, $\alpha = 0, 1, 2$:

$$\partial_t N^0 + c\nabla\circ N^1 + 2W\circ N^1 = 0, \tag{2.5a}$$

$$\partial_t N^1 + c\nabla\circ N^2 + \tfrac{1}{3}c\nabla N^0 + W\circ N^2 + \tfrac{4}{3}N^0 W = P^1, \tag{2.5b}$$

$$\partial_t N^2 + c\nabla\circ N^3 + \tfrac{2}{5}c\nabla\wedge N^1 + 2W\wedge N^1 = P^2, \tag{2.5c}$$

where the collision integrals $P^1\in\mathbb{E}$ and $P^2\in\mathrm{Ker}_2\,\mathrm{Tr}$ are given by

$$P^\alpha := \int_{\mathbb{E}} Y^\alpha(g)\kappa J(f)\,dk, \quad \alpha = 1,2. \tag{2.6}$$

Let ϵ and Q be the local densities of energy and quasimomentum, respectively, and suppose that q denotes the heat flux. Then it is easy to show that

$$N^0 = \frac{8\pi^3}{\hbar c}\epsilon, \quad N^1 = \frac{8\pi^3}{\hbar}Q = \frac{8\pi^3}{\hbar c^2}q. \tag{2.7}$$

B. Constitutive representations of the second-order theory

There is an aspect of the equations of transfer (2.5) that leaves something to be desired as they stand—in neither case can the above system of balance equations serve as field equations for the gas-state variables $\mathbb{N} := \{N^0, N^1, N^2\}$, because additional quantities have appeared, namely, the flux moment N^3 and the collision integrals (productions) P^α, $\alpha = 1, 2$. Consequently, these excessive quantities should be related to $\mathbb{N}$ and c in a materially dependent manner. Within the framework of the local constitutive assumption, we postulate that the flux moment N^3 and the productions of N^1 and N^2, denoted by P^1 and P^2, do not depend upon the time-space derivatives of $\mathbb{N}$ and c, i.e., that

$$N^3(x,t) = N^3[\mathbb{N}(x,t); c(x,t)], \tag{2.8a}$$

$$P^\alpha(x,t) = P^\alpha[\mathbb{N}(x,t); c(x,t)], \quad \alpha = 1, 2. \tag{2.8b}$$

Essentially the constitutive theory to be exploited here is appropriate for processes close to (local) equilibrium, where N^1 and N^2 vanish. This is why it makes sense to expand the rhs of (2.8) in powers of N^1 and N^2. Dropping those resulting terms in the expansion which are of higher than second order in the fields N^1 and N^2, we see that

$$N^3 = bN^1\wedge N^2, \tag{2.9a}$$

$$P^1 = -\nu N^1 + \nu_0 N^1\circ N^2, \tag{2.9b}$$

$$P^2 = -\mu N^2 + \mu_0 N^1\wedge N^1 + \mu_1 N^2\cap N^2, \tag{2.9c}$$

where the scalar coefficients, $b,\ldots,\mu_1$ may depend on N^0 and c, i.e., on the energy density ϵ and the sound velocity c. In so doing we have made use of

$$N^3|_{\mathrm{eq}} = 0, \quad P^\alpha|_{\mathrm{eq}} = 0, \quad \alpha = 1, 2. \tag{2.10}$$

Clearly the most restrictive result among (2.9) is the strictly nonlinear dependence of N^3 upon N^1 and N^2. However, given the second-order constitutive theory, there exists no way in which $N^3\in\mathrm{Ker}_3\,\mathrm{Tr}$ can depend on $N^1\in\mathbb{E}$ and/or $N^2\in\mathrm{Ker}_2\,\mathrm{Tr}$, if only the relation (2.8a) is to be isotropic.

Of course, insofar as the purely macroscopic theory is concerned, the method just outlined must lack an ordering parameter which is necessary to sort out terms of various powers in the expansion. Therefore, from the physical standpoint it is not at all clear why a second-order approximation, in particular, should be taken. Why not the third or fourth order or even a higher order approximation? Since the unquestionable preference given to (2.9) cannot be explained except on grounds outside the Liu–Müller framework itself,[9,17] for an ultimate answer we shall look more closely both at the quasiparticle analog of Grad's procedure (Sec. IV A) and at the satisfactory improvement or replacement of ordinary low-temperature phonon hydrodynamics in circumstances beyond its range of validity (Sec. VI B).

C. Motivation

As we know from the important memoirs on the subject,[1–5] the collision operator J can be decomposed into its normal part $J_{\mathscr{N}}$, for which the quasimomentum of the colliding phonons is conserved, and a resistive part $J_{\mathscr{R}}$ that includes umklapp ($\mathscr{U}$) processes and other possible momentum-destroying scattering events. Also, at sufficiently low temperature $|J_{\mathscr{N}}(f)| \gg |J_{\mathscr{R}}(f)|$, i.e., $\mathscr{R}$ processes are very infrequent compared to $\mathscr{N}$ collisions; for example, $\mathscr{U}$ events die out exponentially with decreasing the local absolute temperature T.

Without going into any of the details of the derivation which may easily be reconstructed from the original sources, we conclude that

$$\tau_{\mathscr{R}} := 1/\nu \gg \tau_{\mathscr{N}} := 1/\mu, \tag{2.11}$$

in virtue of the above-mentioned properties of J. We interpret the positive quantities $\tau_{\mathscr{R}}$ and $\tau_{\mathscr{N}}$ as the mean (k-independent) relaxation times. Indeed, in a purely time-dependent problem with no space variation the values of N^1 and N^2 tend, without exception, to the values they would have according to thermostatics, namely, 0, and the rates at which they do so are governed by the parameters $\tau_{\mathscr{R}}$ and $\tau_{\mathscr{N}}$, respectively. Clearly, due to the inequality (2.11), N^1 decays slowly (the slow moment), whereas N^2 decays rapidly (the fast moment).

The underlying idea of linear irreversible thermodynamics is to take the physically relevant fields, viz. the slow moments N^0 and N^1, as independent gas-state variables and supplement the balance equations for them by the constitutive relations of the form (Sec. VI B)

$$N^2 = -\tfrac{2}{5}(c/\mu)\nabla \wedge N^1, \quad P^1 = -\nu N^1. \tag{2.12}$$

The differential equations for N^0 and N^1 arising from such an approach are of first order in t and of second order in x.

While ordinary low-temperature phonon hydrodynamics has been undeniably useful for the analysis of many phenomena in dielectric crystals, it presents some well-known limitations. The first one is that in this theory the response of the system is instantaneous; in fact, the systems have a certain inertia that produces a delay or retardation in their response to any driving force. On the other side, if the mean collision time $\tau_{\mathscr{N}}$ is not very small compared to the time τ_h required for any appreciable macroscopic changes of N, then classical hydrodynamiclike behavior of a low-temperature phonon gas breaks down at all and Müller's extended thermodynamics[9,17] including the fast nonconserved moment N^2 as a gas-state variable could potentially become a subject of intense interest.[23] [The choice of N^2 is in a sense obvious, because the moment N^2 appears naturally in Eq. (2.5b) and, insofar as (2.12) is concerned, plays a crucial role also in ordinary description.]

Further undeniable advantages of dealing with (2.5) rather than (2.5a) and (2.5b) will be recognized and explained in Sec. VI B.

III. EXTENDED IRREVERSIBLE THERMODYNAMICS OF PHONON GASES

A. Entropy principle

If the explicit form of the basic constitutive relations (2.8) is known,[24] we can obtain an explicit set of field equations from (2.5). Every solution of this system is called a *thermodynamic process*. In continuum mechanics the entropy density h and the entropy flux Φ are the auxiliary constitutive quantities which should be determined from N and c in a materially dependent manner. Hence we set,[24] as in (2.8),

$$h(x,t) = h[\mathrm{N}(x,t);c(x,t)], \tag{3.1a}$$

$$\Phi(x,t) = \Phi[\mathrm{N}(x,t);c(x,t)]. \tag{3.1b}$$

The entropy principle states that the inequality

$$\partial_t h + \nabla \circ \Phi \geqslant 0 \tag{3.2}$$

holds for all thermodynamic processes. Thus the entropy principle must restrict the possible constitutive equations. Since every thermodynamic process gives rise to N, c, N^3, P^α ($\alpha = 1, 2$), h, and Φ such as to satisfy (2.5), (2.8), (3.1), and (3.2) simultaneously, the axiom of entropy growth (3.2) can prove nothing about fields that are not so related. But according to Liu's theorem,[25] the new [with respect to (3.2)] inequality

$$\begin{aligned}
&\partial_t h + \nabla \circ \Phi - \Delta^0 \circ (\partial_t N^0 + c\nabla \circ N^1 + 2W \circ N^1) \\
&\quad - \Delta^1 \circ (\partial_t N^1 + c\nabla \circ N^2 + \tfrac{1}{3} c \nabla N^0 \\
&\quad + W \circ N^2 + \tfrac{4}{3} N^0 W - P^1) \\
&\quad - \Delta^2 \circ (\partial_t N^2 + c\nabla \circ N^3 + \tfrac{2}{5} c \nabla \\
&\quad \wedge N^1 + 2W \wedge N^1 - P^2) \geqslant 0
\end{aligned} \tag{3.3}$$

holds for *all*[26] fields $\mathrm{N}(x,t)$ and $c(x,t)$. The tensorial variables $\Delta^\alpha \in \mathrm{Ker}_\alpha\ \mathrm{Tr}$, $\alpha = 0, 1, 2$ just introduced depend locally upon N and c,[24]

$$\Delta^\alpha(x,t) = \Delta^\alpha[\mathrm{N}(x,t);\, c(x,t)], \quad \alpha = 0, 1, 2; \tag{3.4}$$

we call Δ^α Lagrange's multipliers.

Now, using the definition (A3), the inequality (3.3) may be written in an essentially different form as

$$\begin{aligned}
&\left(\frac{\partial h}{\partial N^0} - \Delta^0\right) \circ \partial_t N^0 + \left(\frac{\partial h}{\partial N^1} - \Delta^1\right) \circ \partial_t N^1 + \left(\left\langle \Pi \frac{\partial h}{\partial N^2} \right\rangle - \Delta^2\right) \circ \partial_t N^2 + \left(\frac{\partial h}{\partial c}\right) \circ \partial_t c + \left(\frac{\partial \Phi}{\partial N^0} - \frac{1}{3} c\Delta^1 - c\Delta^2 \circ \frac{\partial N^3}{\partial N^0}\right) \circ \nabla N^0 \\
&\quad + \left(\frac{\partial \Phi}{\partial N^1} - c\Delta^0 I - \frac{2}{5} c\Delta^2 - c\Delta^2 \circ \frac{\partial N^3}{\partial N^1}\right) \circ \nabla N^1 + \left(\sum_{r=1}^{3} e_r \otimes \left\langle \Pi \left[e_r \circ \left\{ \frac{\partial \Phi}{\partial N^2} - cI \otimes \Delta^1 - c\Delta^2 \circ \frac{\partial N^3}{\partial N^2} \right\} \right] \right\rangle \right) \circ \nabla N^2 \\
&\quad + \left(\frac{\partial \Phi}{\partial c} - 2\Delta^0 N^1 - \frac{4}{3} N^0 \Delta^1 - \Delta^1 \circ N^2 - 2N^1 \circ \Delta^2 - c\Delta^2 \circ \frac{\partial N^3}{\partial c}\right) \circ \nabla c + \Delta^1 \circ P^1 + \Delta^2 \circ P^2 \geqslant 0,
\end{aligned} \tag{3.5}$$

where $I \in \mathbb{E}_s^2$ means the contravariant metric (unit) tensor[22] and $\{e_r;\ r = 1, 2, 3\}$, Π, and $\langle M^\alpha \rangle$ stand for the orthonormal basis in $\mathbb{E}$, the symmetrizer, and the traceless part of $M^\alpha \in \mathbb{E}_s^\alpha$, respectively.[22] Since the time-space derivatives $\partial_t N^\alpha$, ∇N^α ($\alpha = 0, 1, 2$), $\partial_t c$, and ∇c can be chosen *arbitrarily* and *independently* of any other term in (3.5), it follows that the quantities in parentheses must vanish lest the inequality (3.5) be violated by some such choices and, in view of these observations, we arrive at

$$\frac{\partial h}{\partial N^0} - \Delta^0 = 0, \quad \frac{\partial h}{\partial N^1} - \Delta^1 = 0, \tag{3.6a}$$

$$\left\langle \Pi \frac{\partial h}{\partial N^2} \right\rangle - \Delta^2 = 0, \quad \frac{\partial h}{\partial c} = 0, \tag{3.6b}$$

$$\frac{\partial \Phi}{\partial N^0} - \frac{1}{3} c\Delta^1 - c\Delta^2 \circ \frac{\partial N^3}{\partial N^0} = 0, \tag{3.6c}$$

$$\frac{\partial \Phi}{\partial N^1} - c\Delta^0 I - \frac{2}{5} c\Delta^2 - c\Delta^2 \circ \frac{\partial N^3}{\partial N^1} = 0, \tag{3.6d}$$

$$\left\langle \Pi \left[e_r \circ \left(\frac{\partial \Phi}{\partial N^2} - cI \otimes \Delta^1 - c\Delta^2 \circ \frac{\partial N^3}{\partial N^2} \right) \right] \right\rangle = 0 \quad (r = 1, 2, 3), \tag{3.6e}$$

$$\frac{\partial \Phi}{\partial c} - 2\Delta^0 N^1 - \frac{4}{3} N^0 \Delta^1 - \Delta^1 \circ N^2 - 2N^1 \circ \Delta^2 - c\Delta^2 \circ \frac{\partial N^3}{\partial c} = 0. \tag{3.6f}$$

There remains the residual inequality:

$$\Delta^1 \circ P^1 + \Delta^2 \circ P^2 \geqslant 0. \tag{3.7}$$

One final point to note is as follows. While the classical Gibbs equation has proved its efficiency as a mathematical basis for proofs of plausible and useful results in thermostatics, the nonequilibrium generalization of it, namely,

$$dh = \Delta^0 \circ dN^0 + \Delta^1 \circ dN^1 + \Delta^2 \circ dN^2, \tag{3.8}$$

cannot be viewed as an essential building block for extended irreversible thermodynamics. Indeed, Eq. (3.8) is a trivial consequence of (3.6a) and (3.6b).

B. Intermediate formulas

For a theory of the constitutive equations (2.9) to be internally consistent, the functions on the rhs of (3.4) should be expanded about the "point" $\{N^0, 0, 0\}$ up to second-order terms in N^1 and N^2 and the relations for h and Φ must include also those parts of the Taylor's expansions which are cubic in N^1 and N^2. Hence Eqs. (3.4) and (3.1) are replaced by

$$\Delta^0 = \lambda_0 + \lambda_1 N^1 \circ N^1 + \lambda_2 N^2 \circ N^2, \tag{3.9a}$$

$$\Delta^1 = \lambda_3 N^1 + \lambda_4 N^1 \circ N^2, \tag{3.9b}$$

$$\Delta^2 = \lambda_5 N^2 + \lambda_6 N^1 \wedge N^1 + \lambda_7 N^2 \cap N^2, \tag{3.9c}$$

$$h = h_0 + h_1 N^1 \circ N^1 + h_2 N^2 \circ N^2 + h_3 (N^1 \wedge N^1) \circ N^2 + h_4 (N^2 \cap N^2) \circ N^2, \tag{3.9d}$$

$$\Phi = \phi_1 N^1 + \phi_2 N^1 \circ N^2 + \phi_3 (N^1 \circ N^1) N^1 + \phi_4 (N^2 \circ N^2) N^1 + \phi_5 (N^1 \circ N^2) \circ N^2, \tag{3.9e}$$

where again the scalar coefficients $\lambda_0, \ldots, \phi_5$ may depend on N^0 and c.[27]

We insert (2.9a) and (3.9) into (3.6) and drop[28] in the result all terms of higher than second order in the fields N^1 and N^2. Then we obtain certain tensorial polynomials in N^1 and N^2 of degree 2, and we conclude from their vanishing that the coefficients of the constitutive representations (2.9a) and (3.9) must obey the following conditions:

$$\frac{\partial h_0}{\partial N^0} = \lambda_0, \quad \frac{\partial h_1}{\partial N^0} = \lambda_1, \quad \frac{\partial h_2}{\partial N^0} = \lambda_2, \tag{3.10a}$$

$$h_1 = \tfrac{1}{2}\lambda_3, \quad h_2 = \tfrac{1}{2}\lambda_5, \tag{3.10b}$$

$$h_3 = \tfrac{1}{2}\lambda_4 = \lambda_6, \quad h_4 = \tfrac{1}{3}\lambda_7, \tag{3.10c}$$

$$\frac{\partial \phi_1}{\partial N^0} = \frac{1}{3} c\lambda_3, \quad \frac{\partial \phi_2}{\partial N^0} = \frac{1}{3} c\lambda_4, \tag{3.10d}$$

$$\phi_1 = c\lambda_0, \quad \phi_2 = c\lambda_3 = \tfrac{2}{5} c\lambda_5, \tag{3.10e}$$

$$\phi_3 = \tfrac{3}{5} c\lambda_1 = \tfrac{1}{5} c\lambda_6, \tag{3.10f}$$

$$\phi_4 = \tfrac{1}{6} cb\lambda_5 = c\lambda_2 - \tfrac{2}{15} c\lambda_7 + \tfrac{1}{3} cb\lambda_5, \tag{3.10g}$$

$$\phi_5 = -\tfrac{4}{15} cb\lambda_5 = c\lambda_4 + \tfrac{2}{3} cb\lambda_5 = \tfrac{2}{5} c\lambda_7 + \tfrac{2}{5} cb\lambda_5, \tag{3.10h}$$

$$\lambda_0 + \tfrac{4}{3} N^0 \lambda_3 = 0, \quad 5\lambda_3 + \tfrac{4}{3} N^0 \lambda_4 = 0. \tag{3.10i}$$

As explicitly stated by Müller *et al.*,[9,17,29] the residual inequality (3.7) cannot be exploited except for up to quadratic terms in N^1 and N^2, because beyond those it is not reliable. It is thus an easy matter to arrive at, using (2.9b), (2.9c), (3.9b), (3.9c), and (3.7),

$$\lambda_3 \nu \leqslant 0, \quad \lambda_5 \mu \leqslant 0. \tag{3.11}$$

In the foregoing, it has been supposed that, when the general constitutive representations (2.8), (3.1), and (3.4) exist—at least for certain definite choices of the values of N and c—and are exact solutions of (3.6) and (3.7), the Taylor's equations (2.9) and (3.9) satisfy (3.6) and (3.7) only approximately, i.e., up to terms that are not of higher than second order in N^1 and N^2. Obvious as these facts are, they can be overlooked in an effort to understand the Liu–Müller theory.

C. Explicit results according to continuum mechanics itself

We postulate, with no loss of physical generality, that

$$\lambda_0 = \frac{\hbar c}{8\pi^3} \frac{1}{T}, \tag{3.12a}$$

$$\phi_1 = \frac{\hbar c^2}{8\pi^3} \frac{1}{T}. \tag{3.12b}$$

Indeed, the determination of λ_0 and ϕ_1 from the local absolute temperature T bears a considerable resemblance to the role that was assigned to some of the constitutive coefficients in the work by Liu and Müller,[9] where, in order to be in complete accord with both thermostatics and the Clausius–Duhem inequality [Truesdell's terminology[30]], very similar identification rules were proposed. The condition (3.12a) for λ_0, when combined with (2.7) and (3.10a), corresponds to the universal requirement of thermostatics that $\partial h_0 / \partial \epsilon = 1/T$. The identification rule (3.12b), immediately resulting from (3.10e) and (3.12a), asserts in turn that in the neighborhood of the state of local equilibrium the entropy flux Φ is approximately equal to q/T, as it should be; in this context, see Eqs. (2.7), (3.9e), and (3.12b).

Before we can proceed, we agree to accept the existence of a relation that gives N^0 as a function of T and c, i.e., of the so-called equation of state:

$$N^0 = N^0(T,c); \tag{3.13}$$

Eq. (3.13) is assumed differentiable and invertible for T as a function of N^0 and c.

By taking (3.12a) and (3.13) in (3.10) we obtain, after a very tedious calculus, the following fairly explicit expressions for the basic and auxiliary constitutive coefficients.

The excessive flux moment,

$$N^3\colon\ b = \frac{45}{28}\frac{1}{N^0}; \tag{3.14}$$

Lagrange's multipliers,

$$\Delta^0\colon\ \lambda_0 = \frac{1}{8}\,\mathbb{A},\quad \lambda_1 = \frac{15}{256}\frac{1}{N^0N^0}\,\mathbb{A}, \tag{3.15a}$$

$$\lambda_2 = \frac{75}{512}\frac{1}{N^0N^0}\,\mathbb{A}; \tag{3.15b}$$

$$\Delta^1\colon\ \lambda_3 = -\frac{3}{32}\frac{1}{N^0}\,\mathbb{A},\quad \lambda_4 = \frac{45}{128}\frac{1}{N^0N^0}\,\mathbb{A}; \tag{3.16}$$

$$\Delta^2\colon\ \lambda_5 = -\frac{15}{64}\frac{1}{N^0}\,\mathbb{A},\quad \lambda_6 = \frac{45}{256}\frac{1}{N^0N^0}\,\mathbb{A}, \tag{3.17a}$$

$$\lambda_7 = \frac{1125}{1792}\frac{1}{N^0N^0}\,\mathbb{A}; \tag{3.17b}$$

Entropy,

$$h\colon\ h_0 = \frac{1}{6}N^0\mathbb{A} + \mathbb{B},\quad h_1 = -\frac{3}{64}\frac{1}{N^0}\,\mathbb{A}, \tag{3.18a}$$

$$h_2 = -\frac{15}{128}\frac{1}{N^0}\,\mathbb{A},\quad h_3 = \frac{45}{256}\frac{1}{N^0N^0}\,\mathbb{A}, \tag{3.18b}$$

$$h_4 = \frac{375}{1792}\frac{1}{N^0N^0}\,\mathbb{A}; \tag{3.18c}$$

Entropy flux,

$$\Phi\colon\ \phi_1 = \frac{1}{8}c\mathbb{A},\quad \phi_2 = -\frac{3}{32}\frac{c}{N^0}\,\mathbb{A}, \tag{3.19a}$$

$$\phi_3 = \frac{9}{256}\frac{c}{N^0N^0}\,\mathbb{A},\quad \phi_4 = -\frac{225}{3584}\frac{c}{N^0N^0}\,\mathbb{A}, \tag{3.19b}$$

$$\phi_5 = \frac{45}{448}\frac{1}{N^0N^0}\,\mathbb{A}; \tag{3.19c}$$

The equation of state,

$$\frac{\partial N^0}{\partial T} - \frac{4}{T}N^0 = 0,\quad N^0(T,c) = \mathbb{C}\left(\frac{\pi^3}{k_B\Theta}\right)^4, \tag{3.20}$$

where

$$\mathbb{A}\colon = \left(\frac{\mathbb{C}}{N^0}\right)^{1/4},\quad \Theta\colon = \frac{\hbar c}{k_B T}. \tag{3.21}$$

Here $\mathbb{B}$ and $\mathbb{C}$ stand for constants of integration and k_B is one of Boltzmann.

It is now evident that the Liu–Müller restrictions are so stringent that even the equation of state, which relates N^0 or ϵ to T and c, cannot be arbitrary. Having obtained (3.14)–(3.21), one should like to be able to say that the kinetic theory, by itself, is needless, insofar as the scalar constitutive coefficients for N^3, Δ^α, h, and Φ are concerned. This is almost true, but not quite so. First, the nine-moment system of balance equations (2.5), which is of great interest in deriving (3.14)–(3.21) from the axiom of entropy growth (3.2), in all probability cannot be demonstrated to be valid on the grounds of the laws of continuum mechanics alone; for the kinetic origin of the nine-moment system (2.5), see Sec. II A. Also, the obscurity of all that relates to the unknown constants $\mathbb{B}$ and $\mathbb{C}$ [and to the unspecified coefficients ν, ν_0, μ, μ_0, and μ_1 in Eqs. (2.9b) and (2.9c)], and what is physically meant by a second-order perturbation theory, etc., calls for other approaches as well.

IV. CONSISTENCY OF THE MOMENT METHOD OF GRAD'S TYPE WITH CONTINUUM MECHANICS

A. The Tchebychef–Ikenberry representation of the one-particle density f

Define a function of k, x, and t by setting

$$f_0(z)\colon = (e^z - 1)^{-1},\quad z\colon = \Theta\cancel{k}, \tag{4.1a}$$

and suppose that

$$\int_{\mathbb{E}} \cancel{k} f_0\, dk = \int_{\mathbb{E}} \cancel{k} f\, dk. \tag{4.1b}$$

We may call the function f_0 as given by (4.1) the local Bose–Einstein density that corresponds to f. By use of (2.4) and (4.1b) we easily calculate N^0 in terms of Θ:

$$N^0 = 4\pi^5/15\Theta^4. \tag{4.2}$$

Then comparison of (4.2) with (3.20) shows that

$$\mathbb{C} = 4k_B^4/15\pi^7. \tag{4.3}$$

In the Appendix of Ref. 12 we have considered in more detail only such nonequilibrium occupation probabilities f as can be expanded in a series of one-dimensional Tchebychef polynomials $A_\beta(z)$ and Ikenberry's harmonics $Y^\alpha(g)$. We have found there that

$$f = f_0\left[1 + \frac{ze^z}{e^z - 1}\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty} a^{\alpha,\beta}(x,t)\circ Y^\alpha(g)A_\beta(z)\right], \tag{4.4}$$

with the scalar function $A_\beta(z)$ being a polynomial of precise degree β in the dimensionless variable $z\in\mathbb{R}^+\colon = (0, +\infty)$. The tensorial coefficients $a^{\alpha,\beta}\in\mathrm{Ker}_\alpha\,\mathrm{Tr}$ occurring in Eq. (4.4) may be called the Tchebychef–Ikenberry expansion coefficients of the distribution function f. The Tchebychef polynomials $A_\beta(z)$ are orthonormal with respect to the weight $\mathbb{W}(z)$ given by

$$\mathbb{W}(z)\colon = z^4e^z/(e^z - 1)^2. \tag{4.5}$$

In other words, if $\delta_{\beta\gamma}$ denotes the Kronecker symbol, then

$$\int_{\mathbb{R}^+} A_\beta(z)A_\gamma(z)\mathbb{W}(z)dz = \delta_{\beta\gamma}. \tag{4.6}$$

The orthogonality conditions for $Y^\alpha(g)$ are[22]

$$\int_{\mathbb{K}} Y^\alpha(g)\otimes Y^\nu(g)dg = \frac{4\pi\alpha!}{(2\alpha+1)!!}\delta_{\alpha\nu}E(\alpha|\alpha), \tag{4.7}$$

where $\mathbb{K}$ is the unit sphere and $E(\alpha|\alpha)\in\mathbb{E}^{2\alpha}$ represents the natural projection.[22]

The Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}$ have a simple representation in terms of $f-f_0$. Beginning with (4.6) and (4.7), we find that

$$a^{\alpha,\beta}=\frac{(2\alpha+1)!!}{4\pi\alpha!}\Theta^4\int_{\mathrm{E}}Y^\alpha(g)\not{k}A_\beta(z)(f-f_0)dk. \quad (4.8)$$

By (2.4), (4.1), (4.2), (4.7), (4.8), and $A_0=\sqrt{15}/2\pi^2$, then, we arrive at

$$a^{0,0}=0, \quad (4.9a)$$

$$a^{\alpha,0}=\frac{\sqrt{15}}{8\pi^3}\frac{(2\alpha+1)!!}{\alpha!}\Theta^4N^\alpha$$
$$=\frac{\pi^2}{2\sqrt{15}}\frac{(2\alpha+1)!!}{\alpha!}\frac{1}{N^0}N^\alpha,$$
$$\alpha=1,2,3. \quad (4.9b)$$

In this way the quantities N^0 and $a^{\alpha,0}$, $\alpha=1,2,3$ are simply an alternative set of moments of f, just as were the moments N^α, $\alpha=0,\ldots,3$ defined in Sec. II A.

It is now natural to ask whether we can derive from (2.8a) a relation of the form (2.9a), say as a limit, *the first term in a series expansion*, or otherwise. The left-hand side of (2.8a) is defined from a solution N^3 of (4.9b) for $\alpha=3$; so are N^1 and N^2 on the right-hand side. Thus, if we are willing to support (2.9a), then we must assume that in the dimensionless version of (2.8a) the following inequality holds[22]:

$$|a^{\alpha,0}|\ll 1, \quad \alpha=1,2. \quad (4.10)$$

Concerning the range of validity of (2.9b) and (2.9c) for Callaway's relaxation model,[21] see Sec. VI B.

B. Implications of the Tchebychef–Ikenberry expansion of *f* on constitutive relations for *h* and Φ

By making use of the Boltzmann–Peierls equation we easily obtain the kinetic analog to the entropy principle (3.2) of just the same form, in which, of course, h and Φ are calculated from f:

$$h=\frac{k_{\mathrm{B}}}{8\pi^3}\int_{\mathrm{E}}\mathbb{F}\,dk, \quad (4.11a)$$

$$\Phi=\frac{k_{\mathrm{B}}c}{8\pi^3}\int_{\mathrm{E}}g\mathbb{F}\,dk, \quad (4.11b)$$

$$\mathbb{F}:=(1+f)\ln(1+f)-f\ln f. \quad (4.11c)$$

The Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}$ in (4.4) are functions of position x and time t, i.e., fields in the ordinary sense of continuum mechanics. In order for the entropy density h and the entropy flux Φ that are defined by (4.11) and (4.4) to satisfy the requirements (3.1) of extended thermodynamics, it is sufficient that the *generators of constitutive equations* $a^{\alpha,\beta}$ be uniquely determined from N and c:

$$a^{\alpha,\beta}(x,t)=a^{\alpha,\beta}[\mathrm{N}(x,t);\ c(x,t)]. \quad (4.12)$$

Implicitly, we have introduced so far the restrictions imposed upon (4.12) by (4.9), but these are merely definitions compatible with (4.12), not sufficient for (4.12) to hold. Clearly, regardless of the problem concerning the nature and origin of (4.12), the existence of the relations (4.12) as they stand is one of many postulates of the present formulation of extended thermodynamics and here, in true Müller fashion,[17] we make no question of them.

Henceforth by the constitutive representations (3.1) we shall mean (3.9d) and (3.9e), i.e., the constitutive representations of the second-order perturbation theory, as they are practically the only cases of (3.1) to which the moment method of Grad's type could conceivably be expected to provide a precise counterpart. However, for a theory of the constitutive relations (4.12) that are not of higher than second order in the fields N^1 and N^2, we see almost immediately that

$$(\alpha>3)\Rightarrow(a^{\alpha,\beta}=0), \quad (4.13)$$

because $a^{\alpha,\beta}\in\mathrm{Ker}_\alpha$ Tr, $N^1\in\mathrm{E}$, and $N^2\in\mathrm{Ker}_2$ Tr. Also, applying (2.9a), (3.14), and (4.9b), we obtain

$$a^{3,0}=\frac{15\sqrt{15}\pi^2}{16}\frac{1}{N^0N^0}N^1\wedge N^2. \quad (4.14)$$

The universal reason for using the simplified version (2.9a) of (2.8a) in deriving (4.14) is as follows. The moment method cannot deliver (4.12), as well as (4.14), although it may—and does—deliver (3.1), as we shall soon see, of course after the constitutive relations (4.12) have been proposed on grounds outside the kinetic theory and the moment method themselves. Thus we set down the following hypothesis.

Hypothesis I: Expanding the rhs of (4.12) in powers of N^1 and N^2 and dropping in the result those terms which are of higher than first order in N^1 and N^2, we expect to find that[31]

$$[(\alpha=0,\ldots,3)\wedge(\beta>0)]\Rightarrow(a^{\alpha,\beta}=0). \quad (4.15)$$

There is certainly no sense in saying that this hypothesis must be derived, for deciding whether it is correct or not, from the Boltzmann–Peierls equation. Rather, we should regard Hypothesis I as a proposition, at the same time compatible with the moment method of Grad's type and independent of the kinetic theory, motivated, however, by the entropy maximum principle (Sec. V) and reinforced maybe by the success of its consequences, which we now proceed to consider.

To this end, let us introduce the following abbreviations:

$$\mathbb{F}_0:=(1+f_0)\ln(1+f_0)-f_0\ln f_0, \quad (4.16a)$$

$$F(z):=\frac{z(e^z+1)}{e^z-1}, \quad (4.16b)$$

$$\vartheta:=\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}a^{\alpha,\beta}(x,t)\circ Y^\alpha(g)A_\beta(z). \quad (4.16c)$$

We first began to study the dependence of $\mathbb{F}$ upon ϑ in Ref. 12. The compact formulas in terms of $a^{\alpha,\beta}$ for both h and Φ, valid in the neighborhood of the state of local equilibrium, were some of the new results we found there. Looking back at Sec. VII of Part II,[12] we can easily show that

$$\mathbb{F}=\mathbb{F}_0+(\mathbb{F}-\mathbb{F}_0)_1+(\mathbb{F}-\mathbb{F}_0)_2+\mathscr{O}(\vartheta^4), \quad (4.17a)$$

where

$$(\mathbb{F}-\mathbb{F}_0)_1:=z^{-2}\mathbb{W}(z)\vartheta-\tfrac{1}{2}z^{-2}\mathbb{W}(z)\vartheta^2, \quad (4.17b)$$

$$(\mathbb{F}-\mathbb{F}_0)_2:=\tfrac{1}{6}z^{-2}F(z)\mathbb{W}(z)\vartheta^3. \quad (4.17c)$$

Then, replacing $\mathbb{F}$ by (4.17a) in (4.11a) and (4.11b), we arrive, after a direct calculus, at

$$h = \tfrac{1}{6}N^0\mathbb{A} + (h-h_0)_1 + (h-h_0)_2 + \cdots, \tag{4.18a}$$

$$\Phi = \tfrac{1}{8}c\mathbb{A}N^1 + (\Phi)_1 + (\Phi)_2 + \cdots, \tag{4.18b}$$

where

$$(h-h_0)_1 := \frac{k_B}{8\pi^3}\int_E (\mathbb{F}-\mathbb{F}_0)_1\, dk = -\frac{15}{16\pi^4}N^0\mathbb{A}\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\frac{\alpha!}{(2\alpha+1)!!} \times a^{\alpha,\beta}\circ a^{\alpha,\beta}, \tag{4.18c}$$

$$(\Phi)_1 := \frac{k_B c}{8\pi^3}\int_E g(\mathbb{F}-\mathbb{F}_0)_1\, dk = -\frac{15c}{8\pi^4}N^0\mathbb{A}\sum_{\alpha=1}^{\infty}\sum_{\beta=0}^{\infty}\frac{\alpha!}{(2\alpha+1)!!} \times a^{\alpha-1,\beta}\circ a^{\alpha,\beta}, \tag{4.18d}$$

$$(h-h_0)_2 := \frac{k_B}{8\pi^3}\int_E (\mathbb{F}-\mathbb{F}_0)_2\, dk = \frac{k_B}{48\pi^3}\int_E z^{-2}F(z)\mathbb{W}(z)\vartheta^3\, dk, \tag{4.18e}$$

$$(\Phi)_2 := \frac{k_B c}{8\pi^3}\int_E g(\mathbb{F}-\mathbb{F}_0)_2\, dk = \frac{k_B c}{48\pi^3}\int_E gz^{-2}F(z)\mathbb{W}(z)\vartheta^3\, dk, \tag{4.18f}$$

$\mathbb{A}$ being a function of N^0, which is derivable from (3.21) and (4.3).

Given (4.13)–(4.15), we may now identify and drop all terms in (4.18c)–(4.18f) of higher than third order in N^1 and N^2. It is, then, no surprise that we are justified in replacing the quantity ϑ, which is calculated from (4.16c), by the simplified one

$$\vartheta_L := (3/4N^0)(N^1\circ Y^1 + \tfrac{5}{2}N^2\circ Y^2), \tag{4.19}$$

insofar as Eqs. (4.18e) and (4.18f) are concerned; of course, in extracting (4.19) from (4.16c) we have utilized (4.9) and $A_0 = \sqrt{15}/2\pi^2$. Using (4.9) and (4.14) with respect to (4.18c) and (4.18d) and applying the identity

$$F(z)\mathbb{W}(z) = -z^{-5}\frac{d}{dz}[z^{-4}\mathbb{W}(z)] \tag{4.20}$$

in (4.18e) and (4.18f), after some analysis based upon (4.13)–(4.15) we obtain the transformation rules for (4.18c)–(4.18f) of the form

$$(h-h_0)_1 \Rightarrow -\frac{15}{16\pi^4}N^0\mathbb{A}\sum_{\alpha=0}^{2}\frac{\alpha!}{(2\alpha+1)!!}a^{\alpha,0}\circ a^{\alpha,0} = -\frac{3}{64}\frac{1}{N^0}\mathbb{A}N^1\circ N^1 - \frac{15}{128}\frac{1}{N^0}\mathbb{A}\,N^2\circ N^2, \tag{4.21a}$$

$$(\Phi)_1 \Rightarrow -\frac{15c}{8\pi^4}N^0\mathbb{A}\sum_{\alpha=1}^{3}\frac{\alpha!}{(2\alpha+1)!!}a^{\alpha-1,0}\circ a^{\alpha,0} = -\frac{3}{32}\frac{c}{N^0}\mathbb{A}N^1\circ N^2 - \frac{225}{1792}\frac{c}{N^0N^0}\mathbb{A}\,(N^2\circ N^2)N^1 - \frac{135}{896}\frac{c}{N^0N^0}\mathbb{A}(N^1\circ N^2)\circ N^2, \tag{4.21b}$$

$$(h-h_0)_2 \Rightarrow \frac{k_B}{48\pi^3}\int_E z^{-2}F(z)\mathbb{W}(z)\vartheta_L^3\, dk = \frac{45}{256}\frac{1}{N^0N^0}\mathbb{A}(N^1\wedge N^1)\circ N^2 + \frac{375}{1792}\frac{1}{N^0N^0}\mathbb{A}(N^2\cap N^2)\circ N^2, \tag{4.21c}$$

$$(\Phi)_2 \Rightarrow \frac{k_B c}{48\pi^3}\int_E gz^{-2}F(z)\mathbb{W}(z)\vartheta_L^3\, dk = \frac{9}{256}\frac{c}{N^0N^0}\mathbb{A}(N^1\circ N^1)N^1 + \frac{225}{3584}\frac{c}{N^0N^0}\mathbb{A}(N^2\circ N^2)N^1 + \frac{225}{896}\frac{c}{N^0N^0}\mathbb{A}\,(N^1\circ N^2)\circ N^2. \tag{4.21d}$$

We should expect the constitutive formulas for h and Φ obtained by the present method and that of Sec. III to agree. Such is the case. In order to see that this conjecture is true, we must substitute (4.21) into (4.18) and compare the resulting expressions with Eqs. (3.9d), (3.9e), (3.18), and (3.19) of Sec. III. Then, if only $\mathbb{B}$ is equal to zero in (3.18a) and $\mathbb{C}$ is identified with $4k_B^4/15\pi^7$ in (3.21), we find that (4.18) and (4.21) deliver just the same constitutive representations for h and Φ as those which follow from a macroscopic approach of the Liu–Müller type. Since Sec. III appeals neither to the functional dependence of h and Φ upon f nor to the Tchebychef–Ikenberry expansion of f, the proof of consistency of both procedures as completed here is by no means of purely academic value.

V. THE ENTROPY MAXIMUM PRINCIPLE

A. Formulation of the problem

According to the tenets of information theory, the estimate of the basic and auxiliary consitutive relations, such as, for example, (2.8), (3.1), (4.12), etc., can be obtained from the distribution function which gives the known moments N^α, $\alpha = 0,1,2$ *correctly* and also *maximizes* the following entropy functional:[13–16]

$$U := \int_E \mathbb{F}(f)dk - \Theta(\Lambda^0 - 1)\circ\left[N^0 - \int_E Y^0(g)\not{k}f\,dk\right] - \Theta\Lambda^1\circ\left[N^1 - \int_E Y^1(g)\not{k}f\,dk\right] - \Theta\Lambda_2\circ\left[N^2 - \int_E Y^2(g)\not{k}f\,dk\right], \tag{5.1}$$

$\Lambda^\alpha(x,t)\in\mathrm{Ker}_\alpha\,\mathrm{Tr}$, $\alpha = 0,1,2$, being nonequilibrium and dimensionless multipliers, which are associated with the constraints

$$N^\alpha = \int_E Y^\alpha(g)\not{k}f\,dk, \quad \alpha = 0,1,2. \tag{5.2}$$

Then, due to the well-known formula

$$\delta U = \int_{\mathbb{K}} \left[\ln\left(\frac{1+f}{f}\right) - z(1-\mathbb{H}) \right] \delta f \, dk \tag{5.3}$$

in which

$$\mathbb{H} := \Lambda^0 \circ Y^0 + \Lambda^1 \circ Y^1 + \Lambda^2 \circ Y^2, \tag{5.4}$$

one obtains

$$f_w := [e^{z(1-\mathbb{H})} - 1]^{-1}, \quad \mathbb{H} < 1, \tag{5.5}$$

for f that makes the entropy functional U maximal. Given the definition (5.4), without further information we may conclude that[22]

$$|\mathbb{H}| \leqslant |\Lambda^0| + |\Lambda^1| + |\Lambda^2|. \tag{5.6}$$

Inserting $f = f_w$ into (5.2) yields, of course, the explicit dependence of $\mathbb{N}$ upon Λ^α, $\alpha = 0,1,2$, practically unmanageable although (under certain additional but judicious restrictions) invertible in principle for the Λ multipliers as nonlinear functions of $\mathbb{N}$:

$$\Lambda^\alpha(x,t) = \Lambda^\alpha[\mathbb{N}(x,t)], \quad \alpha = 0,1,2. \tag{5.7}$$

With regard to the status of (5.5), we have the following (adjustment of) theorem of Dreyer.[13,14]

Theorem: In a gas of quasiparticles with $\Omega = c\cancel{k}$, if for some constant $\mathbb{D}$, $\mathbb{D} < 1$,

$$|\Lambda^0| + |\Lambda^1| + |\Lambda^2| \leqslant \mathbb{D}, \tag{5.8}$$

then[32] the Müller's Lagrange multipliers Δ^α are given by

$$\Delta^\alpha = \frac{k_B}{8\pi^3} \Theta(\delta_{\alpha 0} - \Lambda^\alpha), \quad \alpha = 0,1,2, \tag{5.9}$$

$\delta_{\alpha 0}$ being Kronecker's delta. If in addition the constitutive relations (2.8), (3.1), and (3.4) are calculated from (2.4), (2.6), (4.11), $f = f_w$, and (5.9), then they satisfy the Liu–Müller equations (3.6) and the residual inequality (3.7) is also valid. [For the proof of Dreyer's theorem, see Refs. 13 and 14.]

In a recent paper[20] Eu analyzed the entropy maximum principle in the form of (Müller–Dreyer–)Banach.[17,14,16] His two concluding statements are as follows:

(I) "It...must be concluded that Banach's (variational) method[16] is not useful from the standpoint of irreversible thermodynamics. It...can be inferred that his interpretation of" the rhs of (5.9) "as Lagrange's multipliers is nothing to count on thermodynamically."

(II) "The (variational) method proposed by Banach does not lead to an internally consistent formulation of extended irreversible thermodynamics."

Taking Dreyer's theorem,[13,14] it is now evident that items I and II are *incorrect*. Since the remaining statements of Eu[20] do not refer to the subject of this paper but only to that of Ref. 16, they will be discussed elsewhere.

Although the distribution function as written in the exponential form does not satisfy the Boltzmann (–Peierls) equation, there is a literature in which some exact (unknown) solutions of the equations of moments generated by it have been labeled incorrectly "exact (thermodynamic) solutions" according to the kinetic theory itself.

B. The nonequilibrium phase density f_w in the second-order theory

Since Λ's are the undetermined multipliers, the variational constraints (5.2) must be used to determine them from $\mathbb{N}$. But since they appear in the exponential function for the one-particle density, the variational constraints furnish exponentially nonlinear equations for the undetermined multipliers which cannot be solved exactly and analytically. Thus *regardless of the important fact that by placing $f = f_w$ into (5.2) we do not obtain divergent integrals,*[33] (!!) Eq. (5.5) as it stands must be viewed as a rather untractable proposition that prevents us from pushing the explicit calculations very far. Fortunately, even if the second-order constitutive functions for the rhs of (5.7) were taken, they would give rise to some important features associated with the *nature* and *origin* of Hypothesis I not available in other approaches. Indeed, granted that $|\mathbb{H}| \ll 1$, Eqs. (5.9) and (3.9a)–(3.9c) impose an expansion of (5.5) in a series with respect to $\mathbb{H}$ which is correct to within second-order terms in

$$\overline{N}^\alpha := (3/4N^0) N^\alpha, \quad \alpha = 1,2, \tag{5.10}$$

inclusive ($|\overline{N}^\alpha| \ll 1$), so that by use of (3.15)–(3.17), (3.20), (3.21), and (4.3), we arrive, after a very tedious but straightforward calculus, at

$$f_w \Rightarrow f_N := f_0 \left[1 + \frac{z e^z}{e^z - 1} (\vartheta_L + \vartheta_N) \right] \tag{5.11a}$$

with ϑ_L and ϑ_N given by

$$\vartheta_L := \overline{N}^1 \circ Y^1(g) + \tfrac{5}{2} \overline{N}^2 \circ Y^2(g), \tag{5.11b}$$

$$\begin{aligned} \vartheta_N := {} & \tfrac{1}{6} G(z) (\overline{N}^1 \circ \overline{N}^1 + \tfrac{5}{2} \overline{N}^2 \circ \overline{N}^2) \\ & + G(z) (\overline{N}^1 \circ \overline{N}^2) \circ Y^1(g) \\ & + \tfrac{1}{2} G(z) (\overline{N}^1 \wedge \overline{N}^1 + \tfrac{25}{7} \overline{N}^2 \cap \overline{N}^2) \circ Y^2(g) \\ & + \tfrac{5}{2} F(z) (\overline{N}^1 \wedge \overline{N}^2) \circ Y^3(g) \\ & + \tfrac{25}{8} F(z) (\overline{N}^2 \wedge \overline{N}^2) \circ Y^4(g), \end{aligned} \tag{5.11c}$$

where

$$G(z) := F(z) - 5. \tag{5.11d}$$

It is not difficult to prove that f_N leads to the constitutive relations (2.9a), (3.9d), and (3.9e) which are in *excellent conformability* with the Liu–Müller results (3.14), (3.18), and (3.19). Of course, from the standpoint of extended irreversible thermodynamics, the second-order approximation to f_w *can never be exact*. But we have explained already these problems, and of general interest in this regard is our discussion at the end of Sec. III B.

Elementary inspection shows that f_N satisfies Condition I of Ref. 12, i.e., that

$$\int_{\mathbf{R}^+} \int_{\mathbf{K}} |\hat{f}_N|^2 \mathbb{W}(z) \, dz \, dg < +\infty \tag{5.12a}$$

for

$$\hat{f}_N := \frac{(e^z - 1)}{z e^z} \frac{(f_N - f_0)}{f_0} = \vartheta_L + \vartheta_N. \tag{5.12b}$$

Thus $f = f_N$ has a unique expansion (4.4) and this expansion converges in the *mean* to f_N. Even more, it follows from (4.8) and $f = f_N$ that, insofar as the first-order constitutive functions for the rhs of (4.12) are concerned, the Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}$ of f_N other than $a^{1,0}$ and $a^{2,0}$ are null. The agreeement of the prediction of the variational method with that based upon the Tchebychef–Ikenberry representation (4.4) of the one-particle density and Hypothesis I cannot therefore be regarded as largely coincidental.

VI. CALLAWAY'S MODEL

A. Definition of the collision operator J

In any case, it is clear that the difficulty in estimating the collision coefficients in Eqs. (2.9b) and (2.9c), whether in the nonlinear or in the linear regime, is largely due both to the *intricacy* of the collision operator $J = J_{\mathcal{N}} + J_{\mathcal{R}}$ and to the *necessity*[34] of replacing f by f_N in (2.6). Hence it is very tempting to try guessing model equations with the same basic features as the original Boltzmann–Peierls equation, but allowing for a precise specification of the "nonlinear" coefficients ν_0, μ_0, and μ_1 (in terms of ν and μ) without first having to determine f as an exhibited function of $\mathbb{N}$. One of these equations is that of Callaway,[21] which amounts to writing

$$J_{\mathcal{N}}(f) = -(\mu-\nu)(f-f_1), \quad \mu \gg \nu > 0, \tag{6.1a}$$

$$J_{\mathcal{R}}(f) = -\nu(f-f_0), \tag{6.1b}$$

in order to imitate the rhs of the Boltzmann–Peierls equation and those terms in (2.9b) and (2.9c) which are linear in N^1 and N^2. Of course, in (6.1b) f_0 is the local Bose–Einstein density that corresponds to f, and f_1 is defined by

$$f_1 := [e^{z(1-\mathbb{H}_1)} - 1]^{-1}, \tag{6.2a}$$

where

$$\mathbb{H}_1 := \Lambda_1^0 \circ Y^0 + \Lambda_1^1 \circ Y^1, \quad \mathbb{H}_1 < 1. \tag{6.2b}$$

Here the quantities $\Lambda_1^0(x,t) \in \mathbb{R}$ and $\Lambda_1^1(x,t) \in \mathbb{E}$ are such that f_1 gives the moments N^0 and N^1 of f,

$$\int_{\mathbb{E}} Y^\alpha(g) k (f-f_1) dk = N^\alpha - \int_{\mathbb{E}} Y^\alpha(g) k f_1 \, dk = 0, \quad \alpha = 0,1. \tag{6.3}$$

The integral relations (4.1b) and (6.3) show us immediately that

$$J(f_0) = 0, \quad \int_{\mathbb{E}} k J(f) dk = 0. \tag{6.4}$$

Thus we have established some of the needful properties of $J = J_{\mathcal{N}} + J_{\mathcal{R}}$. More important, however, is the fact that, by

$$\begin{aligned}\int_{\mathbb{E}} J(f) \ln\left(\frac{1+f}{f}\right) dk &= (\mu-\nu) \int_{\mathbb{E}} [f_1(1+f) - f(1+f_1)] \\ &\quad \times \ln\left[\frac{f_1(1+f)}{f(1+f_1)}\right] dk \\ &\quad + \nu \int_{\mathbb{E}} [f_0(1+f) - f(1+f_0)] \\ &\quad \times \ln\left[\frac{f_0(1+f)}{f(1+f_0)}\right] dk \geqslant 0,\end{aligned} \tag{6.5}$$

Callaway's equation

$$\partial_t f + c g \circ \nabla_x f - k W \circ \nabla_k f = -(\mu-\nu)(f-f_1) - \nu(f-f_0) \tag{6.6}$$

satisfies an H theorem automatically.

The simplicity of (6.6) is only apparent. Indeed, since, for example, the fields Λ_1^0 and Λ_1^1 have to be determined from the moments N^0 and N^1 of f, f_1 depends functionally upon f, i.e., $f_1 = \bar{f}_1[f]$ for some $\bar{f}_1[\circ]$, and Callaway's equation is extremely nonlinear. Thus it is not yet clear whether (6.6) really brings a decisive simplification in dealing with complicated flow problems. In the neighborhood of the state of local equilibrium, however, Callaway's model offers definite advantages. To this end, let us expand $\bar{f}_1[f_0 + (f-f_0)]$ about the "point" $f_0 = \bar{f}_1[f_0]$ by writing[35]

$$f_1 = f_0\{1 + [z e^z/(e^z-1)] [\cdots + \tfrac{1}{2} F(z) (\bar{N}^1 \wedge \bar{N}^1) \circ Y^2(g) + \cdots]\}, \tag{6.7}$$

where the ellipses in (6.7) stand for the remaining unspecified terms of no greater interest in calculating the collision coefficients ν_0, μ_0, and μ_1 in (2.9b) and (2.9c). With the help of (2.6), $J = J_{\mathcal{N}} + J_{\mathcal{R}}$, (6.1), and (6.7), it is then easy to show that

$$\nu_0 = \mu_1 = 0, \quad \mu_0 = (3/4N^0)(\mu-\nu). \tag{6.8}$$

B. Transition to ordinary low-temperature phonon hydrodynamics

If we recall (3.14) and make use of (2.5), (2.9), and (6.8), the nine-moment system of field equations for $\mathbb{N}$ can be shown to take the following form:

$$\partial_t N^0 + c\nabla \circ N^1 + 2W \circ N^1 = 0, \tag{6.9a}$$

$$\partial_t N^1 + c\nabla \circ N^2 + \tfrac{1}{3} c \nabla N^0 + W \circ N^2 + \tfrac{4}{3} N^0 W = -\nu N^1, \tag{6.9b}$$

$$\begin{aligned}\partial_t N^2 + \tfrac{45}{28} c \nabla \circ ((1/N^0) N^1 \wedge N^2) + \tfrac{2}{5} c \nabla \wedge N^1 + 2W \wedge N^1 \\ = -\mu N^2 + (3/4N^0)(\mu-\nu) N^1 \wedge N^1.\end{aligned} \tag{6.9c}$$

The balance equations (6.9a) and (6.9b) remain valid in ordinary low-temperature phonon hydrodynamics, in which, of course,

$$\tau_{\mathcal{N}} = 1/\mu \ll \tau_{\mathcal{R}} = 1/\nu, \tag{6.10}$$

but (6.9c) allows us to formulate a constitutive relation for N^2. To see this, let us consider a flow problem in which the relaxation (reference) time $\tau_{\mathcal{N}}$ is short compared to the time τ_h required for any appreciable macroscopic changes of $\mathbb{N}$ (then $\ell_h := c\tau_h$ represents some macroscopic length characterizing the spatial variation of $\mathbb{N}$):

$$\tau_{\mathcal{N}} \ll \tau_h. \tag{6.11}$$

Write Eq. (6.9c) in the form

$$\partial_t N^2 + M^2 + \mu N^2 = 0, \tag{6.12a}$$

where

$$M^2 := \tfrac{2}{5}c\nabla\wedge N^1 - \frac{3}{4N^0}\mu\left(1-\frac{\nu}{\mu}\right)N^1\wedge N^1 + \tfrac{45}{28}c\nabla\circ\left(\frac{1}{N^0}N^1\wedge N^2\right) + 2W\wedge N^1 \quad (6.12b)$$

is, by (6.11), almost constant in a time comparable to $\tau_{\mathcal{N}}$. Over such a short period of time, we obtain, except for a very small error, the solution of the differential equation (6.12a) as consisting of a transient plus a steady state, i.e.,

$$N^2(t) = \mathbb{C}^2\exp(-\mu t) - (1/\mu)M^2,$$
$$\mathbb{C}^2 \in \mathrm{Ker}_2\,\mathrm{Tr}, \quad (6.13)$$

$\mathbb{C}^2$ being a "quasiconstant." But the important scale of $\bar{t} := t/\tau_{\mathcal{N}}$ is of order unity, and, no matter what value N^2 has initially, within a short time $t \cong \tau_{\mathcal{N}}$, the N^2 approaches its quasiequilibrium value. Thus (6.13) becomes

$$N^2 = -(1/\mu)M^2. \quad (6.14)$$

Now, if $|\overline{N}^2| \ll 1$ and the variation of the sound velocity c over ℓ_h is very small ($c^{-1}\ell_h|W| = c^{-1}\ell_h|\nabla_x c| \ll 1$)—as is almost always the case—then we can in the first approximation neglect the last two terms on the rhs of (6.12b). If in addition we replace the second term on the rhs of (6.12b), in which $\nu/\mu \ll 1$, by $-(3/4N^0)\mu N^1\wedge N^1$, Eq. (6.14) just reduces to[36]

$$N^2 = -\tfrac{2}{5}(c/\mu)\nabla\wedge N^1 + (3/4N^0)N^1\wedge N^1. \quad (6.15)$$

Since

$$\tfrac{2}{5}(c/\mu)|\nabla\wedge N^1| \propto (\tau_{\mathcal{N}}/\tau_h)N^0|\overline{N}^1| \quad (6.16a)$$

and

$$(3/4N^0)|N^1\wedge N^1| \propto N^0|\overline{N}^1|^2, \quad (6.16b)$$

there is no reason at all to conclude, as typically in other approaches on the subject, that the leading term on the rhs of (6.15) is the first one. Only on the restrictive assumption that $|\overline{N}^1| \ll \tau_{\mathcal{N}}/\tau_h$, the well-known relation (2.12) of ordinary low-temperature phonon thermodynamics[1–6] and the nonlinear constitutive equation (6.15) become practically indistinguishable. The far weaker inequality $|\overline{N}^1| \leqslant \tau_{\mathcal{N}}/\tau_h$ provides in turn not only a range of validity of (6.15), but also a domain of disagreement ($|\overline{N}^1| \cong \tau_{\mathcal{N}}/\tau_h$) between (2.12) and (6.15).

Given extended irreversible thermodynamics of far-from-equilibrium processes, the question of contraction can be considered in some generality. However, while the approximations (2.9b) and (2.9c) which yield (6.15) are likely to significantly increase the accuracy, other possibilities, such as higher order expansions or the *literal utilization* of the distribution function as written in the exponential form, would not seem to be justified, and one should consider the Boltzmann (–Peierls) equation and its "very hydrodynamic solutions." But at the present time this is a matter of taste rather than mathematics and physics. [By way of digression, our circumspect generalization (6.15) of (2.12) is not to be confused with the concept of nonlinear transport coefficients.]

Since in general $\tau_{\mathcal{N}}$ is not small compared to a representative time τ_h, one would expect Eqs. (6.9a), (6.9b), and (6.15) to demonstrate, at least to some extent, the effect of an overlap between the two time scales, viz., the collision time $\tau_{\mathcal{N}}$ and the macroscopic decay time τ_h. Of course, if also the condition $\tau_{\mathcal{N}} \ll \tau_h$ is not satisfied, then the nine-moment system differs considerably from that just mentioned and should give better results!

VII. FINAL REMARKS

In this paper we have treated only a simple quasiparticle gas: a gas composed of phonons, each characterized by $\Omega = c\not{k}$. As the dispersionless case is certainly the easiest to handle mathematically as well as the simplest in concept, we have a very persuasive reason[12] to conjecture that many results concerning it will have their strict analogs for *other* isotropic dispersion relations with the group velocity $|\nabla_k\Omega|$ essentially depending upon $\not{k}$. It must be stressed, however, that such a generalized discussion is expected to be exceedingly complex.

One of the central ideas for drawing inferencies from our axioms consists in an extension of the scope of Grad's moment procedure. We hope not to be too far off the mark in naming the expansion (4.4) of f, and the expansion (6.8) in Ref. 12, a quasiparticle counterpart to that of Grad fashioned by mathematical apparatus such as Hermite functions or Laguerre polynomials and Ikenberry's harmonics.[22] Indeed, if just as in Grad's approach we wish to "diagonalize" the entropy density h, then by the choice of a local Bose–Einstein density f_0 we have simply *no* alternative to the use of the weight (4.5) and of the Tchebychef–Ikenberry representation (4.4) of f.

Comparing the present work with the older one, we should take note of a difference stemming from the fact that the series (4.4) is completely independent of the exact (unknown) form of the collision operator J which changes, to some extent accidentally, from one quasiparticle system to another. Finally, it is important to remark that the general results established in Secs. VI and VII of Part II[12] give us a more "microscopic" basis for extended irreversible thermodynamics, mainly as a direct or indirect consequence of the above-mentioned effect of diagonalization of h.

If the theory of classical and quasiparticle gases, as developed in our three papers,[22,12] serves as a stimulus for further research, the efforts put forth in the preparation of the text will have been rewarded.

APPENDIX: DEFINITION OF A MAPPING $\varphi_{\alpha\beta}$: $\mathrm{E}^\beta \to \mathrm{E}^{\alpha+\beta}$

Choose an orthonormal basis $\{e_1, e_2, e_3\}$ in E and set

$$e_{r_1\cdots r_\alpha} := e_{r_1}\otimes\cdots\otimes e_{r_\alpha}. \quad (\mathrm{A1})$$

Let M^α: $\mathrm{E}^\beta \Rightarrow \mathrm{E}^\alpha$ be a differentiable tensor function. Then a mapping $\varphi_{\alpha\beta}$: $\mathrm{E}^\beta \Rightarrow \mathrm{E}^{\alpha+\beta}$, for which we write

$$\varphi_{\alpha\beta}(M^\beta) := \frac{\partial M^\alpha}{\partial M^\beta}, \quad (\mathrm{A2})$$

is defined by

$$\frac{\partial M^\alpha}{\partial M^\beta} := \sum_{r_1\cdots s_\beta=1}^{3} \frac{\partial M^\alpha_{r_1\cdots r_\alpha}}{\partial M^\beta_{s_1\cdots s_\beta}} e_{r_1\cdots s_\beta}, \quad (\mathrm{A3})$$

where the coefficients $M^\alpha_{r_1\cdots r_\alpha}$ and $M^\beta_{s_1\cdots s_\beta}$ are components of M^α and M^β, respectively.

[1]R. E. Peierls, Ann. Phys. **3**, 1055 (1929).
[2]R. E. Peierls, *Quantum Theory of Solids* (Oxford U.P., Oxford, 1955).
[3]R. A. Guyer and J. A. Krumhansl, Phys. Rev. **148**, 766 (1966).
[4]H. Beck, P. F. Meier, and A. Thellung, Phys. Status Solidi A **24**, 11 (1974).
[5]V. L. Gurevich, *Kinetika Fononnykh Sistem* (Nauka, Moscow, 1980).
[6]H. Beck, Phys. Kondens. Materie **12**, 330 (1971).
[7]K. H. Michel and F. Schwabl, Phys. Kondens. Materie **11**, 144 (1970).
[8]The most significant application of the microscopic theory, being outside the direct scope of these investigations, would seem to be in problems which require some details concerning the precise derivation of the dispersion relations and of the transition rates.
[9]I.-S. Liu and I. Müller, Arch. Rat. Mech. Anal. **83**, 285 (1983).
[10]H. Grad, Commun. Pure Appl. Math. **2**, 331 (1949).
[11]H. Grad, *Principles of the Kinetic Theory of Gases, Handbuch der Physik*, edited by S. Flugge (Springer, Berlin, 1958), Vol. XII, p. 205.
[12]Z. Banach and S. Piekarski, J. Math. Phys. **30**, 1816 (1989).
[13]W. Dreyer, Maximization of Entropy for Relativistic and Degenerate Gases in Non-Equilibrium, in *Proceedings of the ISIMM Symposium on Kinetic Theory and Extended Thermodynamics, I*, edited by I. Müller and T. Ruggeri (Pitagora, Bologna, 1987), p. 107.
[14]W. Dreyer, J. Phys. A: Math. Gen. **20**, 6505 (1987).
[15]Z. Banach, "The maximum entropy principle and the moment truncation procedure, in *Proceedings of the 5th International Symposium on Continuum Models of Discrete Systems*, edited by A. J. M. Spencer (Balkema, Rotterdam, 1987), p. 111.
[16]Z. Banach, Physica A **145**, 105 (1987).
[17]I. Müller, *Thermodynamics* (Pitman, London, 1985).
[18]L. S. García-Colín, M. López de Haro, R. F. Rodríguez, J. S. Casas-Vázquez, and D. Jou, J. Stat. Phys. **37**, 465 (1984).
[19]D. Jou, J. Casas-Vázquez, and G. Lebon, Rep. Prog. Phys. **51**, 1105 (1988).
[20]B. C. Eu, Physica A **152**, 420 (1988).
[21]R. A. Guyer and J. A. Krumhansl, Phys. Rev. **148**, 778 (1966).
[22]Z. Banach and S. Piekarski, J. Math. Phys. **30**, 1804 (1989).
[23]Initially, extended irreversible thermodynamics started out in the papers by Müller, Jou, and others as an effort to have the system of field equations which does not imply that the touch of a heat pulse to a body at one point is felt immediately all over the body (the paradox of heat conduction).
[24]The general constitutive relations (2.8), (3.1), and (3.4) are not to be confused with their perturbation theory replacements (2.9) and (3.9).
[25]I.-S. Liu, Arch. Rat. Mech. Anal. **46**, 131 (1972).
[26]Here $c(x,t)$ is regarded as assignable arbitrarily in principle, while in any particular application it should be specified uniquely as a part of the definition of the problem.
[27]It is possible to prove that the coefficients $\lambda_0,...,\lambda_7$, $h_0,...,h_4$, $c^{-1}\phi_1,...,c^{-1}\phi_5$ are functions of N^0 alone. Concerning more details on the subject, see Sec. III C.
[28]Since N^3 and Δ^α, $\alpha = 0,1,2$, are all only given here to within second-order terms inclusive, higher order terms in (3.6) would not be reliable.
[29]I. Müller and K. Wilmanski, Rheol. Acta **25**, 335 (1986).
[30]C. Truesdell, *Six Lectures on Modern Natural Philosophy* (Springer, Berlin, 1966).
[31]Since the Tchebychef–Ikenberry coefficients $a^{\alpha,\beta}$ all vanish in the state of local equilibrium and $a^{\alpha,\beta}\in\mathrm{Ker}_\alpha$ Tr, the implication (4.15) holds trivially for $\alpha = 0$ and $\alpha = 3$.
[32]In fact, the inequality (5.8) is the only reasonable restriction as yet placed upon the values of the Λ multipliers.
[33]Given a gas composed of material molecules, the fundamental question whether or not the exponential form of f leads to divergent integrals will be discussed elsewhere. In all but trivial cases the answer is "yes" (!!).
[34]Define a new quantity f_L by casting away from (5.11a) the nonlinear term ϑ_N. Since the function f_L parallels that of Grad for a classical rarefied gas, we may call it, to some extent incorrectly, the nine-moment approximation to f. The utilization of f_L in place of f given by (4.4) is of interest in determining from (2.6) the collision coefficients ν and μ in Eqs. (2.9b) and (2.9c). If we wish to calculate the remaining collision coefficients, namely, ν_0, μ_0, and μ_1, then the nine-moment approximation to f other than f_L should be proposed, and use of f_N may seem at present unavoidable, except in the case of Callaway's model.
[35]Roughly speaking, Eq. (6.7) tells us that the full Callaway's collision element $-(\mu-\nu)(f-f_1)-\nu(f-f_0)$, i.e., $-\mu(f-f_0)+(\mu-\nu)(f_1-f_0)$, is functionally expanded around local equilibrium up to second-order terms in $\delta f:=f-f_0$. Thus Eq. (6.7) is not to be confused with the transformation rule (5.11a) of unquestionable importance only in discussing extended irreversible thermodynamics.
[36]Considering (6.9a), (6.9b), and (6.15) as a starting point, it is possible to prove that the nonlinear constitutive equation (6.15) is compatible both with the entropy principle (in a sense made precise by a second-order perturbation theory) and with the variational method which appeals to the four-moment representation of a state.

Chemical potential of a D-dimensional free Fermi gas at finite temperatures

M. Howard Lee
Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602

(Received 7 February 1989; accepted for publication 12 April 1989)

The chemical potential of a D-dimensional free Fermi gas at low temperatures has been obtained using some ideas due to Barker and Blankenbechler [J. Math. Phys. **27**, 302 (1986); Am. J. Phys. **25**, 279 (1957)]. There is an interesting even–odd-dimensional effect in the behavior of the chemical potential, also observed in other properties of a free Fermi gas, such as the susceptibility. For $D = 2$, it is possible to give a closed form expression, thus valid both in the high- and low-temperature regions, probably not possible in any other even or odd dimensions.

I. INTRODUCTION

In an interesting paper, Barker[1] shows how to obtain a series representation for, among others, the chemical potential of a three-dimensional free Fermi gas at finite temperatures. He uses a method that is an alternative to the classic one due to Sommerfeld,[2] found in almost every text on statistical mechanics.[3] He has since applied it to study related physical problems.[4] Barker obtains the chemical potential very elegantly, via a contour integration, partly combining an idea due to Blankenbechler.[5,6]

In this brief paper we apply the ideas of Barker and Blankenbechler to obtain the chemical potential of a D-dimensional free Fermi gas at finite temperatures. We find that the behavior of the chemical potential very much depends on whether the dimensions D are even or odd numbered. Such an even–odd effect in the physical behavior has been noted in other properties of a free Fermi gas.[7] If $D = 2$, the expression for the chemical potential can be given in closed form, unlike in any other dimensions. In recent years there has been an interest in the physics of low dimensions, stimulated by device fabrication. Also, high dimensions are regarded as a domain where mean-field theories become valid. Hence our D-dimensional generalization may not be without some intrinsic interest.

II. CHEMICAL POTENTIAL IN D DIMENSIONS

The number of particles N in a volume L^D is given by the following well-known expression[3,7]

$$N = 2\sum_k n_k \to 2\left(\frac{L}{h}\right)^D \int d^Dk\, n(k)$$
$$= 2\left(\frac{L}{h}\right)^D u_D \int_0^\infty dk\, k^{D-1} n(k), \tag{1}$$

where

$$u_D = 2(\Gamma(1/2))^D/\Gamma(D/2) \tag{2}$$

and

$$n(k) = (1 + z^{-1}e^{\beta\epsilon})^{-1}, \tag{3}$$

where $\epsilon = k^2/2m$, m is the particle mass; $z = e^\alpha$, $\alpha = \beta\mu$, where $\beta = 1/k_BT$, T temperature, μ is the chemical potential, and h is the Planck constant.

Writing the density as $\rho_D = N/L^D$, one can express (1) more conveniently as

$$\rho_D = (2m/h^2)^{D/2} u_D \int_0^\infty d\epsilon\, \epsilon^{-1+D/2} n(\epsilon). \tag{4}$$

Since the density is related to the Fermi energy ϵ_F as

$$\rho_D = 2/\Gamma(1 + D/2)\cdot(2\pi m\epsilon_F/h^2)^{D/2}, \tag{5}$$

one may further express (4) as follows:

$$(\beta\epsilon_F)^{D/2} = (D/2)\int_{-\alpha}^\infty dx\,\frac{(\alpha + x)^{-1+D/2}}{1 + e^x} \tag{6a}$$

$$= \int_{-\alpha}^\infty dx\,\frac{e^x(\alpha + x)^{D/2}}{(1 + e^x)^2}. \tag{6b}$$

For D even, (6a) is found more useful for analysis; but for D odd, (6b) is more useful.

III. ODD-NUMBERED DIMENSIONS

First we shall evaluate the integral (6b) when D is an odd number. Following Blankenbechler, we write

$$(\alpha + x)^{D/2} = e^{x\delta}\alpha^{D/2} \equiv e^{x\delta}F(\alpha), \tag{7}$$

where $\delta = \partial/\partial\alpha$. Note that this idea does not work if D is an even number, since then (7) is a polynomial of finite degree. Next, let $s = e^x$. Then, (6b) becomes

$$(\beta\epsilon_F)^{D/2} = \int_{z^{-1}}^\infty ds\,\frac{s^\delta}{(1 + s)^2}F(\alpha). \tag{8}$$

At low temperatures $z^{-1} \approx 0$. Thus one may replace the lower limit by zero. Again this kind of approximation is not permitted if D is an even number. Since D is an odd number, the integrand of (8) has a branch cut along, say, the positive real axis. Otherwise it is analytic, except at $s = -1$. Hence one may evaluate the integral by a contour integration, as was done by Barker. However, it is simpler to change the variable, e.g., $s = y/(1 - y)$, thereby putting (8) into a standard form of the Beta function[8]:

$$(\beta\epsilon_F)^{D/2} = \int_0^1 dy\, y^\delta(1 - y)^{-\delta}\cdot F(\alpha)$$
$$= (\pi\delta/\sin\pi\delta)\cdot\alpha^{D/2}. \tag{9}$$

 0022-2488/89/081837-03$02.50

Finally,

$$\beta\epsilon_F = \alpha[1 + (\pi^2/6.2^2)D(D-2)\alpha^{-2} + (7\pi^4/360.2^4) \times D(D-2)(D-4)(D-6)\alpha^{-4} + \cdots]^{2/D}$$
$$= \alpha[1 + (\pi^2/12)(D-2)\alpha^{-2} + (\pi^4/2.6!)(D-2) \times (D^2 - 25D + 74)\alpha^{-4} + \cdots]. \quad (10)$$

The above expansion may now be reverted:

$$\alpha = x[1 - (\pi^2/12)(D-2)x^{-2} - (\pi^4/2.6!) \times (D-2)(D-6)(D-9)x^{-2} - \cdots], \quad (11)$$

where $x = \beta\epsilon_F$.

Observe the appearance of $(D-2)$ in the coefficients of expansion. Because of this factor, the chemical potential in $D = 1$ is distinguished from that in all other odd-numbered dimensions. Initially it increases with T, whereas the others decrease with T. As $T \to \infty$, they must, however, all become negative. Below we give a few examples, mainly for comparison with known results[9]:

$$\mu/\epsilon_F = 1 + (\pi^2/12)x^{-2} + (\pi^4/36)x^{-4} + \cdots \quad (D=1), \quad (12)$$

$$\mu/\epsilon_F = 1 - (\pi^2/12)x^{-2} - (\pi^4/80)x^{-4} - \cdots \quad (D=3), \quad (13)$$

$$\mu/\epsilon_F = 1 - (\pi^2/4)x^{-2} - (\pi^4/120)x^{-4} - \cdots \quad (D=5). \quad (14)$$

For odd-numbered dimensions, the chemical potential is an analytic function of T and it is regular at $T = 0$.

IV. EVEN-NUMBERED DIMENSIONS

When D is an even number, the idea of Blankenbechler cannot be applied. The factor $(\alpha + x)^{D/2}$ in the integral (6b) is now meromorphic, hence analytic everywhere. It is more convenient to start with (6a). One can then obtain an expression in closed form if $D = 2$ and nearly so if $D = 4$. We shall consider the two cases separately.

A. $D=2$

From (6a), it follows directly that

$$\beta\epsilon_F = \int_{-\alpha}^{\infty} dx \frac{1}{1+e^x} = \int_{z^{-1}}^{\infty} dy \left(\frac{1}{y} - \frac{1}{y+1}\right) = \ln(1+z). \quad (15)$$

Hence

$$e^\alpha = -1 + e^x, \quad (16)$$

where $x = \beta\epsilon_F$. Thus we obtain for low temperatures

$$\beta\mu = x - e^{-x} - (1/2)e^{-2x} - \cdots. \quad (17)$$

The above solution (16) is obtained without any assumption on β, hence it is valid for any β, including $\beta \to 0$. Observe that when $\beta \to 0$ (or $x \to 0$), the chemical potential does become negative. In fact, one gets a simple expansion for high temperatures as well:

$$\beta\mu = \ln x + x/2 + x^2/24 + \cdots, \quad (18)$$

which one may recognize as a classical form if x is replaced by $\rho\lambda^2$, where λ is the thermal de Broglie wavelength. Thus one can obtain (16) also by formally summing a high-temperature expansion solution of (4).

B. $D=4$

From (6a), we have

$$(\beta\epsilon_F)^2 = 2(\ln z)(\ln(1+z)) + 2\int_{-\alpha}^{\infty} dx \frac{x}{1+e^x}. \quad (19)$$

The integral on the right-hand side of (19) cannot be expressed in terms of elementary functions, but it can be given a series expansion as follows:

$$2\int_{-\alpha}^{\infty} dx \frac{x}{1+e^x} = (\ln(1+z^{-1}))^2 - (\ln z)^2 + 2f(z^{-1}), \quad (20)$$

where

$$f(z^{-1}) = \sum_{n=1}^{\infty} n^{-2}(1+z^{-1})^{-n}. \quad (21)$$

Substituting (20) in (19), we get

$$(\beta\epsilon_F)^2 = (\ln(1+z))^2 + 2f(z^{-1}). \quad (22)$$

For $z^{-1} \to 0$, one may replace $f(z^{-1})$ by $f(0) = \pi^2/6$, i.e.,

$$(\ln(1+z))^2 \approx x^2 - \pi^2/3, \quad (23)$$

where $x = \beta\epsilon_F$.

The above approximate solution has nearly the same structure as the $D = 2$ solution, i.e.,

$$e^\alpha = -1 + e^x(1 - \pi^2/6x + \cdots). \quad (24)$$

For higher even-numbered dimensions, one can obtain the chemical potential in essentially the same way. It is clear that the low-temperature behavior of the chemical potential in even-numbered dimensions is quite different from that in odd-numbered dimensions. Unlike in odd-numbered dimensions, $T = 0$ behaves like an essential singular point.

Finally we note that one may also obtain a formal solution of (4) by a small-z expansion as follows:

$$(\beta\epsilon_F)^{D/2} = \frac{D}{2}\int_0^{\infty} dx\, x^{(1-D/2)}(1 + z^{-1}e^x)^{-1} = \Gamma\left(\frac{2+D}{2}\right)\sum_{n=1}^{\infty} \frac{(-)^n z^{n+1}}{(n+1)^{D/2}}. \quad (25)$$

One can easily verify that for, e.g., $D = 4$, the above solution corresponds to (22).

V. CONCLUDING REMARKS

Using the ideas due to Barker and Blankenbechler, we have obtained the chemical potential when dimensions are odd numbered. It is given in the form of a low-temperature expansion. To the leading order in T, we find that $\mu(T) > \mu(0)$ if $D = 1$, but $\mu(T) < \mu(0)$ if $D \geqslant 3$. The difference arises from the $(D-2)$ factor in the coefficient of expansion. Since $\mu(T \to \infty) < 0$, the chemical potential in one dimension must begin to develop downward curvature at some temperature, say T_1, i.e., $\partial\mu(T_1)/\partial T = 0$. This probably is a unique feature of the one-dimensional model.

For even-numbered dimensions, we have shown that the behavior of the chemical potential is rather different. It cannot be given a low-temperature expansion, as for odd-numbered dimensions. For $D = 2$, we have obtained an expression for the chemical potential in closed form. It is surprisingly elementary, but it describes both the high- and low-temperature regions in a useful way. For $D \geqslant 4$ it does not appear possible to obtain a closed form expression. But it seems that the solution is built on the structure of the two-dimensional form.

The different behavior observed in the chemical potential of even- and odd-numbered dimensions was also present in other physical properties of a free Fermi gas, e.g., the susceptibility.

ACKNOWLEDGMENT

This work was supported in part by the NSF and the ARO.

[1] W. A. Barker, J. Math. Phys. **27**, 302 (1986).

[2] A. Sommerfeld, Z. Phys. **47**, 1 (1928).

[3] R. K. Pathria, *Statistical Mechanics* (Pergamon, Oxford, 1972).

[4] W. A. Barker, J. Math. Phys. **28**, 1385, 1389 (1987).

[5] R. Blankenbechler, Am. J. Phys. **25**, 279 (1957).

[6] C. Kittel, *Elementary Statistical Mechanics* (Wiley, New York, 1958), p. 93.

[7] N. L. Sharma and M. H. Lee, J. Math. Phys. **27**, 1618 (1986).

[8] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Appl. Math. Ser. No. 55 (U.S. GPO, Washington, DC, 1970), p. 258.

[9] For $D = 3$, the coefficient of x^{-4} was incorrectly given by Barker [Eq. (29) of Ref. 1]. Ours agrees with, e.g., J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1940), p. 384.

On the approach to statistical equilibrium in an infinite-particle lattice dynamic model

M. Courbage
Laboratoire de Probabilités, Université Paris 6, Tour 56, 4 Place Jussieu, 75230 Paris Cedex 05, France and UFR de Physique, Université de Paris 7, Paris, France

(Received 28 June 1988; accepted for publication 8 February 1989)

A deterministic reversible model dynamic system of infinite interacting particles on a lattice is studied. The model is a modification by de Haan of the Kac ring model. The time evolution of the statistical states can be described by a hierarchy for the reduced distribution functions. It is shown that the system is isomorphic to a Bernoulli shift. The isomorphism is given in terms of the infinite process of observation of the states of one particle in a fixed site. The approach to equilibrium is studied and time-asymmetric initial states that approach equilibrium are constructed. It is shown that these states contain nonvanishing correlations between infinitely distant particles. The Bernoulli property allows the construction of a Markovian irreversible description of the model, realized by a nonunitary transformation. The action of this transformation on the statistical states, the correlations, and the rate of the approach to equilibrium is studied.

I. INTRODUCTION

Kac has devoted an important part of his book[1] to study the paradoxes of irreversibility. This subject is well-known to be very difficult, so he strives for models which illustrate different aspects of the irreversibility. Kac explains that the monotonic statistical approach of a dynamical system to equilibrium with time-increasing entropy is described by the so-called "master equation," analogous to the Ehrenfest urn stochastic model. The main point is to justify the probabilistic description on the basis of a deterministic dynamical description, and to examine the kind of probabilistic "ingredients" which may yield the deterministic description identical, in the limit of an infinite number of particles, to the master equation. Kac has illustrated these ideas in a simple model which can be briefly presented as follows.

On a circle we consider n equidistant sites, m of which are marked. On each site we put a ball which can be either white or black. At regular time intervals, each ball is shifted to the nearest site at the left, changing its color if and only if the site it leaves is marked. The dynamic system is thus described by η_i and ϵ_i, $i = 1,...,n$, where $\eta_i = +1$ or -1 according to the color of the ball in the site i, and $\epsilon_i = -1$ if the site i is marked and $+1$ if not. One may think of the marked sites as occupied by heavy particles; the model is thus reminiscent of the Ehrenfest "wind in trees" model. A microscopic state is represented by a sequence $\underline{\eta} = (\eta_1,...,\eta_n)$ and $\underline{\eta}(t+1)$ is given in terms of $\underline{\eta}(t)$ by

$$\eta_i(t+1) = \epsilon_{i+1}\eta_{i+1}(t) . \tag{1.1}$$

We denote by S the transformation $\underline{\eta}(t) \to \underline{\eta}(t+1) = S\underline{\eta}(t)$. This dynamics is periodic, for the system is finite. Now let us consider the statistical states of the system which correspond to the probability distributions ρ on the phase space Ω of the $\underline{\eta}$'s configurations, where $\rho(\underline{\eta})$ denotes the probability of the configuration $\underline{\eta}$. An initial state $\rho_0(\underline{\eta})$ evolves under S^t by

$$\rho_t(\underline{\eta}) = \rho_0(S^{-t}\underline{\eta}) = (U_t\rho_0)(\underline{\eta}) . \tag{1.2}$$

Note that U_t defined by this relation is the time evolution operator of the distribution ρ_0. We take as a "Liouville measure" the invariant measure P for which all η_i's are independent, equidistributed with $P(\eta_i = +1 \text{ or } -1) = \frac{1}{2}$. It is clear that *this evolution is also periodic and deterministic and cannot lead to any monotonic approach to equilibrium.* Nevertheless one expects intuitively, as a result of the "collisions" with the ϵ particles, that whatever the initial mean density of each color, the initial ensemble will converge as $t \to \infty$ in some sense to an ensemble in which there is a mean equal number of $+1$ and -1. Evidently $\rho_t(\underline{\eta})$ will depend on (ϵ_i) for $t > 0$. Hence in order to get a stochastic and irreversible evolution, Kac, thinking of the "wind in trees" model, supposes that the particles' ϵ_i's are randomly equidistributed variables with $P(\epsilon_i = -1) = p$. The problem is to show that the so-called "coarse-grained" distribution, which is the expectation $\langle \rho_t \rangle(\underline{\eta})$ of ρ_t with respect to ϵ_i, which we shall denote simply $E^\epsilon \rho_t$, evolves monotonically to the equilibrium under a closed Markovian semigroup such that $E^\epsilon \rho_t$ depends only on the initial states and not on the whole past. This means that the $E^\epsilon \rho_t$ evolves under a Markov chain. However, the simple coarse-graining at time t does not lead generally to such evolution. Then Kac replaces it by a coarse-graining repeated after each step, and calls this evolution the master equation,

$$\tilde{\rho}_t = (E^\epsilon U_1)^t \rho_0 = W^t \rho_0 , \tag{1.3}$$

where W is a stochastic matrix transition, that is,

$$\tilde{\rho}_{t+1}(\underline{\eta}) = \sum_{\underline{\delta}\in\Omega} \tilde{\rho}_t(\underline{\delta})\, \Pi_{\underline{\delta},\underline{\eta}} = W\tilde{\rho}_t , \tag{1.4}$$

with

$$\Pi_{\underline{\delta},\underline{\eta}} = \prod_{i=1}^{n} \frac{1}{2}(1 + (1-2p)\delta_{i+1}\eta_i) . \tag{1.5}$$

Then he shows that $\tilde{\rho}_t$ tends monotonically to equilibrium, that is, the expectations of any product of $\eta_{i_1}\, \eta_{i_2} \cdots \eta_{i_k}$ with respect to the probability $\tilde{\rho}_t$, $E_{\tilde{\rho}_t}(\eta_{i_1}\eta_{i_2}\cdots\eta_{i_k})$, tend mono-

 0022-2488/89/081840-11$02.50

tonically to the vanishing equilibrium value as $ke^{-\alpha t}$ and ρ_t verifies an $\mathscr{H}$ theorem (i.e., a law of entropy increase).

To justify this new stochastic description, Kac tried to show that in the limit $n \to \infty$, the solution of the irreversible master equation may be (weakly) approximated by the solution of the Liouville equation (1.2) only for a restricted class of symmetrized coarse-grained initial distributions. This means that the irreversible evolution that comes from a probabilistic hypothesis can be justified if it may coincide with the deterministic evolution under some limits. It is important however to point out that this irreversible evolution of the η particles comes from the random distribution of the surrounding ϵ particles and cannot explain the problem of the irreversible approach to equilibrium of the *isolated systems* (several authors have studied and further developed the Kac model; see for instance Ref. 2).

Recently, de Haan has given another version,[3] which is a system of interacting particles on a lattice evolving under deterministic reversible dynamics. He shows that states with finite range correlations will go asymptotically in the future to the equilibrium state for a system of an infinite number of particles. However, the approach to the equilibrium is not uniform for all local variables and is very slow. The model illustrates some of the problems which are present in the reduced distribution functions approach to nonequilibrium statistical mechanics, such as the problem of creation of correlations between the particles, which was eluded in the Kac model. He also tried to study the problem of the intrinsic irreversibility of the system. This signifies the existence of a class of statistical states that go to equilibrium in the future as $t \to \infty$ but not in the past as $t \to -\infty$. This property has been put forth and studied by Prigogine and the author in the case of unstable dynamical systems.[4] It means that for these systems the set of all statistical states that go to equilibrium in the future is not invariant under the time inversion. The difficulties encountered by de Haan in this respect can be solved if one may prove the instability of the system.

Here we again take the study of the time evolution under the deterministic dynamics of this model with a different approach, in which we start with an infinite number of particles' description and we seek an equivalence of the dynamics with a shift by the well-known technique of symbolic dynamics. If the shift has good ergodic properties, then one may easily characterize a class of states approaching the equilibrium. We have displayed such symbolic dynamics, allowing us to show that the model has very strong ergodic properties (Bernoulli system). This entails several properties: (1) It gives some general characterization of the states which approach the equilibrium. A simple characterization of these states in physical terms is not easy. (2) It allows us to construct a class of statistical states that approach the equilibrium. (3) It allows us to display a class of time asymmetric initial conditions. We show that these states have infinitely long range correlations. This result has to be compared with a general rule in the kinetic theory of dense gases, where one usually postulates that initial states should satisfy a principle of spatial damping of correlations between infinitely distant particles.[5] It appears that this selection rule eliminates states which are time asymmetric.

The isomorphism with the shift brings out the nonuniform nature of the approach to equilibrium, and therefore the nonexistence of a relaxation time uniformly for all local observables. However, the existence of a time relaxation is an important feature of an irreversible description, in addition to the monotonic approach to equilibrium. In this model the transition to such a description is possible in view of the instability of the system. In fact, a stochastic and irreversible description has been previously introduced by Misra, Prigogine, and the author for abstract unstable dynamical systems.[6–8] Let us qualitatively summarize the main ideas. First of all, it is recognized that the phase space of the unstable systems is foliated into two families of exponentially converging trajectories and exponentially diverging ones. This forces us to consider as indiscernable the first family of phase points called the stable manifolds, and to substitute to the density ρ_t a new one, which is in some sense coarse-grained with respect to the stable manifolds and denoted $\Lambda\rho_t$. The main property of this new density is that *it obeys to a closed Markovian evolution such that*

$$\Lambda U_t \rho = W_t^* \Lambda\rho, \quad \text{for } t>0, \tag{1.6}$$

where W_t is a semigroup of Markov processes with an $\mathscr{H}$ theorem, that is, the functional

$$\int_\Gamma (\Lambda\rho_t)(x)\log(\Lambda\rho_t)(x)d\mu(x) \tag{1.7}$$

decreases monotonically to its equilibrium value as $t \to \infty$.

We shall display in this model the meaning of the indiscernability of the configurations, and study some of the main consequences of the introduction of this transformation. First, the Λ transformation drastically changes the states. As an example, we display a class of states with infinitely long range correlations that are transformed into states which are quasilocal perturbations of equilibrium. Second, it modifies the rate of convergence to equilibrium. In other words, there exist states slowly and locally converging to equilibrium under the deterministic dynamics, but they converge globally and exponentially under the stochastic description. Third, the problem of a closed evolution of the one-particle distribution function does not seem to be solved by this transformation, for as in the deterministic dynamics, there is no chaos propagation in the stochastic evolution. However, the damping of the correlations, as realized by the transformation, suggests that this stochastic evolution may be more appropriate to obtain kinetic equations under some limits than the deterministic descriptions. This point deserves some new investigation.

II. THE KAC–DE HAAN MODEL AND ITS STATISTICAL STATES

A. The infinite version of the Kac–de Haan model

On the lattice of the non-negative integers N there are two species of particles. At each site $i \in N$ there is one particle from each kind, the variables η_i and ϵ_i representing their respective states, with each one taking the value $+1$ or -1. A microscopic state is represented by $x = (\eta_i, \epsilon_i)$, $i = 0,1,\ldots$. The phase space Γ is the set of all such sequences. The transformation $S: x(t) \to x(t+1) = Sx(t)$ is given by

$$\eta_i(t+1) = \epsilon_i(t)\eta_{i+1}(t), \quad \epsilon_i(t+1) = \eta_i(t). \tag{2.1}$$

Thus the state at time $t+1$ in the site i depends only of the state of the right neighboring particles (as in cellular automata). S is one-to-one and S^{-1} is given by

$$\eta_i(-1) = \epsilon_i(0), \quad \epsilon_i(-1) = \eta_i(0)\epsilon_{i+1}(0). \tag{2.2}$$

We take as an equilibrium measure μ, the distribution in which all η_i and ϵ_i are independently equidistributed with probability $\frac{1}{2}$. The measure is uniquely defined by its values on the "cylindric sets"

$$A^{(u_1,v_1),\ldots,(u_n,v_n)}_{i_1,\ldots,i_n}, \quad i_1 < i_2 < \cdots < i_n \in N$$

where

$$A^{(u_1,v_1),\ldots,(u_n,v_n)}_{i_1,\ldots,i_n} = \{x \in \Gamma,\ \eta_{i_1} = u_1,\ \epsilon_{i_1} = v_1,\ldots,\epsilon_{i_n} = v_n\}. \tag{2.3}$$

For the sake of simplicity we use the following notations:

$$(i_1,\ldots,i_n) = i, \quad (u_1,\ldots,u_n) = u, \quad \{(u_1,v_1),\ldots,(u_n,v_n)\} = (u,v). \tag{2.4}$$

Thus we have

$$\mu(A_i^{(u,v)}) = (\tfrac{1}{4})^n. \tag{2.5}$$

We denote by $\mathscr{A}$ the σ algebra generated by these sets.

We show next that μ is invariant, i.e., $\mu(S^{-1}A) = \mu(A)$ for any $A \in \mathscr{A}$, and that among all the invariant measures of S, μ is the unique one which maximizes the Kolmogorov–Sinai entropy (the system is the most random in this stationary distribution).

B. Nonequilibrium statistical states

A (nonequilibrium) statistical description of the system is given by any probability distribution ν on $(\Gamma,\mathscr{A})$ defined as above by its finite-dimensional joint distributions. We denote by $\mathscr{S}(\Gamma,\mathscr{A})$ the set of all probability measures on $(\Gamma,\mathscr{A})$.

A remarkable representation of ν can be given in terms of the moments of all finite products of ϵ_i and η_j: let $\rho_n(x)$ be the function

$$\rho_n(x) = \frac{\nu(\eta(x),\epsilon(x))}{\mu(\eta(x),\epsilon(x))}, \tag{2.6}$$

where $\nu(\eta(x),\epsilon(x))$ is the ν measure of the $A^{(\eta,\epsilon)}_{0,1,\ldots,n}$ containing x and $\eta_i = \eta_i(x)$, $\epsilon_i = \epsilon_i(x)$. Thus ρ_n is a function of a finite number of ($+1$ or -1) valued variables. Any function of such variables, say $(u_1,\ldots,u_n) = \underline{u}$, $u_i = +1$ or -1, has the following representation:

$$f(u_1,\ldots,u_n) = \frac{1}{2^n}\Big(c_0 + \sum_{i=1}^{n} c_i u_i + \sum_{i<j} c_{i,j} u_i u_j + \cdots + c_{1,2,\ldots,n} u_1 u_2 \cdots u_n\Big). \tag{2.7}$$

In fact, the space of such functions endowed with the scalar product,

$$\langle f,g\rangle = \sum_{\underline{u}} \Big(\frac{1}{2^n}\Big) f(\underline{u}) g(\underline{u}), \tag{2.8}$$

has an orthonormal basis given by 1 and $u_{i_1} \cdot u_{i_2} \cdots u_{i_r}$, $1 \leqslant i_1 < i_2 < i_r \leqslant n$. It follows that

$$C_{\underline{i}} = \langle f, u_{\underline{i}}\rangle. \tag{2.9}$$

Thus ρ_n has the expansion

$$\rho_n(\eta_0,\epsilon_0,\ldots,\eta_n,\epsilon_n) = 1 + \sum_{\substack{r,l \\ i_k = 0,\ldots,n}} C^{j_1,\ldots,j_l}_{i_1,\ldots,i_r} \eta_{i_1} \cdots \eta_{i_r} \epsilon_{j_1} \cdots \epsilon_{j_l}. \tag{2.10}$$

In this sum, r and l range from 0 to n. The coefficients $C^{\underline{j}}_{\underline{i}}$, hereafter called the Fourier coefficients of ν, are given by

$$C^{\underline{j}}_{\underline{i}} = \langle \rho_n, \eta_{\underline{i}} \epsilon_{\underline{j}}\rangle = \sum_{\underline{\eta},\underline{\epsilon}} \Big(\frac{1}{4}\Big)^{n+1} \eta_{\underline{i}} \epsilon_{\underline{j}} \rho_n(\eta_0,\epsilon_0,\ldots,\eta_n,\epsilon_n) \tag{2.11}$$

$$= E_\nu(\eta_{\underline{i}} \epsilon_{\underline{j}}). \tag{2.12}$$

As a consequence, the coefficients $C^{\underline{j}}_{\underline{i}}$ do not depend on n. The correlations can be computed from (2.10):

$$\nu(\eta_i\eta_j) = 1/2^2(1 + C_i\eta_i + C_j\eta_j + C_{ij}\eta_i\eta_j), \quad \nu(\eta_i) = \tfrac{1}{2}(1 + C_i\eta_i). \tag{2.13}$$

Thus

$$g_2(\eta_i,\eta_j) = \nu(\eta_i\eta_j) - \nu(\eta_i)\nu(\eta_j) = 1/2^2(C_{ij} - C_iC_j)\eta_i\eta_j. \tag{2.14}$$

The term $g_2(\eta_i,\eta_j)$ is called the two-particle correlation function.

The high-order correlations of the cluster representation (see Ref. 9) can be computed similarly. A distribution in which all particles are uncorrelated is characterized by the condition

$$C^{j_1,\ldots,j_l}_{i_1,\ldots,i_r} = C_{i_1} \cdots C_{i_r} C^{j_1} \cdots C^{j_l} \tag{2.15}$$

and called a *chaos state*.

C. Time evolution of the statistical states

Analogous to the Liouville equation is the group of transformations α_t acting on a measure ν according to

$$(\alpha_t\nu)(A) = \nu_t(A) = \nu(S^{-t}A), \quad A \in \mathscr{A}. \tag{2.16}$$

We say that a measure ν weakly approaches the equilibrium for $t \to \pm\infty$ if

$$\nu_t(A) \to \mu(A), \quad t \to \pm\infty \tag{2.17}$$

for every cylindric set A. It can be easily shown that this is equivalent to the convergence of the moments $C^{\underline{j}}_{\underline{i}}(t)$ [and also equivalent to the convergence of the expectation of any continuous function $f(x)$ on Γ:

$$\int_\Gamma f\,d\nu_t \to \int_\Gamma f\,d\mu, \quad t \to \pm\infty, \tag{2.18}$$

with the natural product topology on Γ]. This is a weak and local convergence to equilibrium.

The set of all states that tend to μ will be called the basin of attraction of μ and denoted $\mathscr{S}(\mu,\alpha_t)$.

It is interesting to pursue here the analogy with the nonequilibrium statistical mechanics to show a nonpropagation of the "chaos," that is, creation of correlations for states verifying (2.15). The time evolution of one-particle distribution function is computed from the definition of S:

$$\nu_{t+1}(\eta_i) = \nu_t(S^{-1}A_i^{\eta_i})$$
$$= \nu_t(\{x = (u_i,v_i)\colon u_{i+1}v_i = \eta_i\})$$
$$= \sum_{u_{i+1}} \nu_t(v_i = u_{i+1}\eta_i, u_{i+1}) . \tag{2.19}$$

Thus the one-particle distribution function at time t depends on the two-particle function at time $t-1$ and this will propagate to a many-particle distribution function at time $t = 0$. (See Ref. 9.) More generally, we have the analog of the BBGKY hierarchy,

$$\nu_{t+1}(\eta_0,\epsilon_0,\ldots,\eta_n,\epsilon_n)$$
$$= \sum_u \nu_t((\epsilon_0,\epsilon_1,\eta_0),(\epsilon_1,\epsilon_2\eta_1),\ldots,(\epsilon_n,u\eta_\eta),u) . \tag{2.20}$$

To verify the nonpropagation of chaos we compute the evolution of the Fourier coefficients by using the above formalism (also computed by de Haan directly). From (2.12) it comes,

$$c_{\underline{i}}^{\underline{j}}(t+1) = \int_\Gamma d\nu_{t+1}(x)\eta_{\underline{i}}(x)\epsilon_{\underline{j}}(x)$$
$$= \int_\Gamma d\nu_t(S^{-1}x)\eta_{\underline{i}}(x)\epsilon_{\underline{j}}(x) . \tag{2.21}$$

Now, from the evolution law of η_i's and ϵ_i's (2.1) we get a similar law for any product of them. Inserting this formula in (2.21) yields

$$C_{\underline{i}}^{\underline{j}}(t+1) = E_{\nu_t}(\eta_{\underline{i}+1}(x)\eta_{\underline{j}}(x)\epsilon_{\underline{i}}(x))$$
$$= C^{\underline{i}}_{(\underline{i}+1)\Delta\underline{j}} , \tag{2.22}$$

where we have taken into account that $\eta^2 = 1$. The symmetrical difference between two sets of indices is denoted by Δ. Let us suppose that ν_0 is a chaos state (2.15). From (2.22) we get

$$C_i^j(1) = \begin{cases} C^i, & \text{if } j = i+1, \\ C^iC_jC_{i+1}, & j \neq i+1, \end{cases} \tag{2.23}$$

$$C_i(1)C^j(1) = C^iC_jC_{i+1} .$$

Thus after one transformation we have nonvanishing correlations *only* between η_i and ϵ_{i+1}:

$$g_2(\eta_i = 1, \epsilon_{i+1} = 1) = (1/2^2)(1 - C_{i+1}^2)C^i . \tag{2.24}$$

This illustrates the creation of the correlations under the dynamical evolution even when the initial state is completely uncorrelated and the propagation of the correlations to more and more particles.

III. MIXING PROPERTY

To check the invariance of μ, we have to show that

$$\mu(S^{-1}A_{k,\ldots,k+n}^{(\eta_0,\epsilon_0),\ldots,(\eta_n,\epsilon_n)}) = (\tfrac{1}{4})^{n+1} \tag{3.1}$$

for any $k \geqslant 0$ and $n \geqslant 0$. The same argument as in (2.20) yields the invariance

$$\mu(S^{-1}A_{k,\ldots,k+n}^{(\eta_0,\epsilon_0),\ldots,(\eta_n,\epsilon_n)})$$
$$= \sum_{u=\pm 1} \mu(A_{k,\ldots,k+n,k+n+1}^{(\epsilon_0,\epsilon_1,\eta_0),\ldots,(\epsilon_n,u\eta_n),(u,\cdot)})$$
$$= (\tfrac{1}{4})^{n+1} . \tag{3.2}$$

Now we shall realize an isomorphism between S and a Bernoulli shift by constructing a generating and independent partition $\mathscr{P} = (P_0,P_1)$, that is a partition which verifies, respectively, the properties

(i) $\mu(S^{j_1}P_{i_1} \cap \cdots \cap S^{j_k}P_{ik}) = \prod_{\alpha=1}^{k} \mu(S^{j_\alpha}P_{i_\alpha})$,

(ii) $\vee_{-\infty}^{+\infty} S^i\mathscr{P}$ is the partition of Γ into points, where the product $\mathscr{P} \vee Q$ of two partitions $\mathscr{P}$ and $\mathscr{Q}$ is the partition with elements $P_i \cap Q_j$ (for a short presentation see Ref. 7 and for more details see Ref. 10).

Let us consider the partition

$$\mathscr{P} = \{A_0^{(-1,\cdot)}, A_0^{(1,\cdot)}\} , \tag{3.3}$$

where $A_0^{(\eta,\cdot)} = \{x\colon \eta_0(x) = \eta\}$. We denote $P_u = A_0^u$ where A_0^u stems for $A_0^{(u,\cdot)}$. *In what follows we identify* -1 *with* 0 *and thus* P_{-1} *stems for* P_0. The symbolic dynamics is associated with this partition through the representation of any $x \in \Gamma$ by an infinite sequence $\phi(x) = \{u_n(x)\}$ where $u_n(x)$ is the index of the element of $\mathscr{P}$ containing $S^n(x)$. Here ϕ is a mapping from Γ into $\Omega = \{0,1\}^Z$. We denote by $\underline{u}$ any such double sequence. Sx is mapped into the shifted sequence $\sigma\underline{u} = \{u_{n+1}\}$, that is,

$$\phi Sx = \sigma\phi x . \tag{3.4}$$

Note that $\underline{u} = \{\eta_0(n)\}$ is in fact the "history" of the states of the particle η in the site 0 for a given configuration x. In general, not any $\underline{u}$ of Ω represents such a history for some $x \in \Gamma$. *The main point here is that there is a bijection between* Γ *and* Ω *as shown in the following proposition.*

Proposition 1: Let A_n^α be a set of the form $A_{0,\ldots,\eta}^{(\eta_0,\epsilon_0),\ldots,(\eta_n,\epsilon_n)}$ and Δ_n^β a set of the form $S^{-n}A_0^{u_n} \cap \cdots \cap S^{n+1}A_0^{u_{-n-1}}$. Then any A_n^α set is identical with a unique Δ_n^β set. Therefore the mapping ϕ is one-to-one and $\mathscr{P}$ is generating. Moreover $\mathscr{P}$ is independent.

Proof: A cell of the form

$$\Delta_n^\beta = S^{-n}A_0^{u_n} \cap \cdots \cap S^{n+1}A_0^{u_{-n-1}} \tag{3.5}$$

is the set of $x \in \Gamma$ such that $(S^ix)_0 = u_i$ for $i = -n-1,\ldots,n$. We shall show that for each set A_n^α there exists some Δ_n^β such that

$$A_n^\alpha \subset \Delta_n^\beta , \tag{3.6}$$

and conversely for each Δ_n^γ there exists some A_n^δ such that $\Delta_n^\gamma \subset A_n^\delta$. For this, let us express for each x the double sequence $u_i(x)$ in terms of $\eta_i(x)$ and $\epsilon_j(x)$. This follows from the laws of motion (2.1) and (2.2). The table at the end of this section gives the first elements. It is easy to show by induction that u_i has the following general form:

$$u_k(x) = \eta_{i_1}\cdots\eta_{i_k}\epsilon_{j_1}\cdots\epsilon_{j_k}\eta_k ,$$
$$i_1 < \cdots < i_k < k ,$$
$$j_1 < \cdots < j_k < k , \tag{3.7}$$

$$u_{-k}(x)=\eta_{\alpha_1}\cdots\eta_{\alpha_k}\epsilon_{\beta_1}\cdots\epsilon_{\beta_k}\epsilon_{k-1},$$
$$\alpha_1<\cdots<\alpha_k<k,\qquad \beta_1<\cdots<\beta_k<k-1 \tag{3.8}$$

[the second formula can be deduced from the first by using the time inversion (4.18) and (C6) of Appendix C].

Now for any $x\in A_n^\alpha$, all $\eta_0,...,\eta_n$, $\epsilon_0,...,\epsilon_n$ are fixed. Then the above formulas show that $u_{-n-1},...,u_n$ are also fixed. This entails that A_n^α is included in some Δ_n^β set. Conversely, for any $x\in\Delta_n^\gamma$, $u_i=\eta_0(S^ix)$ are fixed for all $i=-n-1,...,n$. It is clear from (3.7) and (3.8) by induction that all $\eta_0,...,\eta_n$, $\epsilon_0,...,\epsilon_n$ are fixed, thus $x\in A_n^\delta$. This proves that $\Delta_n^\gamma\subset A_n^\delta$. It follows that the sets $\{A_n\}$ and $\{\Delta_n\}$ are identical: $\forall\alpha$, $\exists\beta$,

$$A_n^\alpha=\Delta_n^\beta\,. \tag{3.9}$$

This shows that ϕ is one-to-one from Γ onto Ω.

Thus $\mathscr{P}$ is generating and ϕ is an isomorphism between $(\Gamma,\mathscr{A},S,\mu)$ and (Ω,σ,μ). The relation (3.9) implies also that

$$\mu(S^{-n}A_0^{u_n}\cap\cdots\cap S^{n+1}A_0^{u_{-n-1}})=\mu(A_n^\alpha)=(1/2)^{2(n+1)}\,. \tag{3.10}$$

This shows that $\mathscr{P}$ is independent and that σ is a Bernoulli shift.

This proposition entails that μ is the unique invariant measure which maximizes the Kolmogorov–Sinaï entropy of the (topological) dynamic system S. In fact, as ϕ is one-to-one, it maps the set of all S-invariant measures onto the set of all σ-invariant measures (where Ω has the natural σ-algebra $\mathscr{C}$ of the cylindric sets $C_{i_1,...,i_n}^{u_{i_1},...,u_{i_n}}=\{\underline{u}=u_{i_1},...,u_{i_n}\text{ fixed}\}$) as follows:

$$\mu\leftrightarrow\hat\mu:\ \ \hat\mu(c)=\mu(\phi^{-1}c),\quad c\in\mathscr{C}\,. \tag{3.11}$$

Now if m_0 is the unique measure which maximizes the KS entropy $h_m(\sigma)$ then $m_0\,\phi$ maximizes also uniquely $h_{\hat m}(S)$, for the KS entropy is invariant under ϕ, that is, $h_m(S)=h_{\hat m}(\sigma)$. Now it is well-known (see Ref. 10, p. 194) that the Bernoulli measure $(\frac12,\frac12)$ is the unique measure which maximizes the KS entropy for the shift σ and this maximum is equal to log 2.

Remark: This result on the uniqueness of the maximal entropy measure [a special case of the Parry theorem (Ref. 10, p. 194)] also entails the independence of the partition $\mathscr{P}$ for the measure μ. In fact, one computes directly that $h_\mu(S)$ is equal to log 2:

$$\begin{aligned}h_\mu(S)&=H_\mu(S,\mathscr{P})\\&=\lim_{n\to\infty}\frac{1}{(2n+2)}H_\mu\left(\vee_{-n-1}^{n}S^i\mathscr{P}\right)\\&=\lim_{n\to\infty}\frac{1}{(2n+2)}\sum_{u,v}-\mu(A_{0,...,u}^{u,v})\log\mu(A_{0,...,n}^{u,v})\\&=\log 2\,.\end{aligned} \tag{3.12}$$

Then $h_{\hat\mu}(\sigma)$ is also equal to log 2 which implies that $\hat\mu$ is the Bernoulli measure.

The isomorphism ϕ means that one can equivalently represent the configuration $x=(\eta_i,\epsilon_i)$ by the whole "history" of the particle η at the site 0 when x evolves under S. Knowing this history, one is able to reconstruct the complete configuration. For instance, the first elements of this dictionary are computed from the definition:

$$\begin{matrix}u_{-5}&u_{-4}&u_{-3}&u_{-2}&u_{-1}&u_0&u_1&u_2&u_3&u_4\\ \eta_3\epsilon_0\epsilon_2\epsilon_4&\eta_0\eta_2\epsilon_3&\eta_1\epsilon_0\epsilon_2&\eta_0\epsilon_1&\epsilon_0&\eta_0&\epsilon_0\eta_1&\epsilon_1\eta_0\eta_2&\epsilon_0\epsilon_2\eta_3&\epsilon_3\eta_0\eta_2\eta_4\end{matrix}\,. \tag{3.13}$$

By solving these equations, one determines η_i and ϵ_i in terms of u_i. We have, for instance,

$$\begin{matrix}\eta_1&\eta_2&\eta_3&\eta_4&\epsilon_1&\epsilon_2&\epsilon_3&\epsilon_4\\ u_{-1}u_1&u_{-2}u_2&u_{-3}u_{-1}u_1u_3&u_{-4}u_4&u_0u_{-2}&u_{-3}u_1&u_{-4}u_{-2}u_0u_2&u_{-5}u_3\end{matrix}\,. \tag{3.14}$$

One shows by induction that more generally η_i and ϵ_j are given by formulas of the form

$$\begin{aligned}&\eta_k=u_{-k}u_\alpha\cdots u_\beta u_k,\quad -k<\alpha<\cdots<\beta<k\,,\\&\epsilon_k=u_{-k-1}u_\alpha\cdots u_\beta u_k,\\&\qquad -k-1<\alpha'<\cdots<\beta'<k-1\,.\end{aligned} \tag{3.15}$$

IV. APPROACH TO EQUILIBRIUM AND TIME-ASYMMETRIC DISTRIBUTIONS

de Haan has studied the problem of the existence of a class of states which will approach the equilibrium, and the mechanism of this approach. He shows by using diagrammatic arguments that the subsequence ν_{t_n} of ν_t, with $t_n=2^n$, converges to μ as $n\to+\infty$ if the initial state ν has finite range correlations. By this it is meant that $\exists\delta>0$ such that if $(\underline{i},\underline{j})$ can be decomposed into two sets (k,k') with $\text{dist}(k,k')>\delta$, then $c_{\underline{i}}^{\underline{j}}$ will factorize into $c(\underline{k})\cdot c(\underline{k}')$. This is equivalent to the independence of η_i,ϵ_j when they are separated by more than δ sites.

The above result is probably true for the sequence ν_t but the proof is not complete. Yet the mechanism of the convergence to equilibrium of these states is quite different from the mixing. The mixing and the Bernoulli property of the system allow us to characterize another class of initial states in the basin of attraction of μ. First it follows from the definition of the mixing that any initial normalized measure that is absolutely continuous with respect to μ will converge to μ both for $t\to\infty$ and for $t\to-\infty$. However, this is a restricted class of states that describes only *quasilocal perturbations with respect to equilibrium.* In fact, we shall prove in Appendix A that the expectation of any local observable is vanishing at infinity. More precisely, we prove the following proposition.

Proposition 2: If ν is given by a density probability ρ with respect to μ [i.e., $\nu(A)=\int_A\rho(x)d\mu(x)$ for any $A\in\mathscr{A}$] *then*

$C_{\underline{i}}^{j} \to 0$ *as* $|i| + |j| \to \infty$, *where the symbol* $|i|$ *denotes* $\max(i_1,...,i_r)$.

A simple characterization of all the states of the basin of μ is not easy. (See in this respect some characterization given in Ref. 11.) Moreover, Sigmund[12] has shown that the basin of μ is small and has a bad topological structure in the sense that it is contained in a set of first category.

However, if the system has the K property, then it is possible to construct a class of states that is singular with respect to μ and that converges to equilibrium as $t \to \infty$ but not for $t \to -\infty$. This class of states has been introduced and studied in Ref. 11.

The most important property of the above class of states is that it may distinguish the basin of μ under S^t for $t > 0$ from that corresponding to $t < 0$. When these two basins are not identical we call the system "intrinsically irreversible."[7] The time asymmetry of the system is displayed by the statistical states which tend to equilibrium as $t \to \infty$ but not for $t \to -\infty$.

We shall illustrate this class of states in the model. But we first give a more general criterion based on the shift property of symbolic dynamics. As ϕ is one-to-one, it maps the statistical states $\mathscr{S}(\Gamma,\mathscr{A})$ of S onto $\mathscr{S}(\Omega,\mathscr{C})$ as follows:

$$\hat{\nu}(c) = \nu(\phi^{-1}C), \quad C \in \mathscr{C} . \tag{4.1}$$

We first give a necessary and sufficient condition for the convergence of a measure $\hat{\nu}$ to equilibrium. Let us expand the measure of any cylindric set in its Fourier series as explained in Sec. 11,

$$\hat{\nu}(C^{u_{-n},...,u_n}_{-n,...,n}) = \frac{1}{2^{2n+1}} \left(1 + \sum_{-n}^{n} b_i u_i + \sum_{i<j} b_{ij} u_i u_j + \cdots \right), \tag{4.2}$$

where b_i, b_{ij}, etc., are the expectations of $u_i, u_i u_j$, etc. Then the time evolution of $\hat{\nu}$ is given for any $C \in \mathscr{C}$ by

$$\hat{\nu}_t(C) = \hat{\nu}(\sigma^{-t}C) . \tag{4.3}$$

The action of the shift on a cylindric set is also a shift,

$$\sigma^t C^{v}_{i} = C^{v}_{i-t} , \tag{4.4}$$

and using the definition of b_i as the expectation of u_i, it becomes,

$$\hat{\nu}_t(c^{u_{-n},...,u_n}_{-n,...,n}) = \frac{1}{2^{2n+1}} \left(1 + \sum_{\underline{i}} b_{\underline{i}+t} u_{\underline{i}} \right). \tag{4.5}$$

Here $\underline{i} + t$ denotes $(i_1 + t,...,i_r + t)$. Thus a necessary condition for the convergence of $\hat{\nu}_t$ to $\hat{\mu}$ as $t \to \pm\infty$ is that

$$\lim_{t \to \pm\infty} b_{\underline{i}+t} = 0, \quad \forall \underline{i}, \tag{4.6}$$

for $b_{\underline{i}+t}$ is the expectation of $u_{\underline{i}}$ in the state $\hat{\nu}_t$. Conversely, if (4.6) is verified, then (4.5) entails the convergence of $\hat{\nu}$ to $\hat{\mu}$. We summarize this in the following.

Proposition 3: A necessary and sufficient condition for the convergence of $\hat{\nu}_t$ *to* $\hat{\mu}$, $t \to \pm\infty$ *is that* $\lim_{t \to \pm\infty} b_{\underline{i}+t} = 0$, *for all* $\underline{i}$.

This result is valid for any Bernoulli system.

Remarks: (1) It is clear that if $\hat{\nu}$ is absolutely continuous, then its Fourier coefficients $b_{\underline{i}}$ tend to zero as $|\underline{i}| \to \infty$, and then $b_{\underline{i}+t} \to 0$ as $t \to \infty$. This gives another proof of the approach to equilibrium of this class of states.

(2) The convergence of $b_{\underline{i}+t}$ to 0 as $t \to \infty$ should be distinguished from the convergence of $b_{\underline{i}}$ to 0 as $|\underline{i}| \to \infty$. This can be seen from the two indices' coefficients $b_{\underline{i},j}$: $b_{i+t,j+t} \to 0$ as $t \to \infty$ for any i and j, is equivalent to the convergence to zero of $b_{i,j}$ when (i,j) goes to infinity *along the positive direction of any parallel to the first bisectrix.*

This proposition may also be used to generalize the construction of time-asymmetric distributions in the B systems. *The most general criterion is that* $b_{\underline{i}+t}$ *tend to zero for any* $\underline{i}$ *as* $t \to \infty$ *and not as* $t \to -\infty$. Let us recall our explicit construction for the Bernoulli system $(\Omega,\mathscr{C},\sigma_t,\mu)$ (Ref. 7b).

Let $\mathscr{C}_i$ be the sub-σ-algebra of $\mathscr{C}$ generated by the random variables $\{u_{-i},u_{-i+1},...\}$ (or, equivalently, by C^u_j, $j > -i$) and define $\mathscr{A}_i$ by

$$\mathscr{A}_i = \phi^{-1}\mathscr{C}_i . \tag{4.7}$$

Then $\sigma^t\mathscr{C}_i = \mathscr{C}_{i+t}$, $\mathscr{C}_i$ increases to $\mathscr{C}$ as $i \to \infty$ and decreases to the σ algebra generated by Ω as $i \to -\infty$. Let $\mathscr{C}_i^-$ be the sub-σ-algebra generated by $\{u_{-i-1},...\}$.

A class of time-asymmetric measures is given by measures that are absolutely continuous measures when restricted to some $\mathscr{C}_i$ and singular measures on $\mathscr{C}_i^-$. As an example, let $\hat{\rho}_i$ be the density of $\hat{\nu}_{1\mathscr{C}_i}$ with respect to $\hat{\mu}$:

$$\hat{\nu}_{1\mathscr{C}_i}(C) = \int_C \hat{\rho}_i(\omega) d\hat{\mu}(\omega) . \tag{4.8}$$

We define the measure $\hat{\nu}_d$ by

$$\hat{\nu}_d(C_1 \times C_2) = \delta_{\alpha^-}(C_1) \int_{C_2} \hat{\rho}(\omega) d\hat{\mu}(\omega) , \tag{4.9}$$

where $C_1 \in \mathscr{C}_i^-$, $C_2 \in \mathscr{C}_i$, and α^- is a fixed sequence $(u_{-i-1} = \alpha_{-1}, u_{-i-2} = \alpha_{-2},...)$. This is a measure concentrated on some dilating fiber of the Bernoulli shift. In other words, the measure is defined on the cylindric sets by

$$\nu(C^{j_{-m},...,j_n}_{-m,...,n}) = \delta_{\alpha_{-m},j_{-m}} \cdots \delta_{\alpha_{-i-1},j_{-i-1}} \nu^+(C^{j_{-i},...,n}_{-i,-i+1,...,n}) , \tag{4.10}$$

$$\nu^+(c) = \hat{\nu}_{1\mathscr{C}_i}(c) .$$

We shall show that these states have infinite range correlations (that is, nonvanishing correlations between infinitely distant particles η *or* ϵ*).*

We take in the definition of $\hat{\nu}_d$ $i = 0$. We first compute the relation between the Fourier coefficient of ν_d and $\hat{\nu}_d$, the c's, and the b's, respectively. In general this relation is not simple, except for $i = 2^n$ and 2^{-n}, namely

$$u_{2^n} = \eta_0(2^n) = \eta_{2^n}\eta_{2^n-2}\eta_{2^n-2^2} \cdots \eta_0\epsilon_{2^n-1} , \tag{4.11}$$

$$u_{-2^n} = \eta_{2^n-2}\eta_{2^n-2^2} \cdots \eta_0\epsilon_{2^n-1} . \tag{4.12}$$

The first has been computed by de Haan and the second will be shown in Appendix B. Both yield

$$\eta_{2^n} = u_{-2^n} u_{2^n} . \tag{4.13}$$

The Fourier coefficients of $\hat{\nu}_d$ are computed from the definition in (4.2),

$$\hat{\rho}_n = \frac{\hat{\nu}(c^{u_{-n},\ldots,u_n}_{-n,\ldots,n})}{\hat{\mu}(c^{u_{-n},\ldots,u_n}_{-n,\ldots,n})}$$

$$= 2^n \delta_{\alpha_{-n},u_{-n}} \cdots \delta_{\alpha_{-1},u_{-1}} \left(1 + \sum_{\underline{i}>0} b_{\underline{i}} u_{\underline{i}}\right)$$

$$= (1 + \alpha_{-n} u_n) \cdots (1 + \alpha_{-1} u_{-1}) \left(1 + \sum_{\underline{i}>0} b_{\underline{i}} u_{\underline{i}}\right),$$

and this yields

$$b_{\underline{i}} = \alpha_{\underline{i}}, \quad \underline{i} \leqslant 0 . \tag{4.14}$$

From (4.11) and (4.12) we get

$$c^{2^n-1}_{2^n-2,2^n-2^2,\ldots,0} = E_{\hat{\nu}}(u_{-2^n}) = b_{-2^n} = \alpha_{-2^n} . \tag{4.15}$$

Now we compute the correlation between η and ϵ in the sites $2^n, 2^n - 1$, respectively, and a group of particles distant from them by 2^{n-1} sites,

$$c^{2^n-1}_{2^n-2,2^n-2^2,\ldots,2^n,0} - c_{2^n-2,\ldots,2^{\hat{n}},0}\, c^{2^n-1}_{2^n} = \alpha_{-2^n} \cdots , \tag{4.16}$$

where $2^{\hat{n}}$ means that this index is omitted. To see that the second term of the right-hand side is vanishing as $n \to \infty$, we show that $c^{2^n-1}_{2^n} \to 0$ as $n \to \infty$. In fact, we have from (3.15), (4.13), and the definition of the measure that

$$c^{2^n-1}_{2^n} = \alpha_{\underline{j}} b_{i_1,\ldots,2^n}, \quad \underline{j} \leqslant 0, \; i_1 < \cdots < 2^n . \tag{4.17}$$

As $n \to \infty$, $b_{i_1,\ldots,2^n} \to 0$, the absolute value of the correlation will approach 1.

Remark: It comes from the invariance of the states with finite correlations under the time inversion that ν_d does not have finite correlations. Let us show this.

The time inversion I is a one-to-one transformation which has the following properties: (i) $I^2 = 1$, (ii) $IS^tI = S^{-t}$, (iii) $\mu(IA) = \mu(A)$ $\forall A \in \mathscr{A}$.

One can easily verify in the Kac–de Haan model that the time inversion corresponds to the permutation of the states of the particles η_i and ϵ_i in each site:

$$I(\eta_i, \epsilon_i) = (\epsilon_i, \eta_i) . \tag{4.18}$$

It is clear that the class of states with finite range correlations is invariant under the time inversion. Therefore the state $\hat{\nu}_d$, concentrated on the dilating fiber given by (4.9), does not belong to this class, for its time inverse, which is concentrated on a contracting fiber, cannot converge to μ, even for $t_n = 2^n$, $n \to +\infty$ (see Ref. 11).

V. TRANSITION TO IRREVERSIBLE EVOLUTION

We shall examine in this model the main consequences of the transition to the irreversible description under the Λ transformation which is an operator acting on measurable bounded functions and defined by

$$\Lambda\rho(x) = \sum_{i \in Z} \delta_i E^{\mathscr{A}_i} \rho(x), \quad \delta_i \geqslant 0, \; \sum_i \delta_i = 1 , \tag{5.1}$$

where $E^{\mathscr{A}_i}$ is the conditional expectation with respect to $\mathscr{A}_i$ (4.7) (in more physical terms $E^{\mathscr{A}_i}$ is the coarse graining with respect to some partition formed by the so-called contracting fibers.[7]

A. The contracting fibers

A contracting fiber is the family of the configurations x with the same $u_k(x)$ for all $k > -i$. If a distance is defined between two configurations by

$$d(x,x') = \sum_k |u_k(x) - u_k(x')| / 2^k ,$$

then one can easily see that the $d(x(t),x'(t))$ will decrease to zero as $t \to \infty$ if x and x' belong to the same contracting fiber. For such configurations, the $\eta_0(t)$ coincide for time $t > -i$ so that they do not distinguish the future state of the particle η in the origin.

Here we should note that the partition $\eta_0 = \pm 1$ which allows the construction of the contracting fibers may seem physically insufficient to justify a loss of discernability between the configurations of one fiber. In fact, only future states at a unique site are identical for all these configurations. However, one can also generate contracting fibers with the future observations of the system at any finite number of sites. In that case, it is natural to identify this family of configurations.

B. Change of the correlations

As shown in Ref. 7b, the singular measures constructed in Sec. IV may be transformed into absolutely continuous measures by Λ. Thus it comes from Proposition 2 of Sec. IV that Λ *may transform states with infinite correlations into states with quasilocal correlations.* That is, Λ realizes a renormalization of the correlations. This is reminiscent to some procedures used in the kinetic theory in order to eliminate the divergences. However, Λ does not entail the existence of a closed evolution of the one-particle reduced distribution function nor the propagation of the chaos under the irreversible description (we give in Appendix C explicit expressions of the evolution of the moments of some local observables in the new representation and in Appendix D the form of the new hierarchy).

C. The eigenfunctions of the time operator

The time operator T is a self-adjoint operator on $L^2_\mu(\Gamma)$ having a complete set of eigenfunctions $\psi_{n,\alpha}$ corresponding to the eigenvalues $n \in Z$ and propagated by the dynamical evolutions, i.e., $U_t \psi_{n,\alpha} = \psi_{n+t,\alpha}$, the index α corresponding to a countable degeneracy.

It has been shown in Ref. 7 that the eigenfunctions of $\hat{T}$ for the Bernoulli shift are characterized as follows.

Denote by χ_i the functions from Ω into $\{\pm 1\}$ defined by

$$\begin{aligned} &\chi_0(\underline{u}) = u_0 , \\ &\chi_i(\underline{u}) \equiv (U^i \chi_0)(\underline{u}) = \chi_0(\sigma^{-i}\underline{u}) = u_{-i} . \end{aligned} \tag{5.2}$$

Then the products are $\chi_{i_1} \cdots \chi_{i_p}$, $i_1 < \cdots < i_p$, form a complete orthonormal basis of eigenfunctions of $\hat{T}$ (resp. $\hat{\Lambda}$) corresponding to the eigenvalue i_p (resp. λ_{ip}). In other words, any product $u_{i_1} \cdots u_{i_p}$, $i_1 < \cdots < i_p$, is an eigenfunction of $\hat{T}$ (resp. $\hat{\Lambda}$) for the eigenvalue $-i_1$ (resp. λ_{-i_1}):

$$\hat{\Lambda} u_{i_1} \cdots u_{i_p} = \lambda_{-i_1} u_{i_1} \cdots u_{i_p} . \tag{5.3}$$

It follows from the results of Sec. III, Eqs. (3.7) and (3.8), that any such product corresponds to a product of ϵ_i and η_j and vice versa. Thus *the local observables coincide with the family of the eigenfunctions of the time operator*. This allows us to give an idea of the change of the moments of these observables under Λ. Let $k(\underline{i},\underline{j})$ be the eigenvalue of T for the eigenfunction $\eta_{\underline{i}}\epsilon_{\underline{j}}$. Then

$$\widetilde{C}^{\underline{j}}_{\underline{i}} \equiv \int_\Gamma d\nu\, \Lambda \eta_{\underline{i}}\epsilon_{\underline{j}} = \lambda_k(\underline{i},\underline{j}) \int d\nu\, \eta_{\underline{i}}\epsilon_{\underline{j}}\,, \tag{5.4}$$

$$\widetilde{C}^{\underline{j}}_{\underline{i}} = \lambda_k(\underline{i},\underline{j}) C^{\underline{j}}_{\underline{i}}\,. \tag{5.5}$$

We have, for instance, by using (4.11),

$$\widetilde{C}_{2^n} = \lambda_{2^n} C_{2^n}\,, \tag{5.6}$$

$$\widetilde{C}_{2^n,2^m} = \lambda_{2^m} C_{2^n,2^m}, \quad m > n\,, \tag{5.7}$$

and for any $t > 0$,

$$\widetilde{C}_{2^n}(t) = \lambda_{2^n} C_{2^n}(t)\,. \tag{5.8}$$

These relations, which give the Fourier coefficients of a measure in the new representation, show that Λ leads to a renormalization of the moments of all local variables. This renormalization is, however, time asymmetric. In fact, any observable (from L^2_μ) can be expanded as a superposition of the eigenfunctions of T,

$$f(x) = \int_\Gamma f\, d\mu + \sum_{n,\alpha} d_{n,\alpha}\, \psi_{n,\alpha}\,.$$

For $n > 0$, $\psi_{n,\alpha}(x) = U_n\, \psi_{0,\alpha}(x) = \psi_{0,\alpha}(S^{-n}x)$ represents *the past value of the observable $\psi_{0,\alpha}$ under the deterministic dynamics.* The expectation of $\psi_{n,\alpha}$ in a state ν is damped under the transition to $\tilde{\nu}$:

$$E_{\tilde{\nu}}(\psi_{n,\alpha}) = E_\nu(\Lambda\psi_{n,\alpha}) = \lambda_n E_\nu(\psi_{n,\alpha})\,.$$

Thus Λ introduces a monotonic damping of the excess to equilibrium of the past expectation values of the observables which may be interpreted as a loss of memory of the past deterministic evolution.

D. The rate of the approach to equilibrium

In the irreversible Kac model, the rate of the approach to equilibrium is uniformly exponential. That is, the Fourier coefficients in the master equation description decay as

$$\tilde{c}_{\underline{i}}(t) = e^{-\alpha t}\tilde{c}_{\underline{i}}(0)\,, \tag{5.9}$$

and the rate of the approach to equilibrium is uniform with respect to i:

$$|\tilde{c}_{\underline{i}}(t)| \leqslant \|\tilde{\rho}_t - 1\|_2^2$$

$$= \sum_{\underline{i}} |c_{\underline{i}}(t)|^2 = e^{-2\alpha t}\|\tilde{\rho}_0 - 1\|_2^2\,, \tag{5.10}$$

where α is positive and independent of ρ. This result holds for any initial distribution.

There is no such strong property in the deterministic Kac–de Haan model. First, as said above, the basin of μ under S_t is very small and many initial states do not converge to any limit. This basin is the same as the basin of μ under W_t (as shown in Ref. 11). Moreover, the (weak) approach to equilibrium as defined here is not generally uniform with respect to the local variables; this means that for any $t > 0$ there exist local variables $\eta_{\underline{i}}\epsilon_{\underline{j}}$ such that

$$E_{\nu_t}(\eta_{\underline{i}}\epsilon_{\underline{j}}) - E_\mu(\eta_{\underline{i}}\epsilon_{\underline{j}}) > a\,.$$

This property can be easily checked by using the identity of the family of all local variables with the eigenfunctions ψ_n of the time operator. A simple example can thus be constructed: let ν be a measure such that $E_\nu(\psi_n) = a > 0$, then for any $t > 0$ we have for the expectation at time t of the variable ψ_{n+t},

$$E_{\nu_t}(\psi_{n+t}) = E_\nu(U_{-t}\psi_{n+t}) = E_\nu(\psi_n) = a\,. \tag{5.11}$$

This shows that in the deterministic representation there is no uniform approach to equilibrium. One of the most important properties of W_t is that it entails an L^2_μ monotonic approach to equilibrium in the sense that for any initial state ν with a density $\tilde{\rho}$ from L^2_μ, the density of $\tilde{\nu}_t$, $\tilde{\rho}_t = W_t^*\tilde{\rho}_0$, approaches monotonically the uniform distribution,

$$\Omega_2(\tilde{\rho}_t) = \|\tilde{\rho}_t - 1\|_2 \searrow_0 t \to \infty\,,$$

and thus, as in (5.10), also uniformly for all local observables.

De la Llave[13] has studied the rate of decay of Ω_2. He has shown that there are initial states such that this global rate is slower than any given one, i.e., given $f(t) \to 0$ as $t \to \infty$, there exist initial states $\tilde{\rho} \in L^2_\mu$ such that, for any $M > 0$,

$$(\|\tilde{\rho}_t - 1\|_2)/f(t) > M\,,$$

for sufficiently great t. On the other hand, he displayed a class of states that decay exponentially with respect to this norm. *Thus Λ modifies the nature of the approach to equilibrium and also its rate.* We now investigate the uniform exponential decay of the expectations of the local variables under the usual hypotheses on the coefficients λ_i (see Ref. 8), that is, the sequence $\lambda_i = h(i)$ where $h(t) = e^{-\phi(t)}$, with ϕ differentiable, strictly increasing and convex, and $\phi(-\infty) = 0$ and $\phi(+\infty) = \infty$.

Now let us consider a measure ν which is transformed into square integrable state $\tilde{\rho}$. Recall that the local variables are the eigenfunctions $\psi_{n,\alpha}$ of Λ corresponding to the eigenvalues λ_n. It comes,

$$|\tilde{c}^{\underline{j}}_{\underline{i}}(t)| \leqslant \sum_{\underline{i},\underline{j}} |\tilde{c}^{\underline{j}}_{\underline{i}}(t)|^2$$

$$= \sum_{n,\alpha} |E_{\tilde{\nu}_t}(\psi_{n,\alpha})|^2$$

$$= \sum_{n,\alpha} |E_\nu(U_{-t}\Lambda\psi_{n,\alpha})|^2$$

$$= \sum_{n,\alpha} \lambda_n^2 |E_\nu(\psi_{n-t,\alpha})|^2$$

$$= \sum_{n,\alpha} \lambda^2_{n+t} |d_{n,\alpha}|^2\,, \tag{5.12}$$

where $d_{n,\alpha}$ denotes

$$d_{n,\alpha} = E_\nu(\psi_{n,\alpha})\,. \tag{5.13}$$

We shall show in Appendix E that for any constant k solution of the equation

$$k = \phi'(t)\,, \tag{5.14}$$

there exists a positive constant $A(K)$ such that

$$h(\lambda + t) \leqslant A(k) e^{-\lambda k} e^{-kt}\,. \tag{5.15}$$

In substituting this estimate into (5.12) we get

$$\sum_{\underline{i},\underline{j}} |\tilde{c}_{\underline{i}}^{\underline{j}}(t)|^2 \leqslant A(k)^2 e^{-2kt} \sum_{n,\alpha} e^{-2kn} |d_{n,\alpha}|^2 . \tag{5.16}$$

We may summarize this result in the following.

Proposition 4: There exists a positive constant K such that for any measure ν satisfying the following condition:

$$\sum_{n,\alpha} e^{-2Kn} |d_{n,\alpha}|^2 < +\infty , \tag{5.17}$$

the transformed states decay uniformly exponentially, that is,

$$\sup_{\underline{i},\underline{j}} |\tilde{c}_{\underline{i},\underline{j}}(t)|^2 \leqslant \| \tilde{\rho}_t - 1 \|^2 \leqslant \alpha(K) e^{-2Kt} . \tag{5.18}$$

The condition (5.18) characterizes a class of states which are not necessarily coarse-grained as in the work of de la Llave. The convergence of the series of the right-hand side is however only possible when $|d_{n,\alpha}|$ decays faster than an exponential as $n \to -\infty$.

From the above calculations it is clear that the coefficients λ_i act as a cutoff onto the contributions of the different eigenfunctions of T to ρ and it comes from (5.17) that the uniform and exponential approach to equilibrium results from the exponential damping of Λ to the $d_{n,\alpha}$ for $n > 0$ and the exponential convergence to 0 of these coefficients as $n \to -\infty$.

VI. CONCLUDING REMARKS

The model we have studied represents an infinite lattice gas with collision interaction. The mixing and Bernoulli properties permit us to characterize a class of nonequilibrium statistical states that go to equilibrium. In fact, by using the symbolic dynamics techniques, the model becomes equivalent to the observation of the history of the microscopic state of one particle in a fixed site, reducing the dynamics in this representation *to a Bernoulli shift*. This permits us to study the approach to equilibrium under the mixing. Yet there exists another class of initial states which go to equilibrium owing to their finite range correlations. These two mechanisms seem independent, for there exist states with infinite correlations that go to equilibrium under the mixing.

The Λ transformation which associates to the deterministic dynamics a monotonic irreversible stochastic evolution is constructed in terms of the symbolic dynamics. It realizes a damping of the past states of the "test" particle at the origin to its equilibrium value, thus introducing a renormalization of the mean values of all local observables. Namely, it transforms states with infinite correlations into states with quasilocal correlations.

We have also seen that Λ transforms the local approach to equilibrium into a global (and eventually exponential) approach to equilibrium which is much stronger than the first.

Yet the above study concerns only initial states which become absolutely continuous with finite entropy under the transformation. This is a very restricted class of states of the infinite systems, for which absolutely continuous states are quasilocal perturbations to equilibrium. A global divergence to equilibrium can be measured by mean observables such as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{\infty} E_{\tilde{\nu}}(\eta_i) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{\infty} \widetilde{C}_i .$$

For absolutely continuous states, this global mean excess to equilibrium, as well as the mean density of each kind of particle, is vanishing, for $\lim \widetilde{C}_i = 0$ as $i \to \infty$. An interesting problem is to consider the approach to equilibrium of translation invariant or periodic states, or more generally, states with nonvanishing mean density.

ACKNOWLEDGMENTS

We want to thank M. de Haan, F. Ledrappier, B. Misra, and I. Prigogine for helpful and interesting discussions.

APPENDIX A

We shall prove that if ν is absolutely continuous (a.c.) w.r.t. μ, then $c_{\underline{i}}^{\underline{j}} \to 0$ as $|i| + |j| \to \infty$, $|i|$ denotes $\max(i_1,\ldots,i_r)$.

For the sake of transparency we give the proof only for c_{i_1,i_2}. The same argument extends to the general case. Here we recall a Martingale theorem (see e.g., the Appendix of Ref. 7b) which states that ν is a.c. w.r.t. μ if and only if the Martingale $\rho_n(x)$ defined by (2.6) converges $-L_\mu{}^1$ to the density ρ of ν. Thus

$$\begin{aligned} c_{i_1,i_2} &= \int_\Gamma \eta_{i_1}\eta_{i_2}\rho \, d\mu \\ &= \int_\Gamma \eta_{i_1}\eta_{i_2}(\rho - \rho_n) d\mu + \int_\Gamma \eta_{i_1}\eta_{i_2}\rho_n \, d\mu. \end{aligned} \tag{A1}$$

Now for any $\epsilon > 0$, there is n_0 such that $\|\rho - \rho_n\|_1 < \epsilon$ for any $n > n_0$. Let us fix such an n. Thus in (A1) the first term is bounded by

$$\left| \int_\Gamma \eta_{i_1}\eta_{i_2}(\rho - \rho_n) d\mu \right| \leqslant \|\rho - \rho_n\|_1 < \epsilon. \tag{A2}$$

The remaining term $\langle \eta_{i_1}\eta_{i_2},\rho_n \rangle$ is vanishing for all (i_1,i_2) such that i_1 or i_2 is sufficiently great, for ρ_n contains in its Fourier expansion only products of η_k with $k < n$. Thus for all (i_1,i_2) such that $|\underline{i}| > n$, the scalar product of ρ_n with $\eta_{i_1}\eta_{i_2}$ is vanishing. As ϵ is arbitrary, it comes that $c_{i_1,i_2} \to 0$ as $|i| \to \infty$. The same proof works for $b_{\underline{i}}$.

APPENDIX B

de Haan has given the following formula similar to (4.11):

$$\epsilon_0(S^{2^n}) \equiv \epsilon_0(2^n) = \epsilon_{2^n-2}(0)\epsilon_{2^n-2^2}\cdots\epsilon_0(0)\eta_{2^n-1}(0). \tag{B1}$$

By using the two properties of the time inversion, $Ix = I(\nu,\epsilon) = (\epsilon,\eta)$ and $ISI = S^{-1}$, we obtain from (B1),

$$\begin{aligned} u_{-2^n} &\equiv \eta_0(S^{-2^n}x) = \epsilon_0(S^{2^n}Ix) \\ &= \eta_{2^n-2}(0)\eta_{2^n-2^2}(0)\cdots\eta_0(0)\epsilon_{2^n-1}(0). \end{aligned}$$

This is the desired formula.

APPENDIX C

We shall compute $\tilde{c}^{2^n}(t+1)$, $\tilde{c}_{2^n}(t+1)$ and $\tilde{c}_{2^n}^{2^n}(t+1)$ in terms of $\tilde{c}_\alpha^\beta(t)$. There is no loss of generality in taking $t=0$.

We first give the evolution law of the moments of $\eta_{\underline{i}}\epsilon_{\underline{j}}$ under the stochastic description. We have, from the definition,

$$\tilde{c}_{\underline{i}}^{\underline{j}}(t) = (\tilde{\Lambda}\nu_t)(\eta_{\underline{i}}\epsilon_{\underline{j}}). \tag{C1}$$

[In what follows, the notation $\nu(f)$, for any measure ν, denotes the expectation of the function f w.r.t. ν, $E_\nu(f)$.]

It comes from the relation $U_{-t}\Lambda = \Lambda W_t$ that

$$(\tilde{\Lambda}\nu_t)(\eta_{\underline{i}}\epsilon_{\underline{j}}) = \nu(U_{-t}\Lambda\eta_{\underline{i}}\epsilon_{\underline{j}}) = \tilde{\nu}(W_t\eta_{\underline{i}}\epsilon_{\underline{j}}). \tag{C2}$$

As $\eta_{\underline{i}}\epsilon_{\underline{j}}$ is an eigenfunction of Λ corresponding to some eigenvalue λ_α, we call it χ_α. To compute W_t we use the spectral form (5.6) of Λ in $L_\mu{}^2\Theta 1$,

$$\Lambda^{-1}U_t\Lambda = \Lambda^{-1}\sum_n \lambda_n U_{-t}E_n = \Lambda^{-1}\sum_n \lambda_n E_{n-t}U_{-t} = \sum_n \frac{\lambda_n}{\lambda_{n-t}} U_{-t}E_n. \tag{C3}$$

Inserting this relation into (C2) it becomes,

$$\tilde{C}_{\underline{i}}^{\underline{j}}(t) = (\lambda_\alpha/\lambda_{\alpha-t})\tilde{\nu}(U_{-t}\chi_\alpha) = (\lambda_\alpha/\lambda_{\alpha-t})U_t(\tilde{C}_{\underline{i}}^{\underline{j}}(0)), \tag{C4}$$

where $U_t(\tilde{C}_{\underline{i}}^{\underline{j}}(0))$ means that $\tilde{C}_{\underline{i}}^{\underline{j}}(0)$ evolves under the deterministic law (2.22). *This is the general evolution law for the moments in the stochastic description.* But in general, the difficulty lies in the calculation of λ_α. This however can be done for η_{2^n} and ϵ_{2^n} by writing them as functions of $\{u_i\}$. In (4.13) we found $\eta_{2^n} = \underline{U}_{-2^n}\underline{U}_{2^n}$; it corresponds to the eigenvalue λ_{2^n}. Now we compute the symbolic representation of ϵ_{2^n} by using the time inversion,

$$\epsilon_{2^n}(x) = \eta_{2^n}(Ix) = \underline{U}_{-2^n}(Ix)\cdot\underline{U}_{2^n}(Ix). \tag{C5}$$

Here we shall establish the following formula:

$$\underline{U}_n(Ix) = \underline{U}_{-n-1}(x) \quad \text{for any } n\in Z. \tag{C6}$$

Again using the time inversion, we get

$$\underline{U}_n(Ix) = \eta_0(S^n Ix) = \eta_0(IS^{-n}x) = \epsilon_0(S^{-n}x).$$

But $\epsilon_0(x) = \eta_0(S^{-1}x)$, thus we obtain the desired formula,

$$\underline{U}_n(Ix) = \eta_0(S^{-n-1}x) = \underline{U}_{-n-1}(x).$$

The symbolic expression of ϵ_{2^n} comes by substituting (C6) into (C5),

$$\epsilon_{2^n}(x) = \underline{U}_{2^n-1}(x)\underline{U}_{-2^n-1}(x). \tag{C7}$$

Therefore ϵ_{2^n} corresponds to the eigenvalue λ_{2^n+1}. Similarly, $\epsilon_{2^n}\eta_{2^n} = \underline{U}_{-2^n}\underline{U}_{2^n}\underline{U}_{-2^n-1}\underline{U}_{2^n-1}$ corresponds to λ_{2^n+1}. We finally get the three following formulas:

$$\begin{aligned}
\tilde{C}^{2^n}(1) &= (\lambda_{2^n+1}/\lambda_{2^n})\tilde{C}_{2^n}(0),\\
\tilde{C}_{2^n}(1) &= (\lambda_{2^n}/\lambda_{2^n-1})\tilde{C}_{2^n+1}^{2^n},\\
\tilde{C}_{2^n}^{2^n}(1) &= (\lambda_{2^n+1}/\lambda_{2^n})\tilde{C}_{2^n,2^n+1}^{2^n}(0).
\end{aligned} \tag{C8}$$

APPENDIX D

In order to investigate the new hierarchy, we compute the time evolution of the one-particle distribution function of $\tilde{\nu}_t$:

$$\tilde{\nu}_{t+1}(A_{2^n}^{(\eta,\epsilon)}) = (1/2^2)(1 + \tilde{C}_{2^n}(t+1)\eta + \tilde{C}^{2^n}(t+1)\epsilon + \tilde{C}_{2^n}^{2^n}(t+1)\eta\epsilon). \tag{D1}$$

To compute $\tilde{C}(t+1)$ in terms of $\tilde{C}(t)$, we use (C8) and get by substituting into (D1),

$$\tilde{\nu}_{t+1}(A_{2^n}^{(\eta,\epsilon)}) = \frac{1}{2^2}\left(1 + \frac{\lambda_{2^n}}{\lambda_{2^n}-1}\tilde{C}_{2^n+1}^{2^n}(t)\eta + \frac{\lambda_{2^n+1}}{\lambda_{2^n}}\tilde{C}_{2^n}(t)\epsilon + \frac{\lambda_{2^n+1}}{\lambda_{2^n}}\tilde{C}_{2^n,2^n+1}^{2^n}(t)\eta\epsilon\right). \tag{D2}$$

Here the reduced distribution function at time $t+1$ depends on the three-particle distribution function through $C_{2^n,2^n+1}^{2^n}(t)$.

This hierarchy is similar to the one of the deterministic evolution,

$$\nu_t(A_{2^n}^{(\eta,\epsilon)}) = (1/2^2)\left[1 + C_{2^n+1}^{2^n}(t)\eta + C_{2^n}(t)\epsilon + C_{2^n,2^n+1}^{2^n}(t)\eta\epsilon\right].$$

APPENDIX E

$$h(\lambda+t) = e^{-\phi(\lambda+t)} \leqslant e^{-kt}f(\lambda)$$

if and only if

$$g(\lambda,t) \equiv kt - \phi(\lambda+t) \leqslant \log f(\lambda). \tag{E1}$$

But, $g(\lambda,t)$ may admit an extremum as a function of t $T(\lambda)$ if the following equation has a solution for some constant k:

$$k - \phi'(\lambda + T(\lambda)) = 0. \tag{E2}$$

As Φ is convex, then Φ' is continuous and nondecreasing, thus such k exists, is necessarily positive, and

$$T(\lambda) + \lambda = \Phi'^{-1}(k) \equiv \alpha(k). \tag{E3}$$

In this case,

$$g(\lambda,t) \leqslant k(\alpha-\lambda) - \phi(\alpha(k)) \equiv M(k) - k\lambda. \tag{E4}$$

It follows that (E1) is satisfied if $f(\lambda)$ verifies the inequality,

$$M(k) - k\lambda \leqslant \log f(\lambda),$$

and this is equivalent to the condition

$$e^{M(k)}e^{-k\lambda} \leqslant f(\lambda). \tag{E5}$$

This condition shows that any such uniform estimate is of the exponential type as λ tends to $-\infty$. We conclude by this result, under the usual hypotheses onto $h(t)$:

There exist some positive constants k and $A(k)$ *such that*

$$h(\lambda+t) \leqslant A(k)e^{-kt}e^{-\lambda k}. \tag{E6}$$

Here, $A(k)$ denotes $e^{M(k)}$.

[1]M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience, New York, 1959).
[2]M. Dresden, in *Studies in Statistical Mechanics, I*, edited by de Boer and Uhlenbeck (1962), Vol. I, p. 303; M. Dresden and F. Feiock, J. Stat. Phys. **4**, 111 (1972); M. Coopersmith and G. Mandeville, *ibid.* **10**, 491, 405 (1974); J. Tavernier, *ibid.* **14**, 101 (1976).
[3]M. de Haan, Physica A **122**, 345 (1983).
[4](a) M. Courbage and I. Prigogine, Proc. Natl. Acad. Sci. **80**, 2412 (1983). (b) M. Courbage, Physica A **122**, 459 (1983).
[5]A. Akheizer and S. Peletminski, *Méthodes de la Physique Statistique* (Mir, Moscow, 1980).
[6]B. Misra, I. Prigogine, and M. Courbage, Physica A **98**, 1 (1979).
[7]M. Courbage and B. Misra, Physica A **104**, 359 (1980).
[8]S. Goldstein, B. Misra, and M. Courbage, J. Stat. Phys. **25**, 111 (1980).
[9]R. Balescu, *Equilibrium and Non-Equilibrium Statistical Mechanics* (Wiley–Interscience, New York, 1975).
[10]P. Walters, *An Introduction to Ergodic Theory* (Springer, Berlin, 1982).
[11]S. Martinez, Acta Appl. Math. **3**, 329 (1985).
[12]K. Sigmund, Math. Z. **138**, 183 (1974).
[13]R. de la Llave, J. Stat. Phys. **29**, 17 (1982).

A modified renormalization procedure which may avoid use of bare parameters

Luis P. Chimento and Alejandro S. Jakubi
Departamento de Física de la Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pab. I, Ciudad Universitaria, 1428 Buenos Aires, Argentina

(Received 28 April 1988; accepted for publication 4 January 1989)

A new renormalization procedure is introduced via a method that assigns finite values to divergent expressions in quantum field theory. It is shown that this procedure works at the two-loop level in Feynman diagrams with overlapping divergences. Applications to $\lambda\phi^4$ theory and quantum electrodynamics are made. Also, the Casimir effect is evaluated. Comparison of these results with those from standard renormalization methods proves them to be coincident.

I. INTRODUCTION

The need to extract finite, physical, sensible results out of ultraviolet divergent expressions occurring in perturbative quantum field theory has led to the devising of several renormalization schemes.[1–4] Their purpose is to prove that these divergences may be formally absorbed into the parameters defining the theory while locality, unitarity, and Lorentz invariance are maintained. Whenever the number of these parameters is finite, the theory is called renormalizable. Nonrenormalizable theories lack predictive power in a perturbative framework.

So we see that the "true" theory, that is, the theory that gives finite answers, is expressed in terms of divergent bare parameters. This is a consequence of the idea from classical analysis that convergence and limits are the only acceptable way to give meaning to derivatives, sums, and integrals.[5] For instance, the (finite) value of an integral is defined as the limit, whenever it exists, of a sum. Otherwise, if this sum does not converge, the integral is said to be infinite or divergent.

Divergent expressions are handled by means of a regularization technique. They are generalized as functions of a regulating parameter that makes them converge. Later on, a development in this parameter isolates the divergent terms. The finite remnant gives the renormalized expressions. Examples of regularization procedures are Pauli–Villars',[6] analytic,[7] and dimensional.[8]

The following facts of this standard renormalization procedure deserve to be noted.

(a) Finite results are obtained by reformulating the theory through the introduction of either divergent bare parameters or counterterms.

(b) Regularized expressions sometimes fail to satisfy symmetry properties of the original formal expressions.

(c) The need to carry along the regulating parameter lengthens (in general) the calculations.

So we think that it is legitimate to ask about an alternative scheme, leading from the original theory directly to renormalized expressions. This means to leave aside the unphysical bare parameters and avoid unnecessary symmetry violations. Such a scheme suggests the replacement of the usual mathematical methods by other ones, suitably modified.

In the present work we introduce, as a candidate for this alternative scheme, a technique we name operative continuation (OC). It assigns finite values to divergent integrals by means of the extrapolation of recurrence relations valid between convergent integrals. The objective of this technique is the calculation of renormalized expressions from those finite values.[9]

This paper is organized as follows. In Sec. II we develop the OC method by means of some examples. Section III shows how this technique works in Feynman diagrams of one and two loops (including those with overlapping divergences), and the comparison with the results from the usual renormalization procedures is made. We apply the OC method to $\lambda\phi^4$ theory and quantum electrodynamics, and preservation of gauge invariance is considered. Section IV presents calculation of the Casimir effect by the OC method, and finally in Sec. V the conclusions are presented.

II. OPERATIVE CONTINUATION METHOD

Let us begin with the integral

$$F(a) = \int_{-\infty}^{\infty} dx\, f(a,x), \tag{2.1}$$

where, for the sake of simplicity, f is a scalar function and a is a parameter. The function $F(a)$ is defined when the integral (2.1) exists, so that $f(a,x)$ belongs to a restricted class of integrands. The set of a values where convergence occurs will be called in what follows the convergent region (R_C).

The idea is to generalize the definition (2.1) to integrands f outside this class and for values of a outside R_C (that is, in the divergent region R_D). This means to consider integration as a linear functional relation I, $I: f \to F$ between functions of the parameter a, whether the integral is convergent or not. Operations such as derivation with respect to a parameter in the integrand or shifting the integration variable, which are justified on convergent integrals, are defined for this functional as usual. Whenever the integral is convergent to a finite value, the functional is assigned this value.

Let us see now how finite values may be given to divergent integrals. Take a suitable parameter of the integral as an index of a sequence of integrals. Then look for another parameter such that a linear differential operator in this variable acting on those integrals generates a recurrence relation

 0022-2488/89/081851-10$02.50

between them. This recurrence relation is valid only in R_C, and the operative continuation means to extend its validity to all values of the parameters. Through these relations, the values corresponding to R_D are calculated from those in R_C.

Let us see how this can be done in the simple case of the function $F(a,\alpha)$, defined by the integral

$$F(a,\alpha) \equiv \int_0^a dx\, x^\alpha = \frac{a^{\alpha+1}}{\alpha+1}, \tag{2.2}$$

convergent for $\alpha > -1$. We are interested in calculating $F(a,\alpha)$ in R_D, that is, for $\alpha \leqslant -1$. In R_C, F satisfies the recurrence relation

$$\frac{\partial F}{\partial a}(a,\alpha) = \alpha F(a,\alpha-1). \tag{2.3}$$

The operative continuation means, in this case, to extend the validity of (2.3) to every value of α.

The calculation of $F(a,\alpha)$ in R_D also requires derivation with respect to the limit a:

$$\frac{\partial F}{\partial a}(a,\alpha) = a^\alpha. \tag{2.4}$$

Comparing (2.4) with (2.3) we have

$$F(a,\alpha) = a^{\alpha+1}/(\alpha+1) \tag{2.5}$$

valid for $\alpha < -1$. It is remarkable that one gets an identical result by analytic continuation in α,[10] though the procedure is more complicated.

So to evaluate F at $\alpha = -1$, we integrate the relation (2.5) and get

$$F(a,-1) = \ln a + \ln C,$$

where C is an integration constant which gives an indetermined reference scale for a. To see why, we make a scale change in the integration variable $x \to \lambda x$, with λ real. This gives rise to a multiplicative group of scale transformations. Under it, $F(a,\alpha)$ remains invariant unless $\alpha = -1$. In this case, the transformation is equivalent to the change $C \to C/\lambda$. This means that the arbitrariness in the selection of the integration constant C is equivalent to the arbitrariness in the selection of the scale λ. Physically, this is linked with the arbitrariness in the selection of the unit of measure. So the indetermined constant may be considered as an indicator of the transformation properties $F(a,-1)$ has under scale changes.

Following the same steps, the integral between a and ∞, convergent for $\alpha < -1$, is generalized for $\alpha > -1$. When $\alpha = -1$ we get

$$\int_a^\infty \frac{dx}{x} = -\ln a - \ln C',$$

where C' is another integration constant.

Finally we arrive at

$$G(\alpha) \equiv \int_0^\infty dx\, x^\alpha = \int_0^a dx\, x^\alpha + \int_a^\infty dx\, x^\alpha = 0, \tag{2.6}$$

when $\alpha \neq -1$, the same result is obtained by analytic continuation.[10] Note that the integral $G(-1) = \ln C/C'$ is invariant under the scale group, that is, it has no scale. By continuity, it is given the value zero. Similarly, in the n-dimensional Euclidean space, we find

$$\int d^n x\, x^\alpha = \Omega_n \int_0^\infty dx\, x^{\alpha+n-1} = 0, \tag{2.7}$$

where Ω_n is the solid angle of a spherical hypersurface.

Due to its applications to the renormalization of the energy-momentum tensor in systems with boundaries or any periodicity condition that produces a discrete spectrum (like the Casimir effect, which will be calculated below), it is interesting to evaluate the sum

$$F(\alpha,a) \equiv \sum_{k=0}^{\infty} (k+a)^\alpha, \tag{2.8}$$

which is convergent for $\alpha < -1$ and, by definition, is $\zeta(-\alpha,a)$ (the generalized ζ function[11]). The usual procedure is to continue it analytically for $\alpha > -1$. Let us see how we can do the same by operative continuation. Making use of Bernoulli polynomials $B_n(x)$ (see Ref. 12) we can write

$$\sum_{k=r}^{s} (k+a)^n = \int_{r+a}^{s+a+1} dx\, B_n(x),$$

taking the corresponding limits

$$F(n,a) = \int_a^\infty dx\, B_n(x), \quad n = 0,1,..., \tag{2.9}$$

whose meaning is given as a generalized integral. Using the definition (2.8) we get the relation

$$\frac{\partial F}{\partial a}(\alpha,a) = \alpha F(\alpha-1,a), \tag{2.10}$$

which is operatively continued to $\alpha > -1$. On the other hand, derivating (2.9) with respect to a and comparing it with (2.10) we get

$$F(n,a) = -B_{n+1}(a)/(n+1), \tag{2.11}$$

so that

$$\sum_{k=1}^{\infty} k^n = F(n,0) = -\frac{B_{n+1}}{n+1}, \quad n = 0,1,2,..., \tag{2.12}$$

where $B_n \equiv B_n(0)$ are the Bernoulli numbers.

We are now going to study a four-dimensional example, related to the renormalization of Feynman integrals:

$$\int (N,D) \equiv \int d^4x \frac{\{x\}_N}{[(x-z)^2+\alpha]^D} = \int d^4x \frac{\{x\}_N}{[x^2 - 2x\cdot z + \beta]^D}. \tag{2.13}$$

The symbol $\int (N,D)$ constitutes a compact notation for the integrals, and $\{x\}_N$ denotes the tensorial product of N factors x. A contraction between two of these factors is written as $\int C(N,D)$, so the integral $\int C^n(N,D)$, with n contractions, transforms itself as a tensor of order $N-2n$. We have also $\beta = \alpha + z^2$. These integrals are convergent for $\omega = 4 + N - 2D < 0$, and their values for some usual combinations of N and D are listed in Appendix A.

For $\omega < 0$ we have the recurrence relations

$$\frac{\partial}{\partial \alpha} \int (N,D)|_z = -D \int (N,D+1), \tag{2.14a}$$

$$\frac{\partial}{\partial z} \int (N,D)|_\beta = 2D \int (N+1,D+1), \tag{2.14b}$$

$$\frac{\partial}{\partial \alpha}\frac{\partial}{\partial z}\int (N,D) = -2D(D+1)\int (N+1,D+2)$$
$$= \frac{\partial}{\partial z}\frac{\partial}{\partial \alpha}\int (N,D). \qquad (2.14c)$$

Equivalent relations exist for each contraction, with the restriction that the integrals must satisfy $N \geqslant 2n$. The operative continuation implies extending the validity of these relations for $\omega \geqslant 0$. The interrelation between the integrals in R_D and R_C that this extension produces is represented in Fig. 1 for the integrals with $n = 0,1$.

The calculation of the divergent integrals is based on the iterative integration of the relations (2.14a) and (2.14b), starting from the convergent ones. As we have seen before, each indefinite integration produces an arbitrary constant. Whenever the integral is convergent ($\omega < 0$), this constant is determined by the evaluation of the primitive in the integration limits. But if $\omega \geqslant 0$, those constants of integration are incorporated into the results, and their values are not determined beforehand.

In the case of iterative integration with respect to α, the integration constants turn out to be a polynomial P_ω in z of degree ω [because of (2.14a), the constants can only depend on z]. All the same, using (2.14b), the iterative integration in z produces undetermined polynomials B_ω of degree $\omega/2$ in β. Some results of integrating are shown in the table in Appendix A. Both modes of integration are combined by means of the operative continuation of the condition of compatibility expressed in (2.14c). This relates the coefficients of both polynomials.

III. APPLICATIONS TO QUANTUM FIELD THEORY

Now we are going to show how to use the operative continuation method in the calculation of divergent Feynman integrals up to the two-loop level including overlapping divergences. Application to $\lambda\phi^4$ theory follows, including a calculation of the one-loop effective potential. Also we deduce the renormalization group equation from finite renormalizations and evaluate its coefficients for this interaction.

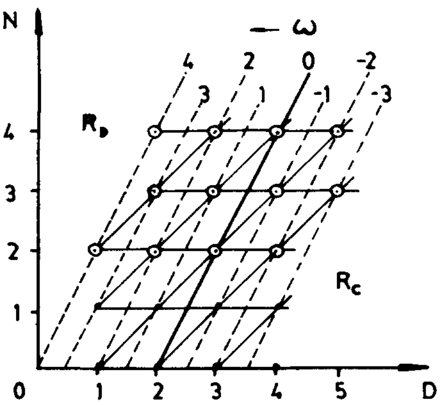


FIG. 1. Graphic representation of Eqs. (2.14a) and (2.14b). The integral $\int(N,D)$ is represented as a point ($\cdot$) of coordinates (N,D). The integral $\int C(N,D)$ is represented by a circle (O). The action of the operator $\partial/\partial z$ is equivalent to a diagonal movement ($\nearrow$), and that corresponding to $\partial/\partial\alpha$ as a horizontal movement ($\rightarrow$).

In addition, we calculate the three divergent diagrams of quantum electrodynamics to one-loop order, and take into account preservation of gauge invariance.

A. Calculation of one-loop Feynman integrals

By means of the change of variables $k = \mu x$, $p = \mu z$, $a^2 = \mu^2\alpha$, and $b^2 = \mu^2\beta$ in (2.13), where μ is a parameter with dimension of mass, we get an integral with dimension μ^ω:

$$\int (N,D) \equiv \int d^4k \frac{\{k\}_N}{[(k-p)^2 + a^2]^D}$$
$$= \int d^4k \frac{\{k\}_N}{[k^2 - 2k\cdot p + b^2]^D}. \qquad (3.1)$$

The new integration polynomial P_ω (B_ω) also gets dimension μ^ω and its argument changes to p (b^2). When the theory has a mass m, the polynomial coefficients turn into adimensional functions of m/μ times powers of m so that $P_\omega(p)$ and $B_\omega(b^2)$ are homogeneous functions of degree ω in momentum and mass. These considerations are valid for any number of contractions in the numerator of the integrand. Clearly the convergent integrals (like those in Appendix A) are determined, that is, they have no explicit dependence on $\mu[\mu^\omega\{z\}_N\alpha^{2-D} = \{p\}_N(a^2)^{2-D}]$. On the other hand, the divergent integrals depend on μ because the argument of the log terms is a/μ. So with the arbitrariness in the scale μ and in the integration polynomials, the divergent integrals become undetermined.

Employing the expressions from Appendix A in terms of dimensional variables and Feynman parametrization,[13] we are able to assign finite values to divergent Feynman integrals. Note that no divergent part must be subtracted so that no divergent counterterm must be added to the Lagrangian. All the calculations are performed using Feynman rules derived from the original Lagrangian, that is, in terms of the finite renormalized parameters. The arbitrariness in the integration polynomial is fixed, as usual, choosing suitable normalization conditions.[14] The OC values will be compared with the finite values we get after subtraction of the divergent parts of the integrals using counterterms by standard renormalization procedures. For instance, the method of Bogoliubov, Parasiuk, and Hepp (BPH) is based on subtracting from the integrand its Taylor development in the external momenta around zero, truncated at order ω. So the finite part is undetermined by a finite polynomial $F(p)$ of degree ω (see Ref. 2). Using the values given in Appendix A it is easy to check that the OC method gives the same expressions for the renormalized magnitudes up to a finite polynomial, which amounts to a finite renormalization.

It is also interesting to make the comparison with the results of dimensional regularization (DR). This implies calculating the integral for convergent values of space dimensionality d, continuing it analytically in the complex d plane and performing a Laurent development around $d = 4$. To keep the coupling constants dimension fixed, an arbitrary constant μ with dimension of mass must be inserted in the rules for calculating Feynman integrals.[8] Then renormalization implies subtraction of poles by counterterms with arbitrary finite parts, dependent in general on m/μ. So we must compare the integrals in Appendix A with the finite parts of dimensional regularized integrals in which those arbitrary finite functions are added. Also, we conclude that they are equal up to a finite renormalization. It is easy to check that the finite parts of DR integrals satisfy recurrence relations like (2.14).

The DR minimal subtraction (MS) scheme, that is, counterterms without finite parts,[15] is equivalent to OC with the integration polynomials set equal to zero. In this scheme each integral in Appendix A with no contraction is strictly equal to the finite part of the DR integral. For those integrals with contraction we get the MS values if we also seek invariance under shift in the integration variable.

B. Two-loop Feynman integrals

We give now a demonstration that OC assigns finite values at the two-loop level. It also shows that there is no difficulty in treating diagrams with overlapping divergences. The general two-loop integral (cf. Ref. 8), written in terms of adimensional variables, has the form

$$I_{ABC}(z) = \int d^4x\, d^4y \times \frac{1}{(x^2+\alpha)^A(y^2+\beta)^B[(z-x-y)^2+\gamma]^C}, \tag{3.2}$$

where A, B, and C are integers. Later on we will see the effect of inserting powers of x and y in the numerator of (3.2).

Now we observe some properties of $I_{ABC}(z)$. By the shift $x \to x-y+z$ we may interchange x^2 and $(z-x-y)^2$ between the first and the third terms. Similarly, we may interchange y^2 and $(z-x-y)^2$. So the role of the exponents A, B, and C is completely equivalent as long as the convergence of integrals is considered. For each pair (AB,BC,AC) there is an associated one-loop degree of divergence [e.g., for BC, $\omega_1 = 4-2(B+C)$]. There is also an overall degree of divergence: $\omega_2 = 8-2(A+B+C)$. The integral I_{ABC} is convergent when all these ω's are less than 0, and there it satisfies these recurrence relations:

$$\begin{aligned}&\frac{\partial}{\partial\alpha} I_{ABC}(z) = -A I_{A+1BC}(z),\\ &A\to B,\quad \alpha\to\beta,\\ &A\to C,\quad \alpha\to\gamma.\end{aligned} \tag{3.3}$$

We continue them to R_D, that is, where at least one $\omega \geqslant 0$.

If any of the exponents are equal to zero, the integral factorizes into two one-loop integrals. So in the following, A, B, and C will be greater than 0. Furthermore, all distinct cases are properly taken into account when $A \geqslant B \geqslant C$.

We may write

$$I_{ABC}(x) = \int d^4x \frac{1}{(x^2+\alpha)^A} J_{BC}(z-x), \tag{3.4}$$

where

$$J_{BC}(z-x) = \frac{\Gamma(B+C)}{\Gamma(B)\Gamma(C)} \int_0^1 du \int d^4y \times \frac{u^{C-1}(1-u)^{B-1}}{[(y-uz)^2+\delta]^{B+C}}, \tag{3.5}$$

and

$$\delta = u\gamma + (1-u)\beta + u(1-u)(z-x)^2. \tag{3.6}$$

We see that $J_{BC} \sim \mathcal{S}(0,B+C)$, and it is convergent unless $B=C=1$. So let us consider first the case $A>2$, $B=C=1$:

$$I_{A11}(z) = -\pi^2 \int_0^1 du \int d^4x \frac{1}{(x^2+\alpha)^A}(\ln\delta + P_0). \tag{3.7}$$

Now using (B4) to express (3.7) in terms of integrals like (2.13), and (B5),

$$I_{A11}(z) = \pi^2 \int_0^1 du\, dv \int d^4x \frac{(1-v)^A}{v} \times \left\{\frac{1}{[(x-vz)^2+\epsilon]^A} - \frac{1}{(x^2+\alpha)^A}\right\} + F_A(\alpha), \tag{3.8}$$

where

$$F_A(\alpha) = \pi^2\left[(2-P_0)\int(0,A) - L_A(\alpha)\right], \tag{3.9}$$

and

$$\epsilon = v(\gamma/(1-u) + \beta/u) + (1-v)\alpha + v(1-v)z^2 \tag{3.10}$$

has the form $v\sigma + (1-v)\alpha$. The integration on x is convergent and gives

$$I_{A11}(z) = \pi^4 \int_0^1 du\, dv \frac{(1-v)^A}{v}\frac{\Gamma(A-2)}{\Gamma(A)\epsilon^{A-2}} \times [1-(\epsilon/\alpha)^{A-2}] + F_A(\alpha). \tag{3.11}$$

The term between square brackets goes as v when $v\to 0$, so we conclude that (3.11) is finite.

Now integrating (3.3) for $A=2$ we get

$$I_{211}(z) = -\pi^4\left(\int_0^1 du\, dv \frac{1-v}{v}\ln\frac{\epsilon}{\alpha} + \frac{1}{2}\ln\alpha\right) + F_2(\alpha), \tag{3.12}$$

so that its integrand is also finite as $v\to 0$. One more integration gives

$$I_{111}(z) = -\pi^4\left\{\int_0^1 du\, dv\left[\sigma(\ln\epsilon - 1) + \frac{1-v}{v}\alpha\ln\frac{\epsilon}{\alpha}\right] + (\alpha/2)(\ln\alpha - 1)\right\} + F_1(\alpha), \tag{3.13}$$

and this integral is also finite.

If we go back to (3.5) we see that if there were a power of y in the numerator of J_{BC}, this would have given after integration a power of uz. All the same, powers of x would have given powers of vz. So we conclude that any two-loop integral is rendered finite by the OC method.

C. $\lambda\phi^4$ theory

As we wish to compare the previous results with those of DR, we proceed now to apply them to the divergent $1PI$ diagrams for the massive scalar field with potential $\lambda\phi^4/4!$, up to order λ^2. The perturbative expansions of proper functions $\Gamma^{(2)}$ and $\Gamma^{(4)}$ have the diagrammatic representation of Fig. 2 (see Ref. 16). Let us begin evaluating the tadpole (Fig. 3) in Euclidean metrics. Looking at the integral (0,1) in Appendix A we get immediately the result

$$-\frac{\lambda}{2}\int\frac{d^4k}{(2\pi)^4}\frac{1}{k^2+m^2}=-\frac{\lambda}{32\pi^2}m^2\left(\ln\frac{m^2}{\mu^2}-1-P_{01}\right). \tag{3.14}$$

All the same, we may evaluate the fish (Fig. 4) introducing a Feynman parameter x and the integral (0,2)

$$\frac{\lambda^2}{2}\int\frac{d^4k}{(2\pi)^4}\frac{1}{(p-k)^2+m^2}\frac{1}{k^2+m^2}=\frac{\lambda^2}{32\pi^2}\left\{-\int_0^1 dx\ln\left[\frac{m^2+x(1-x)p^2}{\mu^2}\right]+P_{02}\right\}. \tag{3.15}$$

At the two-loop level we have the double tadpole (Fig. 5),

$$\frac{\lambda^2}{4(2\pi)^2}\int d^4l\frac{1}{l^2+m^2}\int d^4q\frac{1}{(q^2+m^2)^2}=\frac{\lambda^2}{1024\pi^4}m^2\left(\ln\frac{m^2}{\mu^2}-1-P_{03}\right)\left(-\ln\frac{m^2}{\mu^2}+P_{04}\right), \tag{3.16}$$

and the setting sun (Fig. 6), in the massless limit:

$$\frac{\lambda^2}{6}\int\frac{d^4l}{(2\pi)^4}\int\frac{d^4q}{(2\pi)^4}\frac{1}{l^2q^2(p-1-q)^2}=\frac{1}{12}\left(\frac{\lambda}{16\pi^2}\right)^2p^2\left(\ln\frac{p^2}{\mu^2}-\frac{9}{2}-P_{05}\right), \tag{3.17}$$

where we have used (3.13) with $\alpha=\beta=\gamma=0$. These results are the same that DR gives, up to a finite renormalization.[16,17]

Inserting these results in the developments for proper functions, we get their renormalized values

$$\begin{aligned}\Gamma_R^{(2)}(p^2)=p^2&\left[1-\frac{\hat\lambda^2}{12}\left(\ln\frac{p^2}{\mu^2}-\frac{9}{2}-P_{05}\right)\right]\\&+m^2\left[1+\frac{\hat\lambda}{2}\left(\ln\frac{m^2}{\mu^2}-1-P_{01}\right)\right]\\&+\frac{\hat\lambda}{4}m^2\left(\ln\frac{m^2}{\mu^2}-1-P_{03}\right)\left(\ln\frac{m^2}{\mu^2}-P_{04}\right),\end{aligned} \tag{3.18}$$

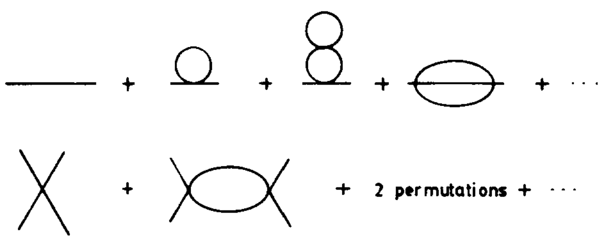


FIG. 2. Diagrammatic representation of proper functions.

FIG. 3. Tadpole diagram.

$$\Gamma_R^{(4)}(p_e)=-\lambda\left\{1+\frac{3}{2}\hat\lambda\left[\ln\frac{m^2}{\mu^2}+A(s,t,u)-P_{02}\right]\right\}, \tag{3.19}$$

where we have put $\hat\lambda\equiv\lambda/(16\pi^2)$ and (cf. Ref. 16)

$$A(s,t,u)=\frac{1}{3}\sum_{z=s,t,u}\left\{\int_0^1 dx\ln\left[1+\frac{x(1-x)z}{m^2}\right]\right\}. \tag{3.20}$$

The form of the arbitrary polynomial of $\Gamma_R^{(2)}(p^2)$ is $\sigma p^2+\rho m^2$, that is, two normalization conditions are required to fix the value of this function. One may require, for instance, the conditions that $\Gamma_R^{(2)}(p^2)$ has a zero at the physical mass m_F^2 and its derivative has unity value there.[14]

Similarly, the arbitrary constant P_{02} in $\Gamma_R^{(4)}(p_e)$ may be fixed by the requirement that it equals $(-\lambda_F)$, minus the physical coupling constant, at a suitable symmetry point in momenta space, determined by scale m_F^2. Otherwise we may choose the minimal prescription so that μ instead of either m_F or any other normalization momentum takes the place of dimensional scale.

D. Effective potential

The effective potential $V_e(\phi)$ is a very important concept in field theory.[18] In a sense, it is the quantum field potential energy. Also, it can be viewed as the generating functional for $1PI$ graphs with vanishing external momenta. So we can calculate it by summing all these graphs and inserting a factor ϕ (where ϕ is a constant average field) for each external line.

We evaluate now the one-loop effective potential of $\lambda\phi^4$ theory. Using either the technique of summing graphs[19] or the functional integral approach,[20] we arrive at the one-loop contribution

$$\frac{1}{2}\int\frac{d^4k}{(2\pi)^4}\ln\left(1+\frac{\lambda\phi^2/2}{k^2+m^2}\right). \tag{3.21}$$

This integral is divergent and we renormalize it via the OC method. First we rewrite it in terms of adimensional variables:

$$I(\alpha,\beta)\equiv\int d^4x\ln\left(1+\frac{\alpha}{x^2+\beta}\right), \tag{3.22}$$

where we have put $k=\mu x$, $\lambda\phi^2/2=\mu^2\alpha$, and $m^2=\mu^2\beta$. Now we note that $\partial I(\alpha,\beta)/\partial\alpha=\int(0,1)$ with the change $\alpha\to\alpha+\beta$. So integrating we have

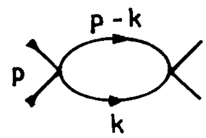


FIG. 4. Fish diagram.

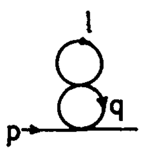


FIG. 5. Double tadpole diagram.

$$I(\alpha,\beta) = \pi^2 \tfrac{1}{2}(\alpha+\beta)^2[\ln(\alpha+\beta) - \tfrac{3}{2} - C] + D(\alpha+\beta) + E, \qquad (3.23)$$

where C, D, and E are integration constants.

Turning back to dimensional variables, adding zero-loop contribution and requiring $V_e(0) = 0$ (that is, setting normalization), we have at one-loop level

$$V_e^{[1]}(\phi) = \frac{\lambda}{4!}\phi^4 + \frac{1}{64\pi^2}\left(\frac{\lambda}{2}\phi^2 + m^2\right)^2 \times\left[\ln\left(\frac{\lambda\phi^2/2 + m^2}{\mu^2}\right) - \frac{3}{2} - C\right] - \frac{m^4}{64\pi^2}\left(\ln\frac{m^2}{\mu^2} - \frac{3}{2} - C\right) + \widetilde{D}\frac{\lambda}{2}\phi^2. \qquad (3.24)$$

If we take the $m = 0$ as a limit and impose the normalization condition $d^2V_e(\phi)/d\phi^2 = 0$ at $\phi = 0$, we get $\widetilde{D} = 0$. If we also require $d^4V_e(\phi)/d\phi^4 = \lambda$ at $\phi = M$, then the integration constant C is fixed at

$$C\left(\frac{M}{\mu}\right) = \ln\frac{\lambda M^2}{2\mu^2} + \frac{8}{3}, \qquad (3.25)$$

so that we obtain the Coleman–Weinberg result[19]

$$V_e^{[1]}(\phi) = \frac{\lambda}{4!}\phi^4 + \frac{\lambda^2}{256\pi^2}\phi^4\left(\ln\frac{\phi^2}{M^2} - \frac{25}{6}\right). \qquad (3.26)$$

E. Renormalization group equation

As we have seen, the OC method assigns to a divergent Feynman integral a finite but undetermined value. For instance, we may add freely to the integral $\int(N,D)$ a polynomial $P(p)$ of degree ω. This means that the integral turns into an equivalence class whose elements have the same (convergent) derivative $\partial^{\omega+1}\int(N,D)$ (here ∂ stands for $\partial/\partial p$ or $\partial/\partial m$).

Let us consider now those integrals contributing to the perturbation development of a proper function $\Gamma^n(p)$. Each choice we make of them gives a renormalization prescription for $\Gamma_R^{(n)}(p)$ by fixing its arbitrary polynomial. This prescription, in turn, relates to the scale κ of momenta at which proper functions are equated to the parameters in the Lagrangian.

The physical results of the theory, being invariant under

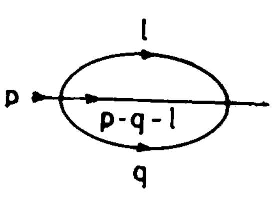


FIG. 6. Setting sun diagram.

transformations between prescriptions, give rise to the renormalization group. This group contains the transformations undergone by the input parameters when the scale κ is changed. The differential expression of this invariance leads to the renormalization group equation.[21]

Note that when all the P_ω's are put equal to zero (an MS-like prescription), we are left with scale μ. It may be freely multiplied by a real positive number s, so that we get another equivalence class: all the dimensional integrals obtained from the same adimensional integral via the rescaling we made in (3.1). Invariance of physical theory under the scale change $\mu \to s\mu$ leads to a particularly useful version of renormalization group equation (cf. Ref. 22).

We are going to consider the following $\lambda\phi^4$ theory, so that the parameters are the coupling λ, the mass m, and the scaling z of the field. Invariance under the scale change $\kappa_1 \to \kappa_2$ is expressed by the relation between proper functions

$$z_1^{-n/2}\Gamma_R^{(n)}(p;\lambda_1,m_1,\kappa_1) = z_2^{-n/2}\Gamma_R^{(n)}(p;\lambda_2,m_2,\kappa_2), \qquad (3.27)$$

where $\lambda_1 = \lambda(\kappa_1)$, $m_1 = m(\kappa_1)$, and $z_1 = z(\kappa_1)$.

To study the effects of this scale change we introduce the following parametrization: $\kappa_1 = \mu_0$, $\kappa_2 = e^t\mu_0 \equiv \mu(t)$. This means that the parameters turn into functions of t, and we rewrite (3.27) as

$$\Gamma_R^{(n)}(p;\lambda_0,m_0,\mu_0) = [z(t)]^{-n/2}\Gamma_R^{(n)}(p;\lambda(t),m(t),\mu(t)), \qquad (3.28)$$

where $\lambda(0) = \lambda_0$, $m(0) = m_0$, and $z(0) = 1$. So this invariance condition takes the form

$$\Gamma_R^{(n)}(0) = [z(t)]^{-n/2}\Gamma_R^{(n)}(t), \qquad (3.29)$$

from which we deduce the differential condition

$$\frac{\partial}{\partial t}\{[z(t)]^{-n/2}\Gamma_R^{(n)}(t)\}\Big|_{t=0} = 0, \qquad (3.30)$$

or

$$\left[\frac{\partial}{\partial t} + \frac{\partial\lambda}{\partial t}\frac{\partial}{\partial\lambda} + \frac{\partial m}{\partial t}\frac{\partial}{\partial m} - \frac{n}{2z}\frac{\partial z}{\partial t}\right]\Gamma_R^{(n)}(p;\lambda,m,\mu)\Big|_{t=0} = 0. \qquad (3.31)$$

This is the renormalization group equation deduced from finite renormalizations, so that it is suitable to use with the OC method as will be shown in what follows.

Now we define the coefficients of the equation as

$$\beta\left(\lambda_0,\frac{m_0}{\mu_0}\right) \equiv \frac{\partial\lambda}{\partial t}\Big|_{t=0}, \qquad (3.32a)$$

$$\gamma_m\left(\lambda_0,\frac{m_0}{\mu_0}\right) \equiv \frac{1}{m}\frac{\partial m}{\partial t}\Big|_{t=0}, \qquad (3.32b)$$

$$\gamma_d\left(\lambda_0,\frac{m_0}{\mu_0}\right) \equiv \frac{1}{2z}\frac{\partial z}{\partial t}\Big|_{t=0}, \qquad (3.32c)$$

so that we can write

$$\left[\frac{\partial}{\partial t} + \beta\frac{\partial}{\partial\lambda} + m_0\gamma_m\frac{\partial}{\partial m} - n\gamma_d\right]\Gamma_R^{(n)}(p;\lambda,m,\mu) = 0, \qquad (3.33)$$

where we perform the derivatives with λ_0, m_0, and μ_0 fixed. Using definitions (3.32) we have the developments of the parameters up to first order in t:

$$\lambda(t) = \lambda_0 + \beta t,$$
$$m(t) = m_0(1 + \gamma_m t),$$
$$z(t) = 1 + 2\gamma_d t,$$
$$\mu(t) = \mu_0(1 + t).$$

In what follows we will employ an MS-like prescription, so that the coefficients β, γ_m, and γ_d will depend only on λ_0. Moreover, they will have perturbative developments in powers of λ_0. We will evaluate the lowest-order terms in these developments employing (3.33) recursively in perturbation order and introducing expressions (3.18) and (3.19). We begin evaluating γ_m; it has a first-order contribution $\gamma_m^{[1]}$ which satisfies (where [1] means first perturbative order)

$$\frac{\partial \Gamma^{(2)[1]}}{\partial \ln \mu^2} + m_0^2 \gamma_m^{[1]} \frac{\partial \Gamma^{(2)[0]}}{\partial m^2} = 0. \tag{3.34}$$

We omit γ_d and β terms in (3.34) because their first significant contribution appears at order 2, as one realizes just looking at (3.18) and (3.19), respectively (they will be evaluated below). So we get

$$\gamma_m^{[1]} = \hat{\lambda}_0/2. \tag{3.35}$$

The coefficient γ_d may be evaluated to order 2 in the massless limit of (3.18); we have

$$\frac{\partial \Gamma^{(2)[2]}}{\partial \ln \mu^2} - \gamma_d^{[2]} \Gamma^{(2)[0]} = 0 \tag{3.36}$$

and

$$\gamma_d^{[2]} = \hat{\lambda}_0^2/12. \tag{3.37}$$

Finally we get β from

$$2\frac{\partial \Gamma^{(4)[2]}}{\partial \ln \mu^2} + \beta^{[2]} \frac{\partial \Gamma^{(4)[1]}}{\partial \lambda} = 0, \tag{3.38}$$

so that

$$\beta^{[2]} = 3\lambda_0 \hat{\lambda}_0. \tag{3.39}$$

It is interesting to verify that the values of the coefficients (3.35), (3.37), and (3.39), calculated by means of the finite renormalization group equation (3.33) and employing OC Feynman integrals, coincide with those calculated from the infinite renormalization group equation, employing invariance of bare parameters (bare theory gives finite results) under a change of renormalization scale.[23]

F. Quantum electrodynamics

We now apply the OC method to a higher spin theory, more specifically to the three serious primitively divergent one-loop graphs of quantum electrodynamics[13]: the vacuum polarization (Fig. 7—we are following conventions of Ref. 14),

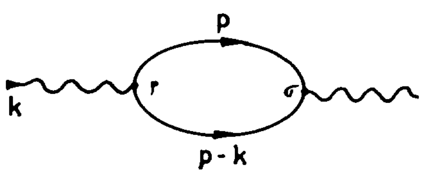


FIG. 7. Vacuum polarization diagram.

$$\overline{\omega}^{\rho\sigma}(k) = -(-ie)^2 \int \frac{d^4p}{(2\pi)^4} \mathrm{Tr}\Big(\gamma^\rho \frac{i}{\not p - m + i\epsilon} \times \gamma^\sigma \frac{i}{\not p - \not k - m + i\epsilon}\Big); \tag{3.40}$$

the electron propagator (Fig. 8),

$$\Sigma(p) = -i(-ie)^2 \int \frac{d^4k}{(2\pi)^4} \frac{-ig_{\rho\sigma}}{k^2 + i\epsilon} \times \gamma^\rho \frac{i}{\not p - \not k - m + i\epsilon} \gamma^\sigma; \tag{3.41}$$

and the vertex function (Fig. 9),

$$\Gamma_\mu(p',p) = (-ie)^2 \int \frac{d^4k}{(2\pi)^4} \frac{-ig_{\rho\sigma}}{k^2 - \overline{m}^2 + i\epsilon} \gamma^\rho \times \frac{i}{\not p' - \not k - m + i\epsilon} \gamma_\mu \frac{i}{\not p - \not k - m + i\epsilon} \gamma^\sigma, \tag{3.42}$$

where $\overline{m}$ is a small mass given to the photon to prevent infrared divergences. Also we take Feynman gauge in the photon propagator.[14] The γ's are Dirac matrices, and $g_{\mu\nu} = \mathrm{diag}\,(1, -1, -1, -1)$. For simplicity we omit, from now on, tensor indices.

Let us turn now to (3.40). Performing the trace and taking into account the properties of γ matrices, we get

$$\overline{\omega}(k) = \frac{\alpha}{\pi^3} \int_0^1 dx \int d^4p \times \frac{(p^2 - p\cdot k - m^2)g - p \vee (p-k)}{[(p - xk)^2 - (m^2 - x(1-x)k^2)]^2}, \tag{3.43}$$

which may be written in the notation of (3.1) as

$$\overline{\omega}(k) = \frac{\alpha}{\pi^3} \int_0^1 dx \Big\{ g\Big[\int C(2,2) - k\cdot \int (1,2) - m^2 \int (0,2)\Big] - 2\int (2,2) + k \vee \int (1,2)\Big\}, \tag{3.44}$$

where α is the fine-structure constant and we use the notation $(a \vee b)_{\mu\nu} = 2a_{[\mu}b_{\nu]}$. Now replacing the values of Appendix A in (3.44), and using an MS-like evaluation of polynomials,

$$\overline{\omega}(k) = \frac{i\alpha}{3\pi}(k^2 g - kk)\Big\{2\Big(1 + \frac{2m^2}{k^2}\Big) \times \Big(\sqrt{\frac{4m^2}{k^2} - 1}\ \mathrm{arccot} \sqrt{\frac{4m^2}{k^2} - 1}\Big) + \frac{1}{3} + \ln \frac{m^2}{\mu^2}\Big\}, \tag{3.45}$$

where $k^2 < 4m^2$. We see that this expression satisfies transversality ($k\cdot\overline{\omega} = 0$).

Next consider the propagator (3.41)

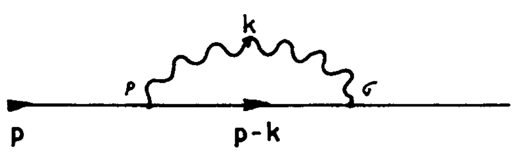


FIG. 8. Electron propagator diagram.

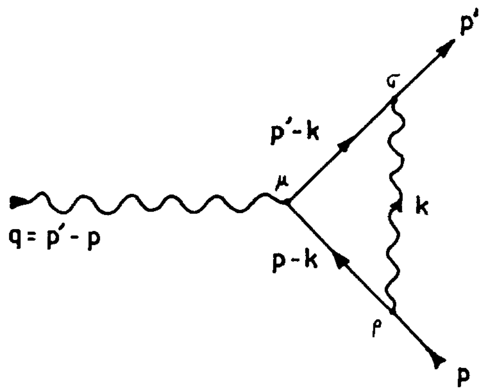


FIG. 9. Vertex function diagram.

$$\Sigma(p) = -\frac{i\alpha}{2\pi^3}\int_0^1 dx\left[(2m - \not{p})\int(0,2) + \int(\not{1},2)\right], \tag{3.46}$$

where the slash in $\int(\not{1},2)$ means contracting the integral with γ. Again, replacing the OC values and performing the x integrations we get

$$\begin{aligned}\Sigma(p) = \frac{\alpha}{2\pi}\Big\{&2m\left[\frac{m^2 - p^2}{p^2}\ln\left(1 - \frac{p^2}{m^2}\right)\right] \\ &- \frac{\not{p}}{2}\left[\frac{m^4 - p^4}{p^4}\ln\left(1 - \frac{p^2}{m^2}\right) + \frac{m^2}{p^2}\right] \\ &+ \left(2 - \ln\frac{m^2}{\mu^2}\right)\left(2m - \frac{\not{p}}{2}\right)\Big\},\end{aligned} \tag{3.47}$$

where $p^2 < m^2$.

Finally, the vertex function (3.42) has the form

$$\Gamma(p,p') = -\frac{i\alpha}{2\pi^3}\int_0^1 dx\int_0^x dy\int d^4k\,\frac{N(p,p')}{[Z(p,p';x,y)]^3}, \tag{3.48}$$

and restricting to on-shell values of p and p', Γ is to be considered as sandwiched between Dirac spinors, so that

$$\begin{aligned}N = 4[&\gamma(p' - k)\cdot(p - k) \\ &- \tfrac{1}{2}k^2\gamma + (p' + p - k)\not{k} - mk].\end{aligned}$$

We may write

$$\begin{aligned}&\int d^4k\,\frac{N}{Z^3} \\ &= 4\Big\{\gamma p'\cdot p\int(0,3) - \gamma(p' + p)\cdot\int(1,3) - m\int(1,3) \\ &\quad + (p + p')\int(\not{1},3) + \frac{1}{2}\gamma\int C(2,3) - \int(\not{1},3)\Big\}.\end{aligned}$$

Integrating over x and y parameters and taking the lowest order in $\overline{m}/m$, we get

$$\begin{aligned}\Gamma(p,p') = \gamma\frac{\alpha}{\pi}\Big[&\left(\ln\frac{\overline{m}}{m} + 1\right)(\theta\coth\theta - 1) \\ &- 2\coth\theta\int_0^{\theta/2} d\phi\,\phi\tanh\phi \\ &- \frac{\theta}{4}\tanh\frac{\theta}{2} + \frac{1}{4}\left(1 - \ln\frac{m^2}{\mu^2}\right)\Big] \\ &+ \frac{i}{2m}\sigma\cdot q\,\frac{\alpha}{2\pi}\,\frac{\theta}{\sinh\theta},\end{aligned} \tag{3.49}$$

where we have used the Gordon identity[16] with $\sigma = (i/2)[\gamma,\gamma]$ and put $q^2/m^2 = -4\sinh^2(\theta/2)$. We note that the values of these three integrals evaluated by the OC method coincide with those corresponding to standard renormalization procedures, up to a finite renormalization.

We also recall that the Green's functions of a gauge-invariant theory satisfy the Ward–Takahashi identities. The operations required to verify these identities are vector manipulations, partial fractioning of the product of two propagators, and shift of the integration variables.[24] All of these are allowed in OC integrals. Otherwise, as the OC method does not require leaving $d = 4$, Ward–Takahashi identities with objects like γ_5 are preserved after renormalization. Using (2.7), massless tadpoles vanish without recourse to *ad hoc* modifications of Gaussian integrals, as required by DR.[25] So we think that OC is a suitable alternative method for renormalization of Abelian and non-Abelian gauge theories.

IV. THE CASIMIR EFFECT

Another problem where renormalization must be carried out to get rid of divergences is known as vacuum polarization.[26] Here we will consider how the insertion of boundaries in space-time (that is, a change of topology) affects the vacuum expectation value of the energy-momentum tensor. More specifically, we will show how to evaluate the Casimir effect[27] by the OC method.

The relevant field in the Casimir effect is the electromagnetic field, and the manifold involved is the plane-parallel slab between two perfect conducting surfaces.

Consider first the vacuum state $|0\rangle$ of Minkowski unbounded space-time, the standard vacuum of particle physics. The expectation value of the energy-momentum tensor in this vacuum $\langle 0|T^{\mu\nu}|0\rangle$ diverges and so must be renormalized subtracting this divergence. The arbitrariness in this procedure is fixed if we require

$$\langle 0|T^{\mu\nu}|0\rangle_R = 0. \tag{4.1}$$

This subtraction corresponds to ignoring the zero-point energy of field oscillators, that is, normal ordering of field operators. The Minkowski vaccum serves as a standard against which all other vacua, whenever possible, are to be compared.

Let us consider now the slab manifold where we introduce coordinates x^μ, $\mu = 0,1,2,3$, oriented so that the x^3 axis is perpendicular to the plane of the conductors. The vacuum state of this quantum system will be denoted $|0_a\rangle$. By symmetry considerations it is clear that $\langle 0_a|T^{\mu\nu}|0_a\rangle$ must be diagonal and independent of x^0, x^1, and x^2. Moreover, because a perfect conductor remains a perfect conductor in any state of motion parallel to its surface, $\langle 0_a|T^{\mu\nu}|0_a\rangle$ must be invariant under Lorentz transformations that correspond to boosts parallel to the (x_1,x_2) plane. This means that the first three rows and columns of $\langle 0_a|T^{\mu\nu}|0_a\rangle$ must be proportional to the metric tensor of a $(2 + 1)$-dimensional Minkowski space, namely diag$(1, -1, -1)$. If to this inference one adds the observation that $T^\mu{}_\mu = 0$ in the case of the electromagnetic field, one concludes that $\langle 0_a|T^{\mu\nu}|0_a\rangle$ has the form

$$\langle 0_a|T^{\mu\nu}|0_a\rangle = f(a)\,\mathrm{diag}(-1,1,1-3). \tag{4.2}$$

The form of the function $f(a)$ may be determined by considering the work required to separate the conductors adiabatically. One obtains[26]

$$f(a) = A/a^4, \tag{4.3}$$

where A is some universal constant. The form of (4.3) may also be inferred by dimensional analysis. The only combination in which $\hbar$, c, and a can be united to yield an energy density is $\hbar c/a^4$. So we may put $A = \alpha\hbar c$, where α is a numerical constant whose evaluation requires explicit calculation of the density of energy $-f(a)$. We verify that (4.3) satisfies the renormalization condition (4.1) in the limit $a \to \infty$, that is, for unbounded Minkowski space.

For convenience we will calculate, instead of the density of energy, the total energy of the oscillators in the ground state inside a finite volume consisting of a rectangular box of transversal dimension L much larger than a (the perpendicular dimension). Then we divide the value of the energy by the box volume L^2a.

The only modes contributing to the sum $\Sigma\hbar\omega_\alpha/2$ are the transversal ones. Due to perfect conductor boundary conditions (**E** perpendicular and **B** parallel to the surface), modes are required to vanish on the surfaces. This implies that if a component k_3 is different from zero, it can take only the discrete values $k_3 = n\pi/a$ $(n = 1,2,...)$, and there are two states of polarization. Otherwise, for a vanishing k_3, only one mode survives. As we take $L \gg a$, the sums over the components parallel to the surfaces $\mathbf{k}_\parallel$ are replaced by integrals.

So the ground-state energy of the system is

$$\begin{aligned} E_0 &= \sum_\alpha \frac{1}{2}\hbar\omega_\alpha = \frac{\hbar c}{2}\sum_\alpha |\mathbf{k}_\alpha| \\ &= \frac{\hbar c}{2}L^2\int\frac{d^2k_\parallel}{(2\pi)^2}\Bigg[|\mathbf{k}_\parallel| \\ &\quad + 2\sum_{n=1}^{\infty}\left(k_\parallel^2 + \frac{n^2\pi^2}{a^2}\right)^{1/2}\Bigg] \\ &= \frac{\hbar c L^2}{4\pi}\sum_{n=-\infty}^{\infty}\int_0^\infty dk_\parallel\, k_\parallel\left(k_\parallel^2 + \frac{n^2\pi^2}{a^2}\right)^{1/2}. \end{aligned} \tag{4.4}$$

This integral diverges, so to renormalize E_0 we evaluate it by OC employing a result analogous to (2.5) and replacing it in (4.4):

$$E_0 = -\frac{\pi^2}{12}\frac{\hbar c L^2}{a^3}2\sum_{n=1}^{\infty}n^3.$$

This sum is also divergent, and we evaluate it using the result (2.12),

$$\sum_{n=1}^{\infty}n^3 = -\frac{B_4}{4} = \frac{1}{120},$$

so that we get

$$-f(a) = \frac{E_0}{L^2a} = -\frac{\pi^2}{720}\frac{\hbar c}{a^4}. \tag{4.5}$$

This determines the value of α as $\pi^2/720$ for the electromagnetic field, and coincides with the Casimir result as evaluated by other methods.[28] Other fields or boundary conditions imply a different numeric coefficient.

As we may appreciate, the OC method gives in this case the physical renormalization condition

$$\lim_{a\to\infty}\langle 0_a|T^{\mu\nu}|0_a\rangle_R = 0.$$

This is linked to the fact that the density of energy in the Minkowski space vacuum of the electromagnetic field is

$$\epsilon = 2\int\frac{d^3k}{(2\pi)^3}\frac{\hbar ck}{2},$$

and this scaleless integral vanishes due to (2.6).

V. CONCLUSIONS

The calculations of Secs. III and IV show that (at least in those cases) the OC method gives the same values for the renormalized quantities as the other standard procedures. No previous regularization was required, and only finite renormalizations had to be taken into account. Also, we found that the calculations get much simpler.

As we have seen, OC renormalization introduces a mass scale like any other method. This breaks symmetries related to absence of scale for massless fields. Otherwise, no additional deformation of the theory seems to be introduced. Invariance of the physical theory under scale changes leads to a finite renormalization group equation, as was shown in Sec. III E.

To qualify as a consistent renormalization scheme, OC should be shown to assign finite values to Feynman integrals of arbitrary perturbative order. In Sec. III B we have shown that this is possible for two-loop integrals. The proof requires one-loop results. This suggests that OC values for integrals of arbitrary perturbative order may be obtained by inserting the results for a given order in the next one. In forthcoming papers we will explore the possibility that OC satisfies a recursive, BPH-like scheme. Also, we will extend OC applications to supersymmetry and quantum field theory in curved space-time.

ACKNOWLEDGMENTS

We would like to thank Andrew Wrigley for his careful reading of the final draft.

L.P.C. and A.S.J. would like to thank, respectively, the Consejo Nacional de Investigaciones Científicas y Técnicas and the Comisión de Investigaciones Científicas de la Provincia de Buenos Aires for financial support.

APPENDIX A: CONVERGENT AND DIVERGENT INTEGRALS

We list now the value of integrals (2.13) for some convergent combinations of N and D (we use the Euclidean metric):

$$\int(0,D) = \frac{\pi^2}{\alpha^{D-2}}\frac{\Gamma(D-2)}{\Gamma(D)},$$

$$\int(1,D) = \frac{\pi^2}{\alpha^{D-2}}\frac{\Gamma(D-2)}{\Gamma(D)}z,$$

$$\int(2,D) = \frac{\pi^2}{\alpha^{D-2}}\frac{\Gamma(D-2)}{\Gamma(D)}\left[zz + \frac{\alpha}{2(D-3)}g\right],$$

$$\int C(2,D) = \frac{\pi^2}{\alpha^{D-2}}\frac{\Gamma(D-2)}{\Gamma(D)}\left[z^2 + \frac{2\alpha}{D-3}\right].$$

In the table below we show the values that the OC method assigns to several divergent integrals of frequent use. We employ integration of (2.14a) and integration constant subindices show their degree in z:

$$\int (0,2) = \pi^2(-\ln\alpha + P_0),$$

$$\int (0,1) = \pi^2[\alpha(\ln\alpha - 1 - P_0) + P_1],$$

$$\int (1,2) = \pi^2(-\ln\alpha + P_0')z,$$

$$\int (1,1) = \pi^2[\alpha(\ln\alpha - 1 - P_0') + P_2']z,$$

$$\int (2,3) = \pi^2\left[\frac{zz}{2\alpha} + g\left(-\frac{1}{4}\ln\alpha + Q_0''\right)\right],$$

$$\int (2,2) = \pi^2\left\{zz(-\ln\alpha + P_0'') + g\left[\frac{\alpha}{2}(\ln\alpha - 1 - 4Q_0'') + P_2''\right]\right\},$$

$$\int C(2,3) = \pi^2\left(\frac{z^2}{2\alpha} - \ln\alpha + P_0^c\right),$$

$$\int C(2,2) = \pi^2[-z^2\ln\alpha + \alpha(\ln\alpha - 1 - P_0^c) + P_2^c].$$

APPENDIX B: CALCULATIONS FOR SEC. III B

We are interested in writing a term like $\ln a/b^\mu$ as an integral over Feynman parameters. So we begin by calculating the integral

$$G(c) = \int_0^1 dx \frac{(1-x)^{\mu-1}}{x}\left[\frac{1}{(1+cx)^\mu} - 1\right] \tag{B1}$$

by expanding $(1+cx)^{-\mu}$ in powers of cx:

$$G(c) = \sum_{k=1}^\infty \frac{\Gamma(1-\mu)}{\Gamma(k+1)\Gamma(1-\mu-k)} \frac{\Gamma(\mu)\Gamma(k)}{\Gamma(\mu+k)} c^k = \sum_{k=1}^\infty \frac{\sin\pi(\mu+k)}{k\sin\pi\mu} c^k, \tag{B2}$$

where we have employed the reflection formula for the Γ functions.[12] The limit of $G(c)$ for the μ integer is

$$G(c) = -\ln(1+c). \tag{B3}$$

So by the replacement $c = (a-b)/b$ we get

$$\frac{\ln a}{b^\mu} = -\int_0^1 dx \frac{(1-x)^{\mu-1}}{x}\left(\frac{1}{z^\mu} - \frac{1}{b^\mu}\right) + \frac{\ln b}{b^\mu}, \tag{B4}$$

where $z = xa + (1-x)b$.

Now we see from the last term of (B4) that we will need integrals of the form

$$L_A(\alpha) \equiv \int d^4x \frac{\ln(x^2+\alpha)}{(x^2+\alpha)^A}. \tag{B5}$$

We may write

$$L_A(\alpha) = -\frac{\partial}{\partial A}\int d^4x \frac{1}{(x^2+\alpha)^A}, \tag{B6}$$

and get, for $A > 2$,

$$L_A(\alpha) = \frac{\pi^2}{(A-2)(A-1)\alpha^{A-2}} \times\left(\ln\alpha + \frac{1}{A-2} + \frac{1}{A-1}\right). \tag{B7}$$

Furthermore, these integrals satisfy, for $A > 2$, the recurrence relation

$$\frac{\partial}{\partial\alpha} L_A(\alpha) = -AL_{A+1}(\alpha). \tag{B8}$$

So by the OC method we get the values

$$L_2(\alpha) = -(\pi^2/2)(\ln^2\alpha + \ln\alpha + C), \tag{B9}$$

$$L_1(\alpha) = (\pi^2/2)[\alpha(\ln^2\alpha + \ln\alpha - 1 + C) + D]. \tag{B10}$$

[1] F. J. Dyson, Phys. Rev. **75**, 486 (1949); A. Salam, **82**, 217 (1951); S. Weinberg, *ibid.* **118**, 838 (1960).
[2] N. N. Bogoliubov and O. S. Parasiuk, Acta Math. **97**, 227 (1957); K. Hepp, Commun. Math. Phys. **2**, 301 (1966).
[3] H. Epstein and V. Glasser, Ann. Inst. H. Poincaré **XIX**, 211 (1973).
[4] C. Callan, Phys. Rev. D **2**, 1451 (1970); A. S. Blaer and K. Young, Nucl. Phys. B **83**, 493 (1974).
[5] J. J. Lodder, Physica A **120**, 30 (1983).
[6] W. Pauli and F. Villars, Rev. Mod. Phys. **21**, 434 (1949).
[7] C. G. Bollini, J. J. Giambiagi, and A. Gonzáles Domínguez, Nuovo Cimento **31**, 550 (1964); E. R. Speer, *Generalized Feynman Amplitudes* (Princeton U.P., Princeton, NJ, 1969).
[8] G. t'Hooft and M. Veltman, Nucl. Phys. B **44**, 189 (1972); C. G. Bollini and J. J. Giambiagi, Nuovo Cimento B **12**, 20 (1972).
[9] L. P. Chimento and A. S. Jakubi, *Proceedings of the VIth Latinamerican Symposium on Relativity and Gravity* (World Scientific, Singapore, 1988).
[10] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964).
[11] *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill, New York, 1953).
[12] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
[13] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, Berlin, 1976).
[14] C. Itzykson and B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[15] G. t'Hooft, Nucl. Phys. B **61**, 455 (1973); S. Weinberg, Phys. Rev. D **8**, 3497 (1973).
[16] P. Raymond, *Field Theory* (Benjamin, New York, 1981).
[17] D. J. Amit, *Field Theory, the Renormalization Group, and Critical Phenomena* (World Scientific, Singapore, 1984).
[18] K. Symanzik, Commun. Math. Phys. **16**, 48 (1970); S. Coleman, in *Laws of Hadronic Matter* (Erice, 1973), edited by A. Zichichi (Plenum, New York, 1977), p. 297.
[19] S. Coleman and E. J. Weinberg, Phys. Rev. D **7**, 1888 (1973).
[20] R. Jakiw, Phys. Rev. D **9**, 1686 (1974); J. Iliopoulos, C. Ittzikson, and A. Martin, Rev. Mod. Phys. **47**, 165 (1975).
[21] M. Gell-Man and F. E. Low, Phys. Rev. **95**, 1300 (1954); E. G. Stueckelberg and A. Peterson, Helv. Phys. Acta **26**, 499 (1953).
[22] G. t'Hooft, Nucl. Phys. **61**, 455 (1973); S. Weinberg, Phys. Rev. D **8**, 3497 (1973).
[23] J. C. Collins, Phys. Rev. D **10**, 1213 (1974).
[24] E. S. Abers and B. W. Lee, Phys. Rep. C **9**, 1 (1973).
[25] G. Leibbrandt, Rev. Mod. Phys. **47**, 849 (1975).
[26] B. S. DeWitt, in *General Relativity*, edited by W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).
[27] H. B. G. Casimir, Proc. K. Ned. Akad. Wet. **51**, 793 (1940).
[28] G. Plunien, B. Mueller, and W. Greiner, Phys. Rep. **134**, 87 (1986).

New higher-dimensional instanton and monopole solutions

Xu Jian-jun and Li Xin-zhou
Centre of Theoretical Physics, China Centre of Advanced Science and Technology (World Laboratory), Beijing, People's Republic of China and Department of Physics, Fudan University, Shanghai, People's Republic of China[a] *and, Institute for Theoretical Physics, Huadong Huagong University, Shanghai, People's Republic of China*

(Received 8 April 1988; accepted for publication 22 February 1989)

Two classes of spherically symmetric solutions to the empty space Einstein equations in even dimensions are presented. The solutions are constructed as radial extensions of circle bundles over a compact homogeneous Kähler manifold M. Type B solutions can be turned into higher-dimensional Kaluza–Klein monopoles. Both the Gross–Perry–Sorkin solution and the Bais–Batenburg solutions are contained as special cases of the general solutions.

I. INTRODUCTION

The discovery of self-dual instanton solutions in Einstein's theory of gravitation[1] has recently stimulated a great deal of interest in monopole solutions of Kaluza–Klein theory. The aim of the so-called Kaluza–Klein theories is to unify the ordinary gauge theories with gravity in a purely geometric way.[2] One postulates the existence of extra dimensions of space-time and interprets gauge transformations as generalized rotations taking place within these extra dimensions. Kaluza–Klein theories contain topological solitons which can be identified as magnetic monopoles.[3] An extensive listing of the general solutions has been made.[4] We have shown the monopole solutions in $(4+k)$-dimensional Abelian theories, which are dependent on the y coordinate of extra dimensions, and the average field strength is defined on the point of four-dimensional space-time.[5]

The first examples of asymptotically locally Euclidean metrics were the self-dual solutions given by Eguchi and Hanson.[1,6] Very general classes of four-dimensional manifolds which could admit self-dual asymptotically locally Euclidean metrics have been identified.[7] Bais and Batenburg[8] have suggested a new class of instanton solutions to Euclidean space Einstein equations in even dimensions. These instantons can be turned into higher-dimensional Kaluza–Klein monopoles by adding the time coordinate in a trivial way.

The aim of this paper is to extend the Taub–NUT solution and the Gross–Perry-Sorkin solution to the higher-dimensional ones. We present two classes of spherically symmetric solutions in $(2n+2)$-dimensional Euclidean space. We also show a new class of generalized monopole solutions in Kaluza–Klein theory. Both the Gross–Perry–Sorkin solution and the Bais–Batenburg solutions are contained as special cases of the general solutions.

This paper is organized as follows. In Sec. II we recall some basic properties of the circle bundles over the Kähler manifold, while in Secs. III and IV the spherically symmetric solutions in $(2n+2)$-dimensional space are studied. A new class of monopole solutions is discussed in Sec. V.

II. CIRCLE BUNDLES OVER KÄHLER MANIFOLDS

We recall some of the basic properties of the circle bundles over the Kähler manifold. Let us consider a Hermitian metric on M given by

$$dS^2 = 2g_{a\bar{b}}\, dz^a \otimes d\bar{z}^b, \tag{2.1}$$

where $g_{a\bar{b}}$ is a Hermitian matrix. One can define the Kähler form

$$K = ig_{a\bar{b}}\, dz^a \wedge d\bar{z}^b. \tag{2.2}$$

Then

$$\bar{K} = -i\bar{g}_{a\bar{b}}\, d\bar{z}^a \wedge dz^b = ig_{b\bar{a}}\, dz^b \wedge d\bar{z}^a = K \tag{2.3}$$

is a real two-form. A metric is said to be a Kähler metric if $dK = 0$, i.e., the Kähler form is closed. Then M is a Kähler manifold if it admits a Kähler metric. It is easy to see that $dK = 0$ is equivalent to

$$\Gamma^a{}_{\bar{b}c} = \Gamma^{\bar{a}}{}_{bc} = 0. \tag{2.4}$$

Since $g_{a\bar{b}}$ is a Hermitian matrix, the only surviving connection coefficients are therefore $\Gamma^a{}_{bc}$ and $\Gamma^{\bar{a}}_{\bar{b}\bar{c}}$. If K is not exact it implies the existence of a nontrivial solution to the source-free Maxwell equations on the Kähler manifold, with the field strength $F_{a\bar{b}} = 2ig_{a\bar{b}}$; this follows from the observation that K is a harmonic form, i.e.,

$$dK = 0, \quad \delta K = 0. \tag{2.5}$$

The Ricci curvature $R_{ab} = R^{\,c}{}_{acb} + R^{\,\bar{c}}{}_{a\bar{c}b}$, etc., takes a particularly simple form

$$R_{a\bar{b}} = -\partial_a \partial_{\bar{b}} \ln\sqrt{\det g}. \tag{2.6}$$

For a compact, homogeneous Kähler manifold, one has in addition that

$$R_{a\bar{b}} = cg_{a\bar{b}}, \tag{2.7}$$

where the coefficient c is some positive constant depending on the size of the manifold. Borel[9] proved that the compact homogeneous Kähler manifolds are of a type G/H, where G is a compact semisimple Lie group and H the centralizer of a torus in G. Given G, these may be obtained from a sequence of maximal embeddings $H_i \times U_1^i \subset H$ $(i = 1,\ldots,\text{rank } G;$ $H_0 = G)$. The Kähler manifolds are then the quotient spaces $M^{(i)} = G/U_1^{(1)} \times \cdots \times U_1^{(i)}$. The second Betti number for the manifold M is $b_2(M^{(i)}) = i$, implying that the second

[a] Mailing address.

cohomology group $H^2(M,Z)$ has at least i generators. As will become clear later, this is basically the topological criterion for the existence of Abelian Kaluza–Klein monopoles, where M corresponds to the spatial boundary of the space-time (base) manifold. The lowest nontrivial bundle E over such M is the manifold

$$E \simeq G/H_i, \tag{2.8}$$

with typical fiber $U_1^{(1)} \times \cdots \times U_1^{(i)}$. Trautman[10] has shown the Riemannian structure on the bundle E ($i = 1$ case). Given a connection ω on the bundle, $\omega = dy + A$, where y ($0 \leqslant y < 2\pi$) is the coordinate along the fiber and A a one-form on M related to the Kähler form $K = F = dA$ (locally), a natural metric is given by

$$dS_E^2 = dS_M^2 + B^2(\omega \otimes \omega), \tag{2.9}$$

where B^2 is some positive function on M.

III. THE SPHERICALLY SYMMETRIC SOLUTION IN (2n+2)-DIMENSIONAL SPACE: TYPE A

The aim of this section and the next section is to study the spherically symmetric solutions in arbitrary even-dimensional space. We consider a $(2n+2)$-dimensional space, where the signature of the metric $g_{\hat\mu\hat\nu}$ is $(+ + \cdots +)$, and the caret denotes a coordinate $x^{\hat\mu} = (r, z^a, \bar z^a, y)$, $a = 1,\ldots, n$ and $0 \leqslant y < 2\pi$. The general solutions are described as radial extensions of circle bundles over a compact homogeneous Kähler manifold M. The starting metric is the radial extension of Eq. (2.9).

$$dS^2 = D^2\,dr^2 + C^2\,dS_M^2 + B^2(\omega \otimes \omega), \tag{3.1}$$

where D, B, and C are functions of r only. Within this ansatz, one can find the following Ricci tensors:

$$R_{rr} = D^2\Big[-2n\Big(\frac{C'}{D}\Big)'\frac{1}{DC} - \Big(\frac{B'}{D}\Big)'\frac{1}{DB}\Big], \tag{3.2}$$

$$R_{yy} = B^2\Big[-\Big(\frac{B'}{D}\Big)'\frac{1}{DB} - 2n\frac{B'}{DB}\frac{C'}{DC} + \frac{1}{4}\frac{B^2}{C^4}(F_{a\bar b}F^{a\bar b} + F_{\bar a b}F^{\bar a b})\Big], \tag{3.3}$$

$$R_{a\bar b} = C^2\Big\{\frac{4g_{a\bar b}}{C^2} - \Big[\Big(\frac{C'}{D}\Big)'\frac{1}{DC} + \Big(\frac{C'}{DC}\Big)^2(2n-1) + \frac{B'}{DB}\frac{C'}{DC}\Big]g_{a\bar b} - \frac{1}{2}\Big(\frac{B}{C^2}\Big)^2 F_{a\bar c}F_{\bar b d}g^{d\bar c}\Big\}, \tag{3.4}$$

where

$$F_{a\bar b} = 2ig_{a\bar b}. \tag{3.5}$$

The Einstein equations for the metric (3.1) are

$$2n\Big(\frac{C'}{D}\Big)'\frac{1}{DC} + \Big(\frac{B'}{D}\Big)'\frac{1}{DB} = 0, \tag{3.6}$$

$$2n\Big(\frac{B^2}{C^4} - \frac{B'}{DB}\frac{C'}{DC}\Big) - \Big(\frac{B'}{D}\Big)'\frac{1}{DB} = 0, \tag{3.7}$$

$$\Big(\frac{B^2}{C^4} - \frac{B'}{DB}\frac{C'}{DC}\Big) - \Big(\frac{C'}{D}\Big)'\frac{1}{DC} + \frac{4}{C^2} - \Big(\frac{C'}{DC}\Big)^2(2n-1) - 3\frac{B^2}{C^4} = 0, \tag{3.8}$$

where we have chosen the constant c in Eq. (2.7) equal to 4. To arrive at these equations the Kähler property is essential. From the sum of Eqs. (3.6) and (3.7) one gets

$$(DB)^2 + [C'' - [(DB)'/DB]C']C^3 = 0. \tag{3.9}$$

To find the type A solution, we may choose

$$DB = qe^{-ar}, \tag{3.10}$$

and set

$$C(r) = f(r)e^{-(1/2)ar}, \tag{3.11}$$

where q and a are some constants and we take $a > 0$. Equation (3.9) then reduces to

$$[f'' - (a^2/4)f]f^3 + q^2 = 0. \tag{3.12}$$

This equation can be solved explicitly. If $q^2 > b^2/a^2$, we have a type AI solution,

$$f^2(r) = \frac{2}{a}\sqrt{q^2 - \frac{b^2}{a^2}}\,\sinh[a(r+d)] - \frac{2b}{a^2}. \tag{3.13}$$

If $q^2 < b^2/a^2$, we have a type AII solution,

$$f^2(r) = \frac{2}{a}\sqrt{\frac{b^2}{a^2} - q^2}\,\cosh[a(r+d)] - \frac{2b}{a^2}. \tag{3.14}$$

If $q^2 = b^2/a^2$, we have a type AIII solution,

$$f^2(r) = e^{a(r+d)} - 2b/a^2, \tag{3.15}$$

where b and d are constants of integration.

Substituting one of the equations, (3.13)–(3.15), into Eq. (3.8), one obtains the following equation for $D(r)$:

$$P(r) + Q(r)(D'/D) - R(r)D^2 = 0, \tag{3.16}$$

where

$$P(r) = \frac{2nq^2}{f^4} + \frac{(2n-1)b}{f^2} - \frac{(2n+1)af'}{f} + \frac{(2n+1)a^2}{2},$$

$$Q(r) = a - 2f'/f, \tag{3.17}$$

$$R(r) = 4e^{ar}/f^2.$$

Equation (3.16) can be solved explicitly: for type AI,

$$D^2 = e^{-3ar}\Big(a - \frac{\sqrt{q^2 - b^2/a^2}\,\cosh[a(r+d)]}{(1/a)\sqrt{q^2 - b^2/a^2}\,\sinh[a(r+d)] - b/a^2}\Big)^{-1}(\alpha - J_1)^{-1}; \tag{3.18}$$

for type AII,

$$D^2 = e^{-3ar}\left(a - \frac{\sqrt{b^2/a^2 - q^2}\,\sinh[a(r+d)]}{(1/a)\sqrt{b^2/a^2 - q^2}\,\cosh[a(r+d)] - b/a^2}\right)^{-1}(\alpha - J_{\rm II})^{-1}; \tag{3.19}$$

for type AIII,

$$D^2 = \frac{1}{a}e^{-3ar}\left[1 - \frac{1}{1-(2b/a^2)e^{-a(r+d)}}\right]^{-1}(\alpha - J_{\rm III})^{-1}; \tag{3.20}$$

where α and d are integral constants, and

$$J_{\rm I} = \frac{2e^{2ad}}{a^2\sqrt{q^2 - b^2/a^2}}\left[e^{-a(r+d)} + \frac{1 + b^2/(a^2q^2 - b^2)}{e^{-a(r+d)} - b/a\sqrt{q^2 - b^2/a^2}}\right], \tag{3.21}$$

$$J_{\rm II} = -\frac{2e^{2ad}}{a^2\sqrt{b^2/a^2 - q^2}}\left[e^{-a(r+d)} + \frac{1 + b^2/(b^2 - a^2q^2)}{e^{-a(r+d)} + b/a\sqrt{b^2/a^2 - q^2}}\right], \tag{3.22}$$

$$J_{\rm III} = \frac{2be^{2ad}}{a^3q^2}\left[e^{-a(r+d)} - \frac{a^2}{2b}\right]^2. \tag{3.23}$$

It is fortunate that the type AI solutions (3.13), (3.18) and (3.10), (3.11) are consistent with the remaining one in the equation system (3.6)–(3.8). Similarly, types AII and AIII solutions are also self-consistent. It could be seen from Eqs. (3.10) and (3.11), (3.13)–(3.15), and (3.18)–(3.20) that as r goes to infinity, both B and D approach zero while C approaches a constant (in the case of $a \neq 0$). So there is no problem of regularity in these solutions.

Next, we will show that the Taub–NUT solution is contained as a special case of the type AII solutions. In the $n = 1$ case, we have the following asymptotic expansions of the parameter:

$$C^2 = b\left[(r+d)^2 - \frac{q^2}{b^2}\right] - abr\left[(r+d)^2 - \frac{q^2}{b^2}\right] + \frac{a^2b}{4}\left[\frac{q^4}{b^4} - \frac{2q^2}{b^2}(r+d)^2 + \frac{1}{3}(r+d)^4 + 2r(r+d)^2 - \frac{2rq^2}{b^2}\right] + \cdots, \tag{3.24}$$

$$D^2 = \frac{b}{4}\,\frac{r + q/b + d}{r - q/b + d}\left[1 + a\left(\frac{q}{b} - 2b - 3r\right) + \cdots\right]. \tag{3.25}$$

If $a \to 0$ and we take the parameters $b = 1$ and $d = 0$, we find

$$D^2 = \tfrac{1}{4}(r+q)/(r-q), \quad C^2 = r^2 - q^2, \quad DB = q, \tag{3.26}$$

where the parameter q can take any real value, which is exactly the self-dual Taub–NUT solution in four dimensions. In general, the Bais–Batenburg solutions[8] are the particular cases in which the parameter a is zero.

IV. THE SPHERICALLY SYMMETRIC SOLUTION IN (2n+2)-DIMENSIONAL SPACE: TYPE B

To find the type B solutions, we may choose

$$DB = q/(ar+b)^2, \tag{4.1}$$

and set

$$C^2(r) = f^2(r)/(ar+b)^2. \tag{4.2}$$

Then Eq. (3.9) is reduced to

$$q^2 + f'' f^3 = 0, \tag{4.3}$$

which can be solved explicitly:

$$f^2(r) = (1/e)[(r+d)^2 - e^2q^2], \tag{4.4}$$

where e and d are constants of integration.

Substituting Eq. (4.4) into Eq. (3.8), on obtains the following equation for $D(r)$:

$$X(r) + Y(r)D'/D - Z(r)D^2 = 0, \tag{4.5}$$

where

$$X(r) = \frac{2nq^2}{f^4} + \frac{(2n+3)a^2}{(ar+b)^2} + \frac{2n-1}{ef^2} - \frac{2(2n+1)af'}{(ar+b)f},$$
$$Y(r) = 2[a/(ar+b) - f'/f], \tag{4.6}$$
$$Z(r) = 4(ar+b)^2/f^2.$$

The solution of Eq. (4.5) is

$$D^2(r) = \frac{F(r)}{(ar+b)^{2(2n+1)}Y^n(r)} \times\left[\gamma - 8\int_{r_0}^{r}\frac{F(s)ds}{(as+b)^{4n}Y^{n+1}(s)f^2}\right]^{-1}, \tag{4.7}$$

where

$$F(r) = \{\beta(ar+b)[(b-ad)(r+d) + ae^2q^2]\}^{1-n}, \tag{4.8}$$

in which β and γ are constants of integration. It is fortunate that the solutions (4.4) and (4.7) are consistent with the remaining one in the equation system (3.6)–(3.8). As r goes to infinity, D approaches zero while both B and C approach a constant (in the case of $a \neq 0$). There is also no problem of regularity in these solutions. So here we present the type B solution, and the Bais–Batenburg solutions[8] are the particular cases in which the parameter a is zero.

V. THE GENERALIZED GROSS-PERRY-SORKIN SOLUTION

The four-dimensional self-dual Taub–NUT solution can be turned into a five-dimensional Kaluza–Klein monopole by adding the time coordinate in a trivial way, which was first given by Sorkin and by Gross and Perry.[3] Type A and B solutions are clearly a higher-dimensional generalization of the self-dual Taub–NUT solution. Indeed, we may add the time coordinate in a trivial way so that

$$dS^2 = -dt^2 + D^2\,dr^2 + C^2 dS_M^2 + B^2(\omega \otimes \omega) \quad (5.1)$$

describes a magnetic type solution in $(2n+3)$-dimensional space-time compactified to some $(2n+2)$-dimensional space-time.

To find explicitly the generalized Gross–Perry–Sorkin solution, we should consider a type B solution; the case $n=1$ is a simple example. In the case of integral constant $b = ad$, we choose another integral constant

$$\gamma = -(2/a^4e^3q^4)(r_0 + d + e^2q^2/(r_0+d) + 2). \quad (5.2)$$

Then we have a generalized Gross–Perry–Sorkin solution

$$D^2(r) = eq^2(r+d-eq)/4(ar+b)^4(r+d+eq), \quad (5.3)$$

$$C^2(r) = [(r+d)^2 - e^2q^2]/e(ar+b)^2, \quad (5.4)$$

$$B^2(r) = 4(r+d+eq)/e(r+d-eq). \quad (5.5)$$

For large r one finds that

$$D^2(r) \sim eq^2/4(ar+b)^4, \quad C^2(r) \sim (r+d)^2/e(ar+b)^2, \quad B^2 \sim 4/e. \quad (5.6)$$

It is clear that the asymptotic behavior of the space is just a U_1 compactification, over some noncompact space with boundary M, where the internal circle has a radius R_∞:

$$R_\infty^2 = B^2(\infty) = 4/e. \quad (5.7)$$

In the case of $b > ad$, we choose

$$\gamma = -\frac{2}{a^2(ad-b)^2} - \frac{2e(r_0+d)}{a(ad-b)(ar_0+b)[(ad-b)(r_0+d) - ae^2q^2]}. \quad (5.8)$$

Then we obtain the second generalized Gross–Perry–Sorkin solution,

$$D^2 = \frac{a(b-ad)[(r+d)^2 - e^2q^2]}{4(ar+b)^4\{(r+d)^2 + e^2q^2 + [b/a - d - ae^2q^2/(ad-b) - e](r+d)\}}, \quad (5.9)$$

$$C^2 = [(r+d)^2 - e^2q^2]/e(ar+b)^2, \quad (5.10)$$

$$B^2 = \frac{4q^2[(r+d)^2 + e^2q^2 + (b/a - d - ae^2q^2/(ad-b) - e)(r+d)]}{a(b-ad)[(r+d)^2 - e^2q^2]}. \quad (5.11)$$

For large r one finds that

$$R_\infty^2 = 4q^2/a(b-ad). \quad (5.12)$$

In the case of $b < ad$, we choose

$$\gamma = \frac{2}{a^2(ad-b)^2} - \frac{2e(r_0+d)}{a(ad-b)(ar_0+b)[(ad-b)(r_0+d) - ae^2q^2]}. \quad (5.13)$$

Then we find

$$D^2 = \frac{a(ad-b)[(r+d)^2 - e^2q^2]}{4(ar+b)^4[(r+d)^2 + e^2q^2 - (d - b/a + ae^2q^2/(ad-b) - e)(r+d)]}, \quad (5.14)$$

$$C^2 = [(r+d)^2 - e^2q^2]/e(ar+b)^2, \quad (5.15)$$

$$B^2 = \frac{4q^2[(r+d)^2 + e^2q^2 - (d - b/a + ae^2q^2/(ad-b) - e)(r+d)]}{a(ad-b)[(r+d)^2 - e^2q^2]}. \quad (5.16)$$

For large r one finds that

$$R_\infty^2 = 4q^2/a(ad-b). \quad (5.17)$$

In the $a=0$, $b=1$, $e=1$, and $d=0$ case, from Eqs. (4.1), (4.2), and (4.7) we have

$$dS^2 = -dt^2 + \frac{1}{4}\frac{r-q}{r+q}\,dr^2 + (r^2 - q^2)dS_M^2 + \frac{4q^2(r+q)}{r-q}(\omega \otimes \omega). \quad (5.18)$$

This is exactly the Kaluza–Klein monopole solution first given by Sorkin and by Gross and Perry.[3]

Finally, the connection between the topological charges and magnetic charges of these solutions shoud be discussed. The classical vacuum solution of the five-dimensional Kaluza–Klein theory is assumed to be $M^4 \times S^1$, the direct product of four-dimensional Minkowski space and the compact manifold S^1. Any classical field configuration that approaches the vacuum solution at spatial infinity thus defines a circle bundle over the sphere at spatial infinity S^2. This bundle has the local structure of a direct product $S^2 \times S^1$; that is, the manifold S^1 sits on the top of every point of S^2. But it need not be a direct product globally. If the circle bundle over S^2 cannot be continuously deformed into the global direct product $S^2 \times S^1$, then the field configuration cannot be continuously deformed into the vacuum solution. To perform the topological classification of a circle bundle over S^2, we cut the sphere S^2 into two hemispheres along the equator. The circle bundle over the two hemispheres $S^2_{\uparrow}$ and $S^2_{\downarrow}$ are then easily deformed into direct product bundles $S^2_{\uparrow} \times S^1$ and $S^2_{\downarrow} \times S^1$ by performing coordinate transformations on each hemisphere. Along the equator, these two coordinate transformations must differ by a transformation that leaves the geometry of the S^1 invariant, that is, an isometry of S^1. Thus we can associate every circle bundle over S^2 with a loop in the isometry group U(1). The S^1 bundle over S^2 is topologically nontrivial if and only if the loop in U(1) has a nontrivial winding number. The solutions (5.3)–(5.5) [or (5.9)–(5.11), etc.] have U(1) magnetic charges.

ACKNOWLEDGMENTS

One of the authors (X.L.) thanks Dr. Ulli Wolff for insightful discussions during his stay at the Institut für Theoretische Physik der Universität Kiel.

This work was supported in part by the National Science Foundation of China under Grant No. KA 12038.

[1] T. Eguchi and A. J. Hanson, Phys. Lett. B **74**, 249 (1978).

[2] Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Math. Phys. Kl 966 (1921); O. Klein, Z. Phys. **37**, 895 (1926).

[3] R. D. Sorkin, Phys. Rev. Lett. **51**, 87 (1983); D. J. Gross and M. J. Perry, Nucl. Phys. B **226**, 29 (1983).

[4] I. G. Angus, Nucl. Phys. B **264**, 349 (1986).

[5] X.-Z. Li, F. Yu, and J.-Z. Zhang, Phys. Rev. D **34**, 1124 (1986).

[6] V. A. Belinskii, G. W. Gibbons, D. N. Page, and C. N. Pope, Phys. Lett. B **76**, 433 (1978).

[7] T. Eguchi and A. J. Hanson, Ann. Phys. (NY) **120**, 82 (1979).

[8] F. A. Bais and P. Batenburg, Nucl. Phys. B **253**, 162 (1985).

[9] A. Borel, Proc. Natl. Acad. Sci. **40**, 1147 (1954).

[10] A. Trautman, Int. J. Theor. Phys. **16**, 561 (1977).

Perturbative coherence in field theory

R. Aldrovandi and R. A. Kraenkel[a)]
Instituto de Física Teórica, Universidade Estadual Paulista, Rua Pamplona, 145-01405, São Paulo, Brazil

(Received 24 November 1987; accepted 9 November 1988)

The basic field equations of a field theory are not always derivable from a Lagrangian. Lagrangian theories are perturbatively coherent, in the sense that they have well-defined vertices. Non-Lagrangian theories may be coherent or not. Coherent theories are, in principle, quantizable by perturbative methods. The general condition for a theory to have well-defined vertices is given.

I. INTRODUCTION

A field model is defined by (i) listing all the fields involved, (ii) specifying their response to the transformations of interest, and (iii) giving the fundamental field equations and boundary conditions corresponding to the physical situation. The last item fixes the supposed dynamical behavior of the system. More frequently, a Lagrangian is given from which the field equations are obtained as the Euler–Lagrange equations. Or the action functional is given, which is global in character and, in principle, includes all the information concerning the system. This Lagrangian formulation has many advantages: it allows a simple treatment of symmetries and gives a workable way to quantization by Feynman path integral methods, which furthermore provide a convenient means to dispose of nonphysical degrees of freedom. Although most quantization methods suppose a Lagrangian, a field model can also be quantized by a perturbative procedure starting directly from the field equations, the Källèn-Yang-Feldman[1] (KYF) method. There is some advantage in working with the field equations, as not all systems of field equations are derivable from a Lagrangian.

It is commonly believed that non-Lagrangian theories are not amenable to quantization. It is one of our objectives here to show that this is not necessarily so. Field models may be perturbatively quantizable provided they satisfy a general condition that we shall find out. All Lagrangian theories will be seen to satisfy automatically that condition so that, as expected, all such theories can be quantized by the perturbative procedure. The condition is, however, less stringent than the condition for the existence of a Lagrangian, so that some non-Lagrangian theories can also be quantized. Notice that this has nothing to do with renormalizability, a question we shall ignore here. A coherent theory may turn out to be nonrenormalizable and consequently nonquantizable in a strict sense, but an incoherent model is defective in a much more primitive way: it has no well-defined vertices. Neither shall we worry about other possible difficulties that a model may come to exhibit in later stages of the quantization procedure. As the point we wish to make is so primitive, we shall resort to a naive formulation of quantization, which is what we are calling here *perturbative quantization*. We refer to the operational formulation of the KYF method as presented in the first volume[2] of Bjorken and Drell's text. It works with wave functions instead of noncommuting fields, but it is shown to be sound in the second volume.[3] Of course, difficulties appear when nonlinearities and/or constraints are present, but at least, in principle, they can be solved along the lines pioneered by Feynman in his "polish" paper.[4] This simple procedure, which roughly reduces quantization to "given the vertices and propagators, draw the Feynman graphs and calculate the S matrix elements" will be enough for our purpose, which is to give a general condition for the vertices to be well-defined. A simple example is the case of two scalar fields φ_1 and φ_2 obeying the equations of motion

$$\begin{aligned}(\Box + m^2)\varphi_1 &= g(\partial^\mu \varphi_1\, \partial_\mu \varphi_2)\varphi_2,\\ (\Box + m^2)\varphi_2 &= g(\partial^\mu \varphi_1\, \partial_\mu \varphi_2)\varphi_1.\end{aligned} \tag{1}$$

This is a non-Lagrangian system, as shown below, but there is no difficulty in quantizing it. Power counting tells us that it is nonrenormalizable in four-dimensional space-time but renormalizable in two dimensions. Starting from the field equations, perturbative quantization is performed in a very simple way. Let us recall the procedure for the model above: first, with the help of the Green's function of the differential operator, we write down the formal expressions for the solutions as

$$\varphi_1 = \varphi_1^{(\text{in})} + (\Box + m^2)^{-1} g(\partial^\mu \varphi_1\, \partial_\mu \varphi_2)\varphi_2, \tag{2}$$

$$\varphi_2 = \varphi_2^{(\text{in})} + (\Box + m^2)^{-1} g(\partial^\mu \varphi_1\, \partial_\mu \varphi_2)\varphi_1. \tag{3}$$

We are using a symbolic notation, omitting integrations. Then, to obtain the S matrix elements for each field, such general solutions, written as a sum of an ingoing free field plus interaction terms, are to be projected[2] on outgoing free fields of the same kind (and again integrated). The perturbative solution to a certain order is obtained by simply iterating the above equations, that is, replacing the fields in the interaction terms by the formal solutions and retaining terms up to the desired order. The Green's operator, once acted on from the left by the outgoing free field, gives again a free field. As a consequence, the vertices are obtained as the first contributions in the series with no remaining Green's functions. For example, projection of (2) on $\varphi_1^{(\text{out})}$ will lead to the vertex $g\varphi_1(\partial^\mu \varphi_1\, \partial_\mu \varphi_2)\varphi_2$. The same comes out from the projection of expression (3) on $\varphi_2^{(\text{out})}$: the vertex appears the same when seen from both (2) and (3). This seems natural enough and embodies what we shall understand by *perturbative coherence*. It would not be a property of the above model if the source terms in (1) were not carefully chosen. A trivial case of vertex incoherence would show up if in (1) the coupling constants in the two equations were different. Another example of this kind is given in Ref. 5. *Per-*

[a)] Fellow of FAPESP, São Paulo, Brazil.

 0022-2488/89/081866-05$02.50

turbative incoherence shows up when a vertex appears different when looked at from different channels. This would seem to be contrary to intuition, but our intuition is based on a familiarity acquired in Lagrangian models which are, as seen below, always coherent. Of course, in the case of incoherent models, the very notion of vertex looses its meaning, but we shall keep using the word "vertex" for simplicity of language. The model (1) will be incoherent if taken with two different coupling constants but coherent (although non-Lagrangian) with a unique coupling constant. In more involved incoherent models, it may even happen that a vertex that is seen in one channel is simply absent when looked at in another channel. This is, for instance, the case of the Poincaré gauge model.[5]

II. THE COHERENCE CONDITIONS

Suppose we have a set $\{\varphi^a\}$ of relativistic fields ($a = 1,2,\ldots,N$) submitted to a set of N equations of which we shall write only two:

$$K^a[\varphi] = J^a[\varphi], \tag{4}$$

$$K^b[\varphi] = J^b[\varphi]. \tag{5}$$

Here, K^a is the kinematical operator (Klein–Gordon, Dirac, etc.) acting on φ^a, K^b that acting on φ^b, and the J's are the source currents. The general form of a current functional involving j_1 fields of kind φ_1, j_2 fields of type $\varphi_2,\ldots, j_p$ fields of type φ_p, with a total of $k = \Sigma j_i$ fields, will be

$$\begin{aligned} J^a[\varphi] = \int d^4x_1\, d^4x_2 \cdots d^4x_k\, &\varphi_1(x_1)\varphi_1(x_2)\cdots \\ &\times\varphi_1(x_{j_1})\varphi_2(x_{j_1+1})\cdots\varphi_2(x_{j_1+j_2}) \\ &\times\varphi_3(x_{j_1+j_2+1})\cdots\varphi_3(x_{j_1+j_2+j_3})\cdots \\ &\times\varphi_p(x_k)C^a{}_{j_1 j_2\cdots j_p}(x_1,x_2,\cdots,x_k). \end{aligned} \tag{6}$$

If the current is a simple monomial in the fields, the coefficient $C^a{}_{j_1 j_2\cdots j_p}$ is a product of Dirac deltas. If the current involves derivatives of the fields, the coefficient will be a product of deltas and derivatives of deltas. When the current is a sum of terms involving different numbers of fields, it will be necessary for our purposes to examine each term separately, as they would correspond to distinct vertices. The general expression for the coefficient is, formally,

$$C^a{}_{j_1 j_2\cdots j_p} = \left.\frac{\delta^k J^a}{\delta\varphi_1(x_1)\cdots\delta\varphi_2(x_{j_1+1})\cdots\delta\varphi_p(x_k)}\right|_{\varphi=0}. \tag{7}$$

We shall again omit the integrations and put together fields of the same type, so as to rewrite (6) symbolically as

$$J^a[\varphi] = C^a{}_{j_1 j_2\cdots j_p}\varphi_1{}^{j_1}\varphi_2{}^{j_2}\ldots\varphi_a{}^{j_a}\ldots\varphi_b{}^{j_b}\ldots\varphi_p{}^{j_p}, \tag{8}$$

and analogously for $J^b[\varphi]$. Here no summation on repeated indices is implied. We have intentionally signaled the presence of the fields φ_a and φ_b. The S matrix element is obtained by projecting Eq. (4) on an outgoing field of type φ_a, or by projecting Eq. (5) on an outgoing field φ_b. The coherence condition will be, in the above compact notation,

$$J^a[\varphi]\varphi_a = J^b[\varphi]\varphi_b \tag{9}$$

(no summation on a, b). Comparing with (8) and using the purely multiplicative character of the coefficients (7), the condition may be put into the form

$$C^a{}_{j_1 j_2\cdots j_a\cdots j_b\cdots j_p} = C^b{}_{j_1 j_2\cdots(j_a+1)\cdots(j_b-1)\cdots j_p}. \tag{10}$$

An analogous reasoning holds for every pair of the N indices a, b, c, etc., so that we have in reality a whole series of $N!/[(N-2)!2!]$ conditions like (10). Use of (7) puts (10) into the form

$$\left.\frac{\delta^{k-1}}{\delta\varphi_1{}^{j_1}\delta\varphi_2{}^{j_2}\cdots\delta\varphi_a{}^{j_a}\cdots\delta\varphi_b{}^{j_b-1}\cdots\delta\varphi_p{}^{j_p}}\left[\frac{\delta J_a}{\delta\varphi_b} - \frac{\delta J_b}{\delta\varphi_a}\right]\right|_{\varphi=0} = 0, \tag{11}$$

for each pair of indices a, b with $a \neq b$. The "derivatives" in this expression are to be taken as functional (Fréchet) derivatives and, as such, as linear operators. When no derivatives on the fields are present in the J^a's, they will have the same algebra as usual derivatives. However, when derivatives are present, integrations by parts are to be carefully considered. Instead of going into these details here, we shall use another, simpler and more powerful formalism, which will allow the whole set of conditions (11) to be put into a simpler form.

III. THE CALCULUS OF FUNCTIONAL FORMS

Let us recall some properties of functional differential forms,[6] of which a less incomplete presentation has been given elsewhere.[7] Consider an action functional $S[\varphi]$, dependent on the fields $\varphi_1, \varphi_2, \cdots, \varphi_N$. Its variation δS will be a functional form of first degree, which can be written as

$$E = E_a \delta\varphi^a, \tag{12}$$

with

$$E_a = \frac{\delta}{\delta\varphi^a} S[\varphi]. \tag{13}$$

This is analogous to the differential $df = (\partial_i f)dx^i$ of a function f. The Euler–Lagrange equations coming from the action S, or from its integrand, the Lagrangian density, are, of course, $E_a = 0$. Expressions like (12) will be called functional *one-forms*, in analogy with the usual differential one-forms of calculus. In special, one-forms related to differential equations will be called *Euler forms*. The analogy with differential calculus goes, in reality, much further. Just as a general one-form $\omega = \omega_i\, dx^i$ is not necessarily the differential of a function (is not necessarily *exact*), a general one-form as (12) is not necessarily the variation of a functional. In this case, the corresponding equations $E_a = 0$ are not related to an action functional and are said to be *non-Lagrangian*. When does a Lagrangian exist for the equations? Once more the analogy with calculus is perfect: for the form ω to be locally the differential of a function, it is necessary and sufficient that $d\omega = 0$. For E to be locally an exact one-form, it is necessary and sufficient that $\delta E = 0$. The algebra of the exterior variations δ is formally the same algebra of the exterior differentials in calculus, the two-form δE being written as

$$\delta E = \left(\frac{1}{2}\right)\left[\frac{\delta E_a}{\delta\varphi^b} - \frac{\delta E_b}{\delta\varphi^a}\right]\delta\varphi^b \wedge \delta\varphi^a. \tag{14}$$

The formal analogy is complete indeed, provided the derivatives are interpreted as Fréchet derivatives. Acting on typical actions, which are 0-forms, such derivatives reduce to the usual Lagrangian derivatives. This analogy leads, in particular, to the boundary-has-no-boundary property $\delta^2 = 0$.

The condition $\delta E = 0$ for the existence of a Lagrangian for the equations $E_a = 0$ becomes, in view of (14), just the vanishing of the bracketed term. This is a new version of Vainberg's theorem,[8] which gives the conditions for a functional to be the functional derivative of another functional. Applied to the Euler form

$$J = J_a\, \delta\varphi^a \tag{15}$$

associated with the currents considered in Sec. II, we see that the Lagrangian condition is just the vanishing of the bracketed term in (11). As a consequence, every Lagrangian model satisfies (11) automatically and can be quantized in a coherent way. More properties of differential forms can be adapted to functional forms. One of them is the Poincaré lemma, which includes the above considerations about the existence of Lagrangians as a special case. Let W be any functional p-form

$$W[\varphi] = W_{a_1 a_2 \cdots a_p}[\varphi]\delta\varphi^{a_1} \wedge \delta\varphi^{a_2} \cdots \wedge \delta\varphi^{a_p}$$

and define its transformed TW as the $(p-1)$-form

$$TW[\varphi] = \sum_{j=1}^{p} (-)^{j-1} \int_0^1 dt\, t^{p-1} W_{a_1 a_2 \cdots a_p}[t\varphi]\varphi^{a_j}\delta\varphi^{a_1} \wedge \delta\varphi^{a_2} \cdots \wedge \delta\varphi^{a_{j-1}} \wedge \delta\varphi^{a_{j+1}} \cdots \wedge \delta\varphi^{a_p}. \tag{16}$$

Then the lemma says that W can always be written locally as

$$W[\varphi] = \delta TW + T\delta W. \tag{17}$$

We see that, when W is an Euler form E, then $\delta E = 0$ implies the existence of a Lagrangian $\Lambda = TW$. Another notion from differential calculus that can be implemented in the calculus of functional forms is that of a Lie derivative. On the space of the φ's, the components φ^a may be used as "functional coordinates." Fields (in the geometrical sense of the word) can be introduced, and the set of derivatives $\{e_a = \delta/\delta\varphi^a\}$ may be used as a "natural" local basis for them. A general field X will be written $X = X^a e_a = X^a\, \delta/\delta\varphi^a$. The Lie derivative L_X, acting on functional forms, will have properties analogous to those found in differential calculus. For example, suppose that X represents a transformation generator on the φ space. On forms, the transformation will be given by the Lie derivative L_X. As Lie derivatives commute with differentials, we have

$$L_X E = L_X\, \delta\Lambda = \delta L_X \Lambda. \tag{18}$$

Consequently, a symmetry of the Lagrangian ($L_X\Lambda = 0$) is a symmetry of the equation ($L_X E = 0$), but the equation may have symmetries which are not symmetries of the Lagrangian, a well-known fact. Other notions of differential calculus translate easily to functional forms, keeping furthermore analogous properties. Such is the case, for example, of the interior product $i_X W$ of a field X by a form W, which has the usual relation to the Lie derivative, $L_X W = i_X(\delta W) + \delta(i_X W)$.

In many of the considerations above some kind of metric is supposed. From (12) on we have been raising and lowering indices. Unless some special metric is at work in the model under consideration (such as the Killing–Cartan form in gauge models for semisimple groups), we shall simply suppose a metric of Euclidean type, which identifies components with higher and lower indices.

IV. A UNIFIED COHERENCE CONDITION

The coherence conditions acquire a simple form in this language. For each set of indices $(j_1, j_2, \ldots, j_p)$ in (11), define the functional one-form,

$$N_{(j_1 j_2 \cdots j_p)} = \sum_a \sum_{b \neq a} \frac{\delta^{k-1} J^a}{\delta\varphi_1{}^{j_1}\, \delta\varphi_2{}^{j_2} \cdots \delta\varphi_a{}^{j_a} \cdots \delta\varphi_b{}^{j_b - 1} \cdots \delta\varphi_p{}^{j_p}}\, \delta\varphi_a. \tag{19}$$

The coherence condition becomes then

$$\delta N_{(j_1, j_2, \cdots j_p)} = 0. \tag{20}$$

For each set $(j_1, j_2, \cdots j_p)$ in the model, the corresponding *coherence form* N must be closed. Notice that, by its very definition, the coefficients of N are linear in the fields (or some of its derivatives) and condition (20) requires N to be derivable from a certain 0-form bilinear in the fiels [which, by the way, is just the tranformed TN calculated by using (16)]. It is not difficult to check that N is a multiple Lie derivative of J with respect to the fields e_a constituting the natural field basis on the functional space:

$$N_{(j_1 j_2 \cdots j_p)} = \sum_a \sum_{b \neq a} \left[(L_{e_1})^{j_1} (L_{e_2})^{j_2} \cdots \times (L_{e_a})^{j_a} \cdots (L_{e_b})^{j_b - 1} \cdots (L_{e_p})^{j_p} \right](J). \tag{21}$$

So, although J is not necessarily derivable from a Lagrangian for the theory to be coherent, some of its multiple Lie derivatives must be. It is easier to interpret N in terms of Feynman graphs: given the vertex graph, to take a Lie derivative L_{e_j} corresponds to extracting a leg of type "j". Each term in (19) or (21) does the following: (i) fix a leg a; (ii) extract all other legs up to one of type b; (iii) exchange the two remaining legs; (iv) impose invariance under such exchange. The summation then spans all possible channels. This is a compact formalization of the "view-from-two-channels" discussion presented in Sec. II.

Consider the model given by Eq. (1). The current Euler-form will be

$$J = g(\partial^\mu \varphi_1\, \partial_\mu \varphi_2)[\varphi_2\, \delta\varphi_1 + \varphi_1\, \delta\varphi_2] \tag{22}$$

or

$$J = g(\partial^\mu \varphi_1\, \partial_\mu \varphi_2)\delta[\varphi_1 \varphi_2].$$

Its variation will be

$$\delta J = g\delta(\partial^\mu \varphi_1 \partial_\mu \varphi_2) \wedge \delta[\varphi_1 \varphi_2] \neq 0.$$

Consequently, there is no Lagrangian for Eqs. (1). Coherence is to be examined from the only set of fields of interest, $(j_1 = 1, j_2 = 1)$, to which corresponds the form

$$N_{(1,1)} = N^i{}_{(1,1)}\, \delta\varphi_i,$$

with

$$N^i{}_{(1,1)} = \frac{\delta^2 J^i}{\delta\varphi_1\, \delta\varphi_2}.$$

We find that

$$N_{(1,1)} = 2g\delta(\partial^\mu \varphi_1 \partial_\mu \varphi_2),$$

so that $\delta N_{(1,1)} = 0$. The model can be coherently quantized, despite its non-Lagrangian character.

V. THE CASE OF GAUGE FIELDS

The Euler form corresponding to the Yang–Mills equations is

$$E = [\partial_\mu F^{a\mu\nu} + f^a{}_{bc} A^b{}_\mu F^{c\mu\nu}]\delta A_{a\nu}, \tag{23}$$

where

$$F^{a\mu\nu} = \partial^\mu A^{a\nu} - \partial^\nu A^{a\mu} + f^a{}_{bc} A^{b\mu} A^{c\nu},$$

the f's being the structure constants of the Lie algebra of the gauge group, the generators T_a satisfying $[T_a, T_b] = f^c{}_{ab} T_c$. We shall rewrite the Euler form as

$$E\,[K^{a\nu} - J_Y{}^{a\nu} - J_X{}^{a\nu}]\delta A_{a\nu}, \tag{24}$$

where

$$K^{a\nu} = \partial_\mu(\partial^\mu A^{a\nu} - \partial^\nu A^{a\mu})$$

is the kinetic term,

$$J_Y{}^{a\nu} = -f^a{}_{bc}[\partial_\mu(A^{b\mu} A^{c\nu}) + A^b{}_\mu(\partial^\mu A^{c\nu} - \partial^\nu A^{c\mu})] \tag{25}$$

is the current leading to three-legged vertices, and

$$J_X{}^{a\nu} = -f^a{}_{bc} f^c{}_{de} A^b{}_\mu A^{d\mu} A^{e\nu} \tag{26}$$

is the current related to four-legged vertices. For each kind of vertex a coherence form must be defined. For the three-legged case,

$$N^Y{}_{(c\sigma)} = \frac{\delta J^{a\nu}}{\delta A^{c\sigma}}\, \delta A_{a\nu} \tag{27}$$

can be put, after a tedious but direct calculation, in the explicit form

$$\begin{aligned} N^Y{}_{(c\sigma)} &= \tfrac{1}{2}\{[f_{(bc)a} - f_{(ac)b} + f_{(ab)c}](\partial_\nu A^b{}_\sigma\, \delta A^{a\nu} - \partial_\nu A^{b\nu} A^a{}_\sigma) \\ &\quad + f_{(ac)b}[\partial_\sigma A^b{}_\nu - \partial_\nu A^b{}_\sigma]\delta A^{a\nu}\}. \end{aligned} \tag{28}$$

Here the symbol (ab) stands for symmetrization. The noticeable fact is that the structure constants appear always symmetrized in the first two indices. As a consequence, the above form will vanish identically for semisimple gauge groups and the coherence condition for three-legged vertices will be satisfied, in agreement with the fact that gauge models for semisimple groups are Lagrangian theories. Such is not the case for nonsemisimple groups,[9] for which the three-legged vertices are well defined only if the above form is closed indeed. The explicit expression of δN is

$$\begin{aligned} \delta N^Y{}_{(c\sigma)} &= \tfrac{1}{2}\{[2f_{(bc)a} - f_{(ac)b}]\delta A^a{}_\sigma \wedge \partial_\nu \delta A^{b\nu} \\ &\quad - \tfrac{1}{2} f_{(ab)c} \partial_\sigma \delta A^{b\nu} \wedge \delta A^{a\nu}\}. \end{aligned} \tag{29}$$

From (25), the coherence form for the four-legged vertices is obtained as

$$\begin{aligned} N^X{}_{(g\rho)(h\sigma)} &= \frac{1}{2} \frac{\delta}{\delta A^{g\rho}\, \delta A^{h\sigma}} \\ &\quad \times \{f_{ac(b} f^c{}_{d)e} A^b{}_\mu A^{d\mu} A^{e\nu}\} \delta A^a{}_\nu \\ &= \delta_{\rho\sigma} f_{ac(h} f^c{}_{g)d} A^{d\nu}\, \delta A^a{}_\nu \\ &\quad + f_{ac(h} f^c{}_{d)g} A^d{}_\sigma\, \delta A^a{}_\rho + f_{dc(g} f^c{}_{a)h} A^a{}_\rho\, \delta A^d{}_\sigma. \end{aligned} \tag{30}$$

Again indices between parentheses ( ) are symmetrized. The four-legged vertices are coherent if the above form is closed, which corresponds to the vanishing of

$$\begin{aligned} \delta N^X{}_{(g\rho)(h\sigma)} &= \tfrac{1}{2}\delta_{\rho\sigma}[f_{(ac)(h} f^c{}_{g)d} - f_{(cd)(g} f^c{}_{h)a}]\delta A^{d\nu} \wedge \delta A^a{}_\nu \\ &\quad + [f_{(ac)h} f^c{}_{dg} - f^c{}_{hg} f_{(ad)c} \\ &\quad - f^c{}_{ah} f_{(dc)g}]\delta A^d{}_\sigma \wedge \delta A^a{}_\rho. \end{aligned}$$

We see once more the coherence of the semisimple case; each term is proportional to a structure constant symmetrized in the two first indices.

In general, gauge models for nonsemisimple groups, besides being non-Lagrangian, cannot be coherently quantized by the perturbative method. An example has been found, by other means, in a model involving the Poincaré group.[10]

VI. FINAL COMMENTS

We have seen that non-Lagrangian models may have well defined vertices, provided they satisfy what we called the *coherence condition*. We have been rather strict in our language: incoherent models cannot be quantized by the usual techniques of perturbation theory because their vertices are not symmetric under the interchange of identical external legs and consequently the usual Feynman rules do not apply. Such asymmetry, however, is not a novelty in physics; it is a well-known property of vertices in dual models,[11] for which specially modified Feynmann rules must be introduced.[12] It is a curious point that gauge models for nonsemisimple groups groups exhibit it. Whether or not they have some relation to dual models is left for future consideration.

ACKNOWLEDGMENTS

This work was supported by FINEP. The work of one of the authors (RA) was partially supported by CNPq, Brazil.

[1] G. Källèn, Ark. Fys. **2**, 187, 371 (1950); C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).

[2] J. D. Bjorken and S. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[3] J. D. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).

[4] R. P. Feynman, Acta Phys. Polon. **24**, 697 (1963); B. S. DeWitt, Phys. Ref. **160**, 1113 (1967).

[5] R. Aldrovandi and J. G. Pereira, J. Math. Phys. **29**, 1472 (1988).

[6] P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).

[7] R. Aldrovandi and R. A. Kraenkel, J. Phys. A **21**, 1329 (1988).
[8] M. M. Vainberg, *Variational Methods for the Study of Non-Linear Operators* (Holden–Day, San Francisco, 1964); R. W. Atherton and G. M. Homsy, Stud. Appl. Math. **54**, 31 (1975); B. A. Finlayson, Phys. Fluids **15**, 963 (1972).
[9] R. Aldrovandi and J. G. Pereira, Rep. Math. Phys. **26**, 81 (1987).
[10] R. Aldrovandi and J. G. Pereira, Phys. Rev. D **33**, 2788 (1986); see also Ref. 5.
[11] S. Sciuto, Lett. Nuovo Cimento **2**, 411 (1969).
[12] J. L. Gervais and A. Neveu, Nucl. Phys. B **46**, 381 (1972).

Generalized Fourier transforms for nonlinear systems

J. Díaz Bejarano and A. Martín Sánchez
Departamento de Fisica, Facultad de Ciencias, Universidad de Extremadura, 06071 Badajoz, Spain

(Received 8 June 1988; accepted for publication 22 February 1989)

A simple generalization is presented of the usual Fourier transforms using generalized exponential and Fourier series developed previously for nonlinear systems. The expressions are given in terms of Jacobi elliptic functions in a form as close as possible to the Fourier transform. There is a pair of transforms for each value of the Jacobi parameter m. Simple applications are presented, one of which gives the generalized Yukawa potential and another that gives the generalized Breit–Wigner potential: Both correspond to the generalized Klein–Gordon equation.

I. INTRODUCTION

In general relativity the usual derivatives are replaced by covariant derivatives. Nonlinear operators and therefore, nonlinear fundamental oscillators, seem to be necessary ingredients of nonlinear quantum mechanics in nonflat space-times. The simplest equation of motion in the Schwarzschild metric implies terms in $(1/r^3)$. Rodríguez[1] showed that orbits in general relativity can be analyzed as equivalent problems of the anharmonic asymmetric oscillator (AAO); we were thus able to approximate the evaporation of primordial black holes by a very simple JWKB type of analysis.[2]

The simplest generalizations of the simple harmonic oscillator (SHO) in special relativity are equivalent to nonlinear ordinary differential equations of the type

$$\dot{x}^2 + Ax^2 + Bx^p = E, \quad \dot{x} \equiv \frac{dx}{dt}, \tag{1}$$

with $p = 4$ for the anharmonic symmetric oscillator[3] (ASO) and $p = 3$ for the AAO.[4] In any case it is evident that if one considers polynomials as the simplest nonlinear extensions of the linear fundamental oscillators, the simplest possible cases are the ASO and AAO.

In a previous paper,[5] we presented wavefunctions in the form of simple generalized real and complex exponentials which are solutions of many linear and nonlinear wave equations. The corresponding generalized Fourier series have also been developed.[6] In this paper, we develop the evident next step of the corresponding generalized Fourier transforms.

The series and transforms that we give are useful in all branches of nonlinear physics. For example, the generalized Fourier series have been used with a generalized harmonic balance method[7] to find approximate solutions of equations of the type

$$\ddot{x} + x^3 = \epsilon(1 - x^2)\dot{x}.$$

The generalized harmonic balance method has proven to be especially suited to the study of limit cycles.[8] In Refs. 1–8 and the present paper the results are presented in a form as similar as possible to the standard presentation in terms of the usual $\sin\phi$, $\cos\phi$ functions which are solutions of Eq. (1) if $\phi = \omega t$ when $B = 0$. Also, as in Refs. 1–8, we avoid mathematical complications as much as possible.

We believe that after reading Sec. V, it should not be difficult to calculate most of the generalized Fourier transform pairs which correspond to the most usual cases: Many are only different in that the argument of ϕ, $\sin\phi$, etc., is the elliptic function $\phi = \operatorname{am}(u,m)$ depending on the parameter $m = k^2$ of the Jacobi elliptic functions instead of the variable $u = \omega t$. The function ϕ changes with the parameter m and as a result there is a pair of transforms for every value of m. It is well known that for $m = 0$, $\operatorname{am}(u,0) = u$: The usual Fourier pair of transforms are the special case of our transforms when $m = 0$.

II. GENERALIZED EXPONENTIAL AND CIRCULAR FUNCTIONS AND THE CORRESPONDING FOURIER SERIES

In Ref. 5 we gave theorems on reciprocal and logarithmic derivatives of solutions of Eq. (1). These theorems are useful for a systematic derivation of different types of generalized exponentials, i.e., solutions of the type

$$x = \mathrm{R}e^{\pm i\phi(\omega t)} \text{ or } x = \mathrm{R}e^{\pm \phi(\omega t)}. \tag{2}$$

The solutions are simple generalizations of

$$x = \mathrm{R}e^{\pm i\omega t} \text{ or } x = \mathrm{R}e^{\pm \omega t}, \tag{3}$$

which are solutions of Eq. (1) when $B = 0$. A less systematic presentation was done in Ref. 6. Rewriting Eq. (1) with the changes $A = \omega^2 - BR^2$ and $E = R^2\omega^2 (p = 4)$, one has

$$\dot{x}^2/[1 + (Bx^2/\omega^2)] + \omega^2x^2 = R^2\omega^2. \tag{4}$$

The change of the independent variable $dT = [1 + (Bx^2/\omega^2)]^{1/2}\,dt$ transforms Eq. (4) into

$$\left(\frac{dx}{dT}\right)^2 + \omega^2x^2 = R^2\omega^2.$$

One sees that one possibility for the function $\phi(\omega t)$ of (2) is

$$\phi(\omega t) = \omega\int\left(1 + \frac{Bx^2}{\omega^2}\right)^{1/2} dt.$$

Now

$$x(t) = Rx_1(u) = R\,\mathrm{sn}\,\omega t, \quad B = -m\omega^2/R^2 \tag{5}$$

with $u = \omega t$; from the well-known properties of the elliptic functions,[9]

 0022-2488/89/081871-06$02.50

$$\phi_3(u) = \omega \int (1 - m\,\mathrm{sn}^2\,\omega t)^{1/2}\,dt$$

$$= \int dn\,u\,du = \int x_3(u)\,du.$$

Therefore,

$$y_3^{\pm} = e^{\pm i\phi_3(u)} = \mathrm{cn}\,u \pm i\,\mathrm{sn}\,u = x_2 \pm i x_1.$$

If $B = 0$, then $m = 0$, cn = cos, sn = sin, dn = 1, and then

$$\phi_3(u) = u = \omega t$$

and one has solutions in the form of Eq. (3). More details are given in Ref. 6.

One can also rewrite Eq. (4) as

$$\dot{x}^2/(R^2 - x^2) - Bx^2 = \omega^2. \tag{6}$$

The change of independent variable

$$dT = (R^2 - x^2)^{1/2}\,dt$$

transforms Eq. (6) into

$$\left(\frac{dx}{dT}\right)^2 - Bx^2 = \omega^2.$$

In this case the function $\phi(\omega t)$ of expression (2) is

$$\phi(\omega t) = B^{1/2} \int (R^2 - x^2)^{1/2}\,dt,$$

with the same values as (5). Using the systematic notation of Ref. 5 one now has

$$\pm i\phi_2(\omega t) = \pm i\omega k \int \mathrm{cn}\,\omega t\,dt = \pm ik \int \mathrm{cn}\,u\,du;$$

consequently,

$$y_2^{\pm} = e^{\pm i\phi_2(u)} = x_3 \pm ikx_1.$$

Observe that in this case

$$\cos\phi_2 = x_3 = \mathrm{dn}\,u, \quad \sin\phi_2 = kx_1 = k\,\mathrm{sn}\,u.$$

An interesting application of this result, used to find limit cycles of

$$\ddot{x} + Ax + 2Bx^3 = \epsilon(1 - x^2)\dot{x},$$

was described in Ref. 10.

In order to use the properties of generalized exponentials, in Ref. 6 we gave an approximation to an arbitrary function $f(\phi)$ defined in the interval $-\pi \leqslant \phi \leqslant \pi$ and following the method in Sommerfeld's well-known textbook,[11] with a $(2n + 1)$-term sum of the trigonometric functions

$$S_n(\phi) = A_0 + A_1\cos\phi + A_2\cos 2\phi + \ldots + A_n\cos n\phi$$
$$+ B_1\sin\phi + B_2\cos 2\phi + \ldots + B_n\sin n\phi.$$

Formally, this is the usual Fourier series, but with $\phi = \mathrm{am}(u)$. One can obtain the coefficients of the series in the usual way $(r > 0)$:

$$A_r = \left(\frac{1}{\pi}\right)\int f(\phi)\cos r\phi\,d\phi, \quad A_0 = \left(\frac{1}{2\pi}\right)\int f(\phi)d\phi,$$

$$B_r = \left(\frac{1}{\pi}\right)\int f(\phi)\sin r\phi\,d\phi,$$

where the limits of integration are $-\pi$ and π. As stated above, applications of these formulas have been made in Refs. 7, 8, and especially 10.

As can be seen from the above arguments or the more detailed discussion of Ref. 6, Eq. (1) with $p = 4$ also has the solution

$$x = Re^{\pm i\phi_3(u)} = R(\mathrm{cn}\,u \pm i\,\mathrm{sn}\,u),$$

but in this case

$$A = (1 - m/2)\omega^2, \quad B = m\omega^2/4R^2, \quad E = -\omega^2 R^2 m/4.$$

III. APPLICATION: GENERALIZED EXPONENTIAL SOLUTIONS FOR THE GENERALIZED KLEIN–GORDON EQUATIONS

Writing the generalized Klein–Gordon equation as

$$\Box\psi + \mu^2\psi + 2\lambda\psi^3 = 0, \quad \Box \equiv \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}, \tag{7}$$

and defining the unitary four-vector

$$\hat{k} \equiv (k^\circ, \mathbf{k}), \tag{8}$$

one has

$$\hat{k}\cdot\hat{k} \equiv k^{\circ 2} - k^2 = \mu^2. \tag{9}$$

Let

$$\tau = \hat{k}\cdot\hat{x}.$$

Then, using definitions (8) and (9), Eq. (7) becomes

$$\frac{d^2\psi}{d\tau^2} + \psi(\tau) + \frac{2\lambda}{\mu^2}\psi^3(\tau) = 0.$$

Multiplying by $d\psi/d\tau$ and integrating, one has

$$\left(\frac{d\psi}{d\tau}\right)^2 + \psi^2(\tau) + \frac{\lambda}{\mu^2}\psi^4(\tau) = R^2\omega^2, \tag{10}$$

where $R^2\omega^2$ is the integration constant. If $\lambda < 0$, $\mu^2 > 0$, and $0 < R^2\omega^2 < 4|\lambda|/\mu^2$ one periodic solution given in Sec. II is

$$\psi = R\,\mathrm{sn}(\omega\tau, m),$$

with $\omega^2 = 1 + (|\lambda| R^2/\mu^2)$ and $m = |\lambda| R^2/(\mu^2\omega^2)$. All the possible solutions for the possible μ^2, λ, and $R^2\omega^2$ signs and values are given in Ref. 5.

In Sec. II we saw that the change of variable

$$\mathcal{T} \equiv \int \left[1 + \left(\frac{\lambda}{\omega^2\mu^2}\right)\psi^2\right]^{1/2} d\tau \tag{11}$$

transforms Eq. (10) into

$$\left(\frac{d\psi}{d\mathcal{T}}\right)^2 + \omega^2\psi^2 = R^2\omega^2$$

or

$$\frac{d^2\psi}{d\mathcal{T}^2} + \omega^2\psi = 0.$$

The last equation can also be obtained from

$$\Box\psi + \omega^2\mu^2\psi = 0, \quad \Box \equiv \frac{\partial^2}{\partial T^2} - \frac{\partial^2}{\partial X^2} - \frac{\partial^2}{\partial Y^2} - \frac{\partial^2}{\partial Z^2},$$

$$\hat{K} = (K^\circ, \mathbf{K}), \tag{12}$$

and

$$\hat{K}\cdot\hat{K} = K^{\circ 2} - K^2 = \mu^2,$$

with the change

$$\mathscr{T} = \hat{K}\cdot\hat{X}$$

and with the same conditions on the constants.

As seen in Sec. II, for Eq. (10) we also have the exponential solutions

$$\psi = R \exp[\pm i\phi(\tau)] = R\,[\mathrm{cn}(\omega\tau,m) \pm i\,\mathrm{sn}(\omega\tau,m)],$$

but now with

$$A = 1 = (1 - m/2)\omega^2, \quad B = \lambda/\mu^2 = -m\omega^2/4R^2,$$
$$E = R^2\omega^2 m/4,$$

i.e., now $\omega^2 = A/(1 - m/2)$, etc.

Using (11), one can go from (7) to (12) with the changes

$$K^\circ T = k^\circ \int \left[1 + \left(\frac{\lambda}{\omega^2\mu^2}\right)\psi^2\right]^{1/2} dt,$$

$$\mathbf{K}\cdot\mathbf{X} = k \int \left[1 + \left(\frac{\lambda}{\omega^2\mu^2}\right)\psi^2\right]^{1/2} dx$$

because

$$d\mathscr{T} = K^\circ\, dT - \mathbf{K}\cdot d\,\mathbf{X}$$
$$= [1 + (\lambda/\omega^2\mu^2)\psi^2]^{1/2}(k^\circ\, dt - \mathbf{k}\cdot d\mathbf{x}).$$

In other words, using the many well-known solutions of Eq. (12), it is possible to transform the nonlinear Klein–Gordon equation into the linear one and vice versa. As explained in Sec. II, this is only one example. A more general change of variable is

$$d\mathscr{T} = (R^2 - \psi^2)^{1/2}\, d\tau.$$

IV. GENERALIZED FOURIER TRANSFORMS

We could follow the textbook of Sommerfeld[11] and other similar presentations to pass from Fourier series to Fourier integrals, but we prefer to show simply that under certain circumstances satisfied by the functions studied in the present paper, it is possible to determine a solution of the integral equation

$$F(\alpha) = \int_a^b f(\phi)K(\alpha,\phi)d\phi$$

in the form

$$f(\phi) = \int_a^b F(\alpha)H(\alpha,\phi)d\alpha,$$

with a and b finite or infinite and $K(\alpha,\phi)$ a Fourier kernel.[12] In the following we shall look at the most usual cases.

(i) We look at the generalized Fourier exponential transform pairs (to be called simply generalized Fourier transforms in what follows)

$$F(\alpha) = \int_{-\infty}^{\infty} f(\phi)e^{-i\alpha\phi}\, d\phi$$

and

$$f(\phi) = \left(\frac{1}{2\pi}\right)\int_{-\infty}^{\infty} F(\alpha)e^{i\alpha\phi}\, d\alpha.$$

(ii) We look at the generalized Fourier cosine transform pairs ($\phi > 0$)

$$F_c(\alpha) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty f(\phi)\cos(\alpha\phi)d\phi$$

and

$$f(\phi) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty F_c(\alpha)\cos(\alpha\phi)d\alpha.$$

(iii) We look at the generalized Fourier sine transform pairs

$$F_s(\alpha) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty f(\phi)\sin(\alpha\phi)d\phi$$

and

$$f(\phi) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty F_s(\alpha)\sin(\alpha\phi)d\alpha.$$

It is a straightforward matter to demonstrate the fundamental properties of the generalized Fourier transforms: They are the same properties as for the usual Fourier transforms (change of sign, conjugate, linearity, scaling, shifting, derivative, etc.)[12] except that now it is necessary to take into account the parameter m, because ϕ and α depend on this parameter.

V. EXAMPLES

The usual Fourier transforms have been used so extensively on mathematical and physical problems of mainly the linear type that it is difficult to predict which ones will correspond to the most interesting nonlinear cases. In our present selection, one criterion was the simplicity of the function. Other criteria were that the Fourier transforms should be common textbook examples[13,14] and nonlinear generalizations of very well-known functions of the linear theories such as, for example, the Lorentzian or Breit–Wigner resonance formula $1/(a^2 + x^2)$. To simplify comparison with the usual transforms, ϕ_3 is used for these examples, i.e., $\cos\phi = \mathrm{cn}\, u$ and $\sin\phi = \mathrm{sn}\, u$.

Figures 1–4 show some examples of the most common Fourier transform pairs. In all cases there are four curves belonging to different values of the parameter: $m = 0.0$ (usual Fourier transformation), $m = 0.5$, $m = 1 - 10^{-3}$, and $m = 1 - 10^{-9}$. The functions are plotted versus the variable $v = \pi u/2K$, where K is the complete elliptic integral of the first kind, i.e., the functions have all been reduced to the fundamental period.

In Fig. 1 one sees the simplest example of Fourier transform studied in any of the textbooks about the subject: the Gaussian curve $f(\phi) = A \exp(-a^2\phi^2)$. To fix our ideas we have chosen $A = 10$ and $a = 0.3$. We could plot $f(\phi)$ vs $\phi = \mathrm{am}\, u$ or u and then the curves in Fig. 1(a) would be Gaussian, but the greater the parameter m the greater the full width at half-maximum. In the plot of $f(\phi)$ vs $v = \pi u/2K$ (reduction to the fundamental period) one sees the nonlinear character of the functions defined using elliptic functions because inflexion points (for $m \neq 0$) appear at regular intervals ($n\pi/2, n = 0, \pm 1, \pm 2, ...$). All the curves are represented by the same function, where the only change is in the value of the parameter m. The corresponding gener-

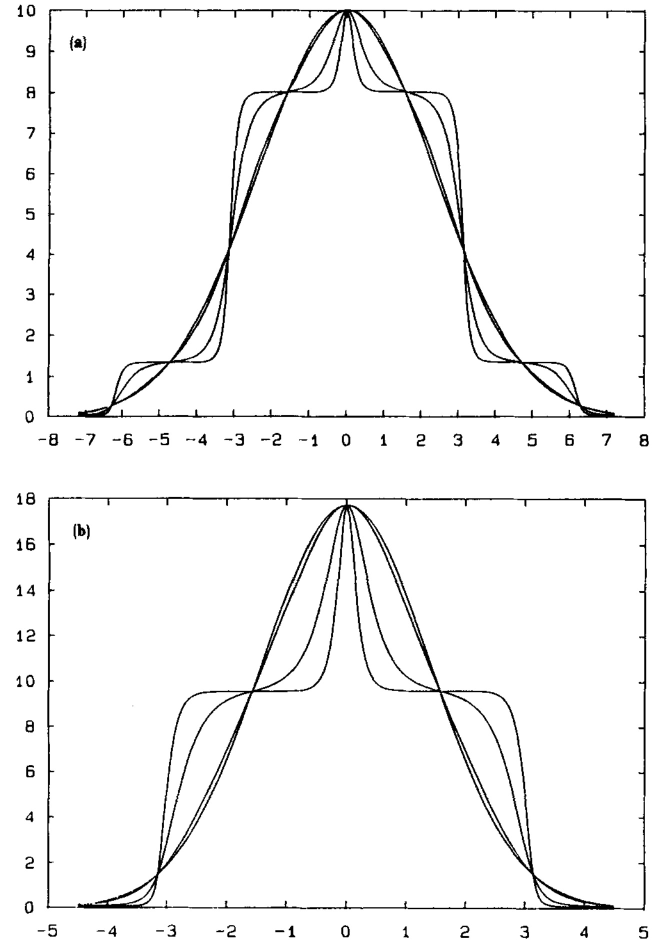


FIG. 1. (a) The function $f(\phi) = A \exp(-a^2\phi^2)$; $A = 10$, $a = 0.3$ vs $v = \pi u/2K$ for $m = 0$, $m = 0.5$, $m = 1\text{–}10^{-3}$, and $m = 1\text{–}10^{-9}$. These values of parameter m are used in all the figures. For $m = 0$ the curve is the usual Gaussian. Notice the intersection points at $n\pi/2$ ($n = 0, \pm 1, \pm 2, \ldots$). (b) Generalized Fourier transform for $f(\phi)$: $F(\alpha) = (A\pi^{1/2}/a)\exp(-\alpha^2/4a^2)$; $A = 10$, $a = 1$ (parameter a has been changed in order to clarify the plot).

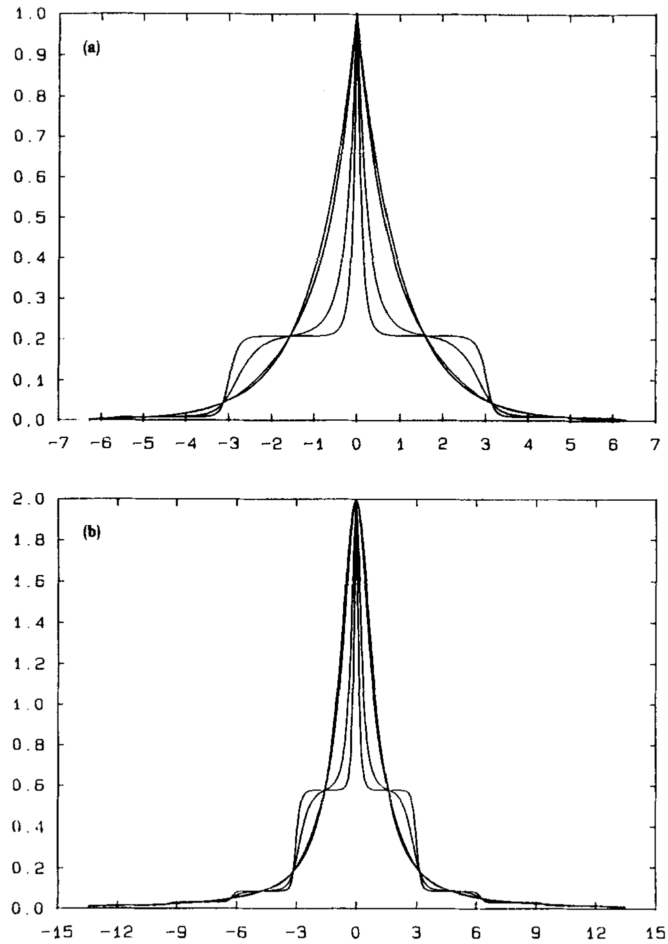


Fig. 2. (a) The function $f(\phi) = A \exp(-a|\phi|)$; $A = a = 1$. (b) Generalized Fourier transforms for $f(\phi)$: $F(\alpha) = (2A/a)[a^2/(a^2+\alpha^2)]$; $A = a = 1$. This function corresponds to the generalized Breit–Wigner curve.

alized Fourier transforms are shown in Fig. 1(b). Each curve of Fig. 1(a) has its corresponding generalized Fourier transform in Fig. 1(b). In order to clarify the plot we have changed the value to $a = 1$. One sees that the inflexion points are the same as in the case of the original functions (without transformation) and all the curves intersect there: this is the effect of reducing to the fundamental period.

To show that a change in parameter a does not alter the theory of generalized transformations and to describe another simple example, we have plotted the functions $f(\phi) = A \exp(-a|\phi|)$ and their corresponding generalized Fourier transforms $F(\alpha) = (2A/a)\,[a^2/(a^2+\alpha^2)]$ in Figs. 2(a) and 2(b), respectively. The last functions are the expression of the generalized Breit–Wigner curves (they are the usual Breit–Wigner or Lorentzian curve when $m = 0$). Notice in Fig. 2 the change of scale on both axes.

In the above examples the Fourier transform of the proposed functions has only a real part because of the symmetry—$f(\phi) = f(-\phi)$—of the functions (and thus of their transforms). We have therefore chosen another example, but now one that is nonsymmetric: its Fourier transform will be complex. This can be seen in Fig. 3, where an odd function and its generalized Fourier transform have been plotted. In the plot of $f(\phi)$ vs u, the values of the inflexion points are the same as before (for $m \neq 0$). Nevertheless, in the imaginary part of the transformed functions one sees that in addition to the inflexion points, there is another intersection between the curves for different values of the parameter m.

As a final example, Fig. 4 shows the generalized Fourier transform for the step function. It is an interesting case because the function $f(\phi)$ is defined piecemeal and is constant within each piece and therefore independent of m; nevertheless, its Fourier transform is dependent on the amplitude of u and there exists a representation for each value of m vs u.

A general comment about the examples mentioned (and many other cases) can be made: When $m \to 1$ all the defined functions tend to step-type functions with the "discontinuities" at the inflexion points; however, in spite of this, the functions shown in Figs. 1–4 are continuous and infinitely differentiable (excluding the case of points of discontinuity or of discontinuities in the first derivative when $m = 0$, i.e., the point zero in Figs. 2 and 3).

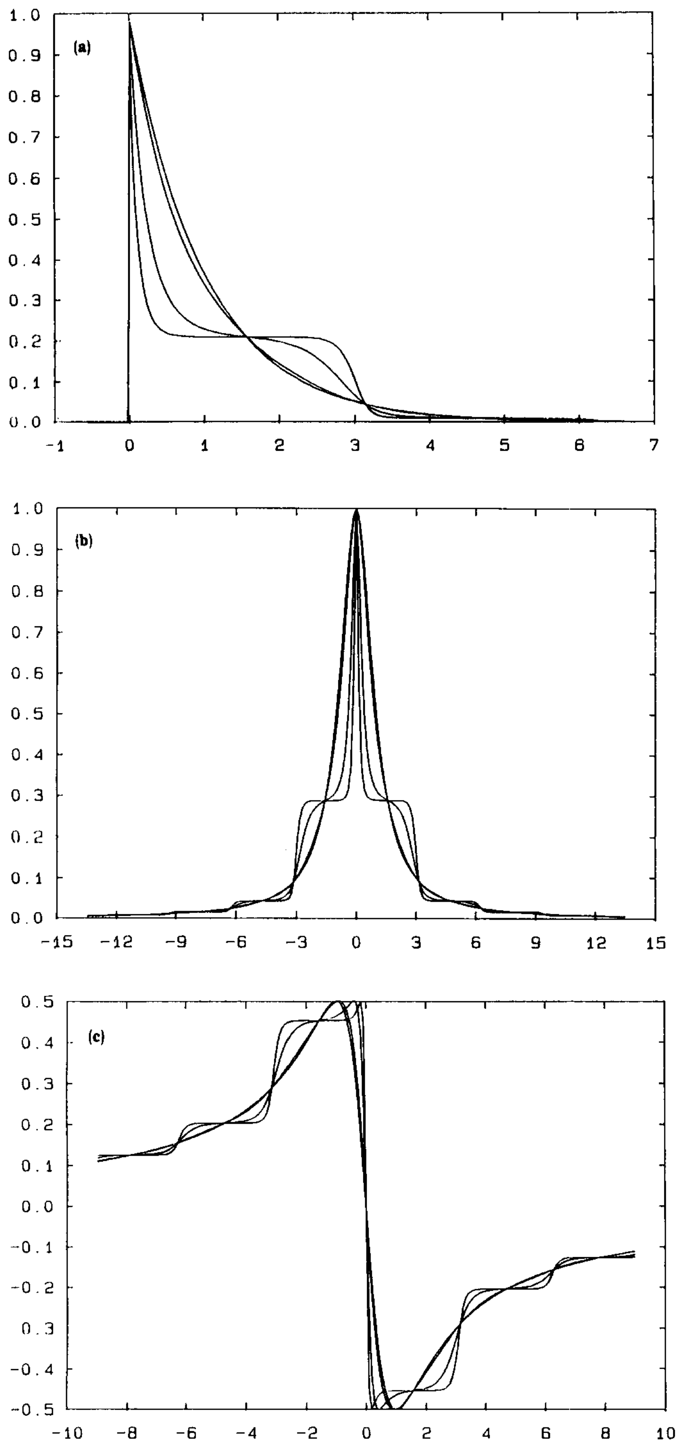


FIG. 3. (a) The function

$$f(\phi) = \begin{cases} A\exp(-a\phi), & \text{if } u>0, \\ 0, & \text{if } u<0, \end{cases} \qquad A=a=1.$$

(b) Real part of the generalized Fourier transform of $f(\phi)$: $F(\alpha) = A[(a-i\alpha)/(a^2+\alpha^2)]$; $A=a=1$. (c) Imaginary part of the generalized Fourier transform of $f(\phi)$.

VI. APPLICATIONS

A. Evaluation of integrals

One of the simplest applications is the evaluation of integrals. Here we follow Ref. 12 and consider as examples

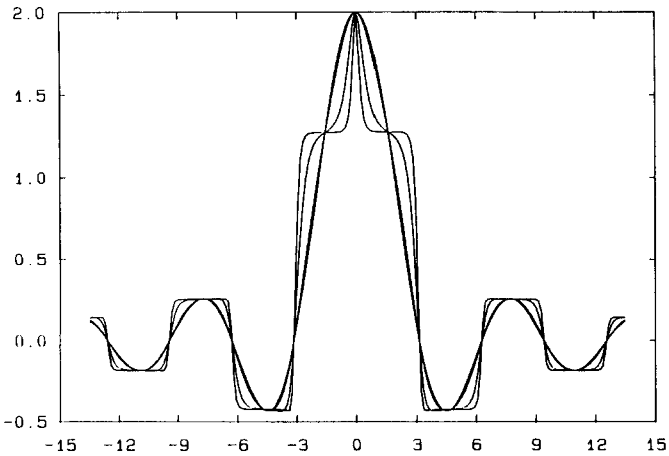


FIG. 4. The function $F(\alpha) = (2A/\alpha)\sin(L\alpha)$; $A = L = 1$ is the generalized Fourier transform of the step function

$$f(\phi) = \begin{cases} A\,|u|<L, \\ 0|u|>L. \end{cases}$$

$$I_1 = \int_0^\infty \exp(-b\phi)\cos(\alpha\phi)d\phi,$$

$$I_2 = \int_0^\infty \exp(-b\phi)\sin(\alpha\phi)d\phi.$$

Integrating I_1 by parts,

$$I_1 = [-\exp(-b\phi)\cos(\alpha\phi)/b]_0^\infty - \left(\frac{\alpha}{b}\right)\int_0^\infty \exp(-b\phi)\sin(\alpha\phi)d\phi,$$

which may be written as

$$I_1 = 1/b - (\alpha/b)I_2.$$

In a similar way,

$$I_2 = (\alpha/b)I_1.$$

Solving these equations for I_1 and I_2 we obtain

$$I_1 = b/(\alpha^2+b^2), \quad I_2 = \alpha/(\alpha^2+b^2),$$

which means that if we write $f(\phi) = \exp(-b\phi)$, then its cosine and sine transforms are, respectively,

$$F_c = \left(\frac{2}{\pi}\right)^{1/2}\left(\frac{b}{\alpha^2+b^2}\right), \quad F_s\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\alpha}{\alpha^2+b^2}\right).$$

Then

$$\int_0^\infty \left(\frac{\cos(\alpha\phi)}{\alpha^2+b^2}\right)d\alpha = \left(\frac{\pi}{2b}\right)\exp(-b\phi),$$

$$\int_0^\infty \left(\frac{\alpha\sin(\alpha\phi)}{\alpha^2+b^2}\right)d\alpha = \left(\frac{\pi}{2}\right)\exp(-b\phi).$$

B. The generalized Yukawa potential

We have seen that the properties of the Fourier transforms can be cast into a form that duplicates the properties and formulas of the usual Fourier series. In fact, we have constructed this and previous papers so that the formulas should be formally the same as the usual ones, where the change is simply in the meaning of the argument of the functions.

The solution of the time-independent Klein–Gordon equation for a time-independent field[15]

$$(\nabla^2 - \mu^2\omega^2)\psi(\mathbf{x}) = 4\pi g\rho(\mathbf{x}),$$

where g determines the strength and ρ is a probability density which vanishes at infinity, is

$$\psi(\mathbf{x}) = -g\int\left(\frac{\exp(-\mu|\mathbf{x}-\mathbf{x}'|)}{|\mathbf{x}-\mathbf{x}'|}\right)\rho(\mathbf{x}')d^3x'.$$

For a point source placed at position $\mathbf{x}' = \mathbf{0}$, this solution becomes the Yukawa potential

$$\psi(r) = -g(e^{-\omega\mu r}/r).$$

C. The generalized Breit–Wigner resonance formula

An important application is illustrated in Fig. 2(b). The generalized Fourier transform of the function $A\exp(-a|\phi|)$ (which in some sense can be considered as the Yukawa potential multiplied by the modulus of the variable) is the generalized Breit–Wigner resonance formula. This function appears in classical mechanics in the forced harmonic oscillator.

VII. CONCLUSIONS

We have presented generalized exponential and circular functions which have enabled us to develop generalized Fourier series and the corresponding integral transforms. These generalized transforms are given in terms of the Jacobi elliptic functions and in consequence are very well suited for problems with polynomial nonlinearities.

The subject of Jacobi elliptic functions is more than 150 years old.[16] Nevertheless, we believe our exponential series and transforms, developed in a form as similar as possible to the usual ones, are new. The usual textbook series and Fourier transforms are only changed by ϕ, $\sin\phi$, etc. being $\phi = \mathrm{am}(u,m)$ of the elliptic integrals of the first kind, where $m = k^2$ is the parameter of Jacobi functions. The usual pair of transforms are the special case of our transforms when $m = 0$.

From the many possible transforms we have selected the most usual and shown how they are useful for the generalized Yukawa and generalized Breit–Wigner potentials corresponding to simple generalizations of the Klein–Gordon equation. These formulas open the way for the development of simple models of nonlinear quantum mechanics in nonflat space-times, but they are also useful in other branches of nonlinear physics, as we have shown in previous papers.[6–8]

ACKNOWLEDGMENTS

JDB thanks the Experimental Physics Division at CERN for its hospitality.

The authors acknowledge the Comisión Asesora (CAICYT, Projects 1179-84 and PB87-007), Spain, for financial support.

[1] C. M. Rodríguez, Nuovo Cimento B **98**, 87 (1987).

[2] J. D. Bejarano and C. M. Rodríguez, J. Math. Phys. **28**, 2665 (1987).

[3] A. M. Sánchez and J. D. Bejarano, J. Phys. A: Math. Gen. **19**, 887 (1986).

[4] J. D. Bejarano, A. M. Sánchez, and C. M. Rodríguez, J. Chem. Phys. **85**, 5128 (1986).

[5] J. D. Bejarano and A. M. Sánchez, J. Math. Phys. **29**, 1847 (1988).

[6] J. D. Bejarano and A. M. Sánchez, to be published in J. Sound Vib.

[7] S. B. Yuste and J. D. Bejarano, J. Sound Vib. **110**, 347 (1986).

[8] J. García-Margallo and J. D. Bejarano, J. Sound Vib. **116**, 591 (1986).

[9] L. M. Milne-Thomson, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).

[10] J. García-Margallo, J. D. Bejarano, and S. B. Yuste, J. Sound Vib. **125**, 13 (1988).

[11] A. Sommerfeld, *Partielle Differential Gleichungen der Physik, Band VI* (Akademie Verlags Gesellschaft, Leipzig, 1954), Vorlesungen über Theoretische Physik.

[12] I. N. Sneddon, *Fourier Transforms* (McGraw-Hill, New York, 1951).

[13] J. D. Gaskill, *Linear Systems, Fourier Transforms and Optics* (Wiley, New York, 1978).

[14] E. O. Brigham, *The Fast Fourier Transform* (Prentice-Hall, Englewood Cliffs, NJ, 1974).

[15] H. Frauenfelder and E. M. Henley, *Subatomic Physics* (Prentice-Hall, Englewood Cliffs, NJ, 1974).

[16] K. G. Jacobi, *Fundamenta Nova Theoriae Functionum Ellipticarum* (Königsberg, 1829); see, also, K. G. Jacobi, Gesammelte Werke, Band 1. 1881, Band 2, 1882, Band 3, 1884.

Hamiltonian reduction for massive fields coupled to sources

C. Marzban,[a] B. F. Whiting, and H. Van Dam
Institute of Field Physics, Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27599-3255

(Received 10 February 1989; accepted for publication 22 March 1989)

The Dirac–Bergmann formalism for handling constrained dynamical systems is applied to massive and massless spin-2 field theories in Minkowski space. In the process, it is shown that an extension of the formalism is required before the true Hamiltonian on the reduced phase space can be calculated. With this extension, the approach then allows a revealing exposition of the $m \to 0$ limit of the massive theory, particularly of the requirement for the conservation of sources, and reconfirms the experimentally observable discontinuity which occurs at $m = 0$.

I. INTRODUCTION, SUMMARY, AND CONCLUSION

The question "Is Einstein's theory an isolated point in a class of theories parametrized by a continuous parameter, m, the mass of the graviton?" has been under study for some time. Historically, the first answer (which was in the affirmative) was given in Ref. 1.

In Ref. 1 the scattering of conserved sources via one graviton exchange (which describes the motion of the planets as well as the bending of light by the sun) was studied. It was found that as the assigned graviton mass m approached zero the results for the scattering disagreed with that for the $m = 0$ case. This discontinuity occurs because the sources of the gravitational field are the energy momentum tensors. These tensors are conserved up to first order in the coupling constant, but not all of them are traceless (as in the energy momentum tensor of the Maxwell field). This leads to an observable discontinuity in the scattering of light by a star (the $m \to 0$ result is $\frac{3}{4}$ of the angle for the $m = 0$ result). The assumptions of Ref. 1 were unitarity, positive energy and that the sources are conserved independently of m.

It is important to note that the discontinuity arises at the level of tree diagrams, i.e., it is a classical result and does not require considerations of the quantization of gravity. Thus, it is of a different nature from that of the discontinuity in the mass, which occurs at the one loop level in a naive theory of massive Yang Mills particles. The latter discontinuity was removed by the Higgs mechanism for introducing mass.[2]

A more extensive treatment of the question "Can gravity have a finite range?" was given by Boulware and Deser in a beautiful paper.[3] They arrived at the conclusion of Ref. 1 from a study of the field equations (and the action) displaying the limiting behavior of the various helicity modes of a massive spin two field. Also, using the ADM mechanism, they extended the results to the full nonlinear theory. Here the discontinuity becomes even more severe; no suitable massive theory appears to exist. This led to their paradoxical conclusion that the gravitational force has an infinite range, even in a finite universe. A recent result by Higuchi[4] points in the same direction. Following earlier work[5] on anti-de Sitter space he investigated massive yields in de Sitter space and found that the square of the mass of a massive spin-2 field cannot lie between zero and $\frac{2}{3} \times$ (cosmological constant). The value zero is allowed for the mass as are mass values above this mass gap.

It has been argued[6] that the discontinuity may be removed by considering sources which are not conserved for $m \neq 0$, but which are conserved in the limit $m \to 0$. This leads to a number of conditions on the limit of the sources as $m \to 0$. These conditions are so severe that no realistic examples can be constructed if one assumes the sources to be local functionals of the matter field variables.[7]

It is amusing that the discontinuity persists for the spin-$\frac{3}{2}$ field, the supersymmetric partner of the graviton.[8]

The motivation for the present work is as follows: since a massive spin-2 field is governed by a constrained theory, the formalism of Dirac–Bergmann,[9] as adopted here, offers an excellent environment for its study. In addition, the structure of the Hamiltonian constraints and the Dirac brackets provides a revealing exposition of the $m \to 0$ limit. Of course, the conclusions are often the same as those already mentioned above, but our approach is sufficiently different to warrant this independent treatment. For instance, in reducing the Hamiltonian we find subtleties that call for amendments to the existing algorithm for reduction. Although the $m \to 0$ discussion is not contingent on this finding, the very construction of a reduced Hamiltonian for a massive spin-2 field is. Furthermore, in contrast to Refs. 1 and 3, we do not assume *a priori* conservation of sources, but rather derive the necessary and sufficient conditions (with emphasis on sufficient) that the sources must satisfy. We also apply the reduction formalism partially to a theory with a different mass term, rederiving some previously well-known results. An outline and the conclusions of this paper follow.

We begin by considering the spin-1 case. This case has been well-studied[10] and it can even be found in certain books! [11] There are two reasons why we nevertheless take the opportunity of presenting it here. First, to introduce our notation and its use in the (by now commonly known) formalism of Dirac. Second, and of more physical significance, to explain our hesitation in the *a priori* use of conservation for the sources. We will derive the field equations as implied by the reduced Hamiltonian (and Lagrangian), and the corresponding Dirac brackets (DB) for the reduced phase space (RPS) variables. In the end, we will be interested in an S matrix, but its inherent "symmetry" (e.g., invariance under field redefinition) renders the $m \to 0$ interpretation of the

[a] Present address: International Centre for Theoretical Physics, P.O. Box 586, I-34100, Trieste, Italy.

 0022-2488/89/081877-16$02.50

field equations slightly ambiguous. The transverse degrees of freedom (helicity ± 1) of the massive spin-1 field smoothly go over to a gauge invariant transverse photon. But the behavior of the longitudinal part—the scalar—is more subtle to probe. We display two scenarios: one in which the $m \to 0$ limit renders the scalar degree of freedom nondynamical (i.e., nonpropagating) and its equation of motion *implies* that the sources have to be conserved. A second scenario (obtained from the first, by field redefinition in the Hamiltonian) still implies that the currents must be conserved; but then the scalar piece decouples from the sources and smoothly becomes a freely propagating massless field as $m \to 0$. Subsequently, we present the reduced *Lagrangian* for the spin-1 case, for here, not only do we arrive at the same conclusions as above, but also the validity of the $m \to 0$ limit becomes clear since one can set $m \equiv 0$ explicitly!

Turning, next, to the spin-2 case, we consider two classes of massive theories, i.e., $\alpha = 1$ and $\alpha \neq 1$ in (24) below. The first of these, corresponding to a pure spin-2 field, has ten second class constraints, from which we construct the DB's (in Appendix B). The constraints and the DB's are used to determine a RPS for which we construct the reduced Hamiltonian which is sufficient for deducing the field equations for the dynamical degrees of freedom of the reduced theory. The necessity of source conservation in the massless case is discussed, at this level. Diverting our attention slightly, but necessarily, a completely general theorem is proved—that a reduced Hamiltonian is *not* necessarily the Hamiltonian on a given RPS. In the proof we also show how to *construct* the correct reduced Hamiltonian. This is followed by two examples that are then used in constructing the true Hamiltonian for the reduced massive spin-2 case. Finally, we derive the equation of motion for the RPS and consider the limit $m \to 0$.

We conclude that, in the limit $m \to 0$, the TT degrees of freedom (helicity ± 2) simply become the coordinate-gauge invariant degrees of freedom of the massless spin-2 field. The TL components (helicity ± 1) behave similarly to the longitudinal component of the massive spin-1 (including the ambiguity due to field redefinition) and are not determined by the field equations in the massless limit. There is one important difference, however. Whereas in the spin-1 case the full conservation of currents can be implied by the massless limit of the equation of motion for the helicity-0 state, in the spin-2 case, the field equations for TL (two degrees of freedom) can at most imply (59), the massless limit of which gives the expected *number* of conditions, which are essentially equivalent to two conservation conditions. The remaining degree of freedom (helicity-0) can be made to imply at most Eq. (62). The condition that a massless limit exists is now not, by itself, a statement of conservation. Furthermore, even though it is unphysical, there is a choice of the helicity-0 field which then remains determined by equations of motion in the limit; *and* it is coupled to the four-trace of the stress energy tensor Φ^{μ}_{μ}, unlike the rescaled longitudinal mode in spin-1, which became a free field. So, even if the sources are entirely conserved, the coupling of this scalar to the trace of the sources produces an extra attractive force that is the reason for the discontinuity. In addition, the complete understanding of conservation requires further investigation.

As in spin-1, we also present the reduced *Lagrangian*, for here the soundness of the $m \to 0$ limit is clearly justified provided the time component of the source conservation equations is satisfied. Additionally, by an alternative choice of the scalar degree of freedom we get a very explicit demonstration of the remaining conservation conditions being consequences of field equations. As in spin-1, *all* the unphysical degrees of freedom now become undetermined in the massless limit.

Finally, we consider an alternative mass term and we illustrate that although the discontinuity disappears it does not give rise to a physically acceptable theory (this is a well-known fact,[3,12] which we do not pursue further).

We will use a flat $(-+++)$ metric. Greek indices are four and Latin indices are three indices.

II. SPIN-1

In this section we apply the Dirac–Bergmann algorithm to the spin-1 field, mainly as preparation for the spin-2 field, and also for establishing the formalism and our notation. The object is to isolate the dynamical degrees of freedom (DDF), and to study the dynamics on the reduced phase space (RPS) in the limit as $m \to 0$. Necessary and sufficient conditions on the currents, in the limit $m \to 0$, will be found.

A. The $m=0$ case

The massless spin-1 field has been under scrutiny for almost a century, and over the years a variety of Lagrangians has been proposed. Let us consider the most orthodox Lagrangian for the spin-1 field A_μ coupled to external currents j^μ, i.e.,

$$L = \int d^3x\, \mathscr{L} = \int d^3x \left[-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \lambda j^\mu A_\mu \right]. \quad (1)$$

The canonical conjugate to A_μ is

$$\pi^\mu = \frac{\delta L}{\delta(\partial_0 A_\mu)} = -F^{0\mu}. \quad (2)$$

The fundamental equal-time Poisson bracket is

$$\{A_\mu(x), \pi^\nu(y)\} = \delta_\mu{}^\nu \delta^{(3)}(x-y).$$

The primary constraint is

$$\phi^0 \equiv \pi^0 \approx 0,$$

and the canonical Hamiltonian density is defined as

$$\begin{aligned} \mathscr{H}_c &= \pi^\mu \partial_0 A_\mu - \mathscr{L}, \\ &= \tfrac{1}{2}\pi_i \pi^i + \pi^i \partial_i A_0 + \tfrac{1}{4} F_{ij} F^{ij} - \lambda j^\mu A_\mu. \end{aligned} \quad (3)$$

Now,

$$H_p \equiv \int d^3x\, \mathscr{H}_p = \int d^3x [\mathscr{H}_c + v(x)\phi^0],$$

is the primary Hamiltonian (with v, the Lagrange multiplier), i.e., the time evolution operator. To get the secondary constraints we demand the conservation of the primary constraint,

$$0 \approx \partial_0 \phi^0 = \{\pi^0, H_p\} = \partial_i \pi^i + \lambda j^0 \equiv \phi^1.$$

Conservation of ϕ^1 leads to the statement of the conservation of sources,

$$0 = \partial_0 \phi^1 = \{\partial_i \pi^i, H_p\} + \lambda \partial_0 j^0 = \lambda \partial_\mu j^\mu. \tag{4}$$

Therefore, in the massless case, the currents must be conserved for consistency. By the use of the fundamental bracket one can see that the constraints

$$\phi^0 = \pi^0,$$
$$\phi^1 = \partial_i \pi^i + \lambda j^0,$$

are first class, i.e.,

$$\{\phi^0(x), \phi^1(y)\} = 0,$$

and that there are no more constraints.

We can anticipate the dimension of the RPS simply by anticipating the effects of gauge fixing. In the Dirac–Bergmann formalism, allowed gauge conditions are those that turn each first-class constraint into a second class one (i.e., those whose Poisson bracket is nonzero). And so, for each first-class constraint one must introduce precisely one gauge-fixing condition (for the second-class case, see below). Therefore,

$$\begin{aligned}&[\text{dim. of reduced phase space at each point}]\\ &\quad = [\text{dim. of full phase space}]\\ &\qquad - 2[\text{number of first-class constraints}]\\ &\quad = 8 - 2(2)\\ &\quad = 4 \equiv 2 + 2.\end{aligned}$$

Thus we expect that only two degrees of freedom are physical. And, of course, two is the dimension of the fundamental irreducible representation (irrep) of the little group $SO(D-2) = SO(2)$.

To isolate the DDF's in a first-class theory it is *not* actually necessary to fix the gauge (contrary to common belief); since the DDF's are the gauge invariant components of the field, and since the first-class constraints are precisely the generators of gauge transformations, the DDF's are simply those components which Poisson commute with all the first-class constraints. Decomposing A_i and π_i into transverse and longitudinal parts (see Appendix A), with

$$A_i = A_i^T + A_i^L \quad \text{and} \quad \pi^i = \pi^{Ti} + \pi^{Li},$$

one can see that only A_i^T and π^{Ti} satisfy the conditions for being DDF's, i.e.,

$$\{A_i^T, \phi^0\} = \{A_i^T, \phi^1\} = \{\pi^{Ti}, \phi^0\} = \{\pi^{Ti}, \phi^1\} = 0.$$

Being a transverse three-vector, A_i^T has only two independent degrees of freedom, corresponding to the two helicity states of the photon. Here A_i^T and π^{Ti} are a canonical pair since

$$\{A_i^T(x), \pi^{Tj}(y)\} = T_i^j \delta^{(3)}(x-y),$$

where in this, and in all similar expressions, operators act only on the first argument of the delta functions. The operator T_i^j is discussed in Appendix A.

In principle, the *form* of the reduced Hamiltonian for the DDF's would depend on the choice of the gauge. However, all that we shall subsequently need is the second-order equation of motion for A_i^T, and of course that is

$$\Box A_i^T = -\lambda j_i^T, \tag{5}$$

where $j_i^T \equiv T_{i}^{k} j_k$.

B. The $m \neq 0$ case

Having established the notation for the massless spin-1 case we go on to consider the massive spin-1 theory (again, given by the most orthodox Lagrangian), and the limiting theory as $m \to 0$. The Lagrangian density is

$$\mathscr{L}^m = \mathscr{L} - \tfrac{1}{2} m^2 A_\mu A^\mu, \tag{6}$$

where $\mathscr{L}$ is given in (1).

The primary constraint is unchanged, i.e.,

$$\chi^0 \equiv \pi^0 \approx 0.$$

The canonical Hamiltonian density becomes

$$\mathscr{H}_c^m = \mathscr{H}_c + \tfrac{1}{2} m^2 A_\mu A^\mu,$$

with $\mathscr{H}_c$ given in (3).

The time evolution of χ^0, as dictated by the primary Hamiltonian

$$H_p^m \equiv \int d^3x \, \mathscr{H}_p^m = \int d^3x [\mathscr{H}_c^m + v(x)\chi^0],$$

gives the secondary constraint

$$0 \approx \partial_0 \chi^0 = \{\pi^0(x), H_p^m\} = \partial_i \pi^i - m^2 A^0 + \lambda j^0 \equiv \chi^1.$$

This is the only secondary constraint, since the time evolution of χ^1 gives a multiplier condition,

$$\begin{aligned}0 \approx \partial_0 \chi^1 &= \{(\partial_i \pi^i - m^2 A^0), H_p^m\} + \lambda \partial_0 j^0\\ &= m^2 v(x) - m^2 \partial_i A^i + \lambda \partial_\mu j^\mu,\end{aligned}$$

telling us what the function $v(x)$ is. Recall that in the massless case there was no such condition, which was only a reflection of the singular nature (first class) of the theory. The function $v(x)$ is the velocity for which we could not solve in the massless case;

$$\partial_0 A^0 = \{A^0(x), H_p^m\} = v(x).$$

There is no point in taking the time evolution of the multiplier condition, for that will not give any additional constraints, only the time evolution of $v(x)$. Note that,

$$\lim_{m \to 0} \chi^a = \phi^a$$

and

$$\lim_{m \to 0} (\text{multiplier condition})$$
$$\Rightarrow (\text{conservation of source}).$$

The significance of this will be discussed below.

As might be expected from the lack of gauge invariance of L^m, all of the constraints

$$\chi^0 \equiv \pi^0,$$
$$\chi^1 \equiv \partial_i \pi^i - m^2 A^0 + \lambda j^0,$$

are second class. And since there is no gauge symmetry to be fixed, we may anticipate the number of DDF's from $8 - 2 = 6 = 3 + 3$; i.e., three independent degrees of freedom. Again, three is, correctly, the dimension of the fundamental irrep of the little group $SO(D-1) = SO(3)$.

Dirac has indicated a method for identifying the RPS. In addition, there are various other ways of isolating the DDF's: however, they all involve calculating the matrix (which to our knowledge has no special name as yet):

$$\Delta(x-y) = \begin{vmatrix} \{\chi^0,\chi^0\} & \{\chi^0,\chi^1\} \\ \{\chi^1,\chi^0\} & \{\chi^1,\chi^1\} \end{vmatrix} = m^2 \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix} \delta^{(3)}(x-y).$$

Note that χ^0 and χ^1 are second class, hence

$$\Delta^{-1}(x-y) = \frac{1}{m^2} \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} \delta^{(3)}(x-y).$$

In Dirac's approach one constructs a new bracket—the Dirac bracket (DB)—based on the requirement that it be the Poisson bracket on the RPS. For any two phase-space functions $A(x)$ and $B(y)$, the DB is defined as

$$\{A(x),B(y)\}_D = \{A(x),B(y)\} - \int du\, dv \{A(x),\chi^a(u)\} \times \Delta_{ab}^{-1}(u-v)\{\chi^b(v),B(y)\}.$$

Apart from the usual properties of Poisson brackets, DB's are constructed also to obey the strong equation

$$\{\text{constraint, any phase-space function}\}_D = 0. \tag{7}$$

Thus, in the search for DDF's all we are looking for is a pair of variables, q and p, which are "conjugate" in the following schematic sense:

$$\{q,p\}_D = 1. \tag{8}$$

For our problem, the only nonzero DB's are calculated to be simply

$$\{A_i(x),\pi^j(y)\}_D = \delta_j^i \delta^{(3)}(x-y), \tag{9a}$$

$$\{A_0(x),A_i(y)\}_D = (1/m^2)\partial_i \delta^{(3)}(x-y). \tag{9b}$$

The fact that $\{A_i(x),\pi^i(y)\}_D = 3\delta(x-y)$ already indicates that the A_i's are three physically independent degrees of freedom.

C. Reduced phase space (RPS)

So we note that, in fact, one can take as the DDF's

$$q_i \equiv A_i(x), \quad \text{and} \quad p^j \equiv \pi^j(x),$$

since,

$$\{q_i(x),p^j(y)\}_D = \delta_i^j \delta^{(3)}(x-y),$$

thereby satisfying (8).

It is possible to describe the dynamics on the reduced phase space (RPS) by a Hamiltonian on the RPS. In most cases which have been discussed in the literature, the Hamiltonian is in fact equal to the reduced Hamiltonian (i.e., one which is obtained from the full Hamiltonian by imposing the constraints strongly). This is also the case in the spin-1 theory, but as we shall discover it is *not* true in the spin-2 theory. See Sec. II B 3.

In reducing the full primary Hamiltonian we may certainly impose the constraints as strong equalities, for time evolution on the RPS will be given by DB's and they obey (7). Hence, for the massive spin-1 case, the reduced Hamiltonian in the RPS variables q, p is

$$H|(q,p) = H_p^m(q,p;\chi=0).$$

Since we have established that A^0 cannot be a DDF we may use the constraint $\chi^0 = 0$, to remove π^0, and $\chi^1 = 0$ to remove A^0 from the full Hamiltonian, i.e.,

$$A^0(x) = (1/m^2)(\partial_i \pi^i + \lambda j^0).$$

The result is

$$H| \equiv H(q_i \equiv A_i, p^i \equiv \pi^i) = \int d^3x \Big[\frac{1}{2}\pi_i\pi^i + \frac{1}{2m^2}(\partial_i\pi^i)^2 + \frac{1}{2}\partial_i A_j \partial^i A^j - \frac{1}{2}(\partial_i A^i)^2 + \frac{1}{2} m^2 A_i A^i + \frac{\lambda}{m^2}(\partial_i \pi^i) j^0 - \lambda A_i j^i \Big] + O(\lambda^2), \tag{10}$$

where we have performed a few partial integrations, and have justified the vanishing of the surface terms. Note that the reduced Hamiltonian appears singular in the massless limit. The Hamiltonian equations of motion on the RPS using (9) and (10) are

$$\partial_0 q_i \equiv \partial_0 A_i(x) = \{A_i(x),H\}_D = \pi_i - (1/m^2)\partial_i\partial_j\pi^j - (\lambda/m^2)\partial_i j^0,$$

$$\partial_0 p^i \equiv \partial_0 \pi^i(x) = \{\pi^i(x),H\}_D = -[(m^2-\nabla^2)A^i + \partial^i\partial_j A^j - \lambda j^i].$$

D. Reduced configuration space and the $m \to 0$ limit

Before we go on to study the $m\to 0$ limit (which at this stage seems formidable), let us first decompose the fields into their transverse and longitudinal "helicity" components. (See Appendix A.)

The equations of motion become

$$\partial_0 A_i^T = \pi_i^T, \tag{11a}$$

$$\partial_0 A_i^L = [1-\nabla^2/m^2]\pi_i^L - (\lambda/m^2)\partial_i j^0, \tag{11b}$$

$$\partial_0 \pi_i^T = -[m^2-\nabla^2]A_i^T + \lambda j_i^T, \tag{11c}$$

$$\partial_0 \pi_i^L = -m^2 A_i^L + \lambda j_i^L. \tag{11d}$$

Now, without appealing to field redefinition, the RPS is clearly ill-behaved as far as the limit $m\to 0$ is concerned. However, one can say more than this by looking at the $m\to 0$ limit of the Lagrangian equations of motion in the reduced configuration space. The dynamics there would be given by a reduced Lagrangian derived from the reduced Hamiltonian via a Legendre transformation. But that requires solving (11a) and (11b) for the π's, i.e.,

$$\pi_i^T = \partial_0 A_i^T, \tag{12}$$

$$\pi_i^L = [1/(m^2-\nabla^2)][m^2\partial_0 A_i^L + \lambda\partial_i j^0]. \tag{13}$$

Taking the time derivative of (12) and (13) and substituting in (11c) and $(\nabla^2 - m^2)$ times (11d), respectively, gives

$$(\Box - m^2)A_i^T = -\lambda j_i^T, \tag{14a}$$

$$m^2(\Box - m^2)A_i^L = -m^2\lambda j_i^L + \lambda\partial_i\partial_\mu j^\mu. \tag{14b}$$

So with A_i^T representing two degrees, and A [see (A.2)] representing the third degree of freedom of a massive spin-1 field, the second-order equations of motion become

$$(\Box - m^2)A_i^T + \lambda j_i^T = 0, \tag{15a}$$

$$m^2[(\Box - m^2)A + \lambda j] = \lambda\partial_\mu j^\mu, \tag{15b}$$

where $j \equiv (1/\nabla^2)\partial_k j^k$. Clearly the equations of motion following from a direct computation with the reduced action will be the same as (15) (see Sec. II E) which were derived by combining the corresponding Hamiltonian equations. Note the covariant expression on the right-hand side of (15b).

It is now clear what happens to the two degrees of freedom A_i^T, in the limit as $m \to 0$; they simply go over to the two degrees of freedom of a transverse photon. See Eq. (5).

The interpretation of (15b) in the limit $m \to 0$ is perhaps more subtle. In the field theory interpretation, the classical equations of motion do not directly represent "physical" quantities. Rather, it is the S matrix which is considered to be physical, and the S matrix is insensitive to field redefinitions (such as rescaling of the fields).

If we were to simply let $m \to 0$ in (15b), not only the source term j would drop out but also, the field A would become nondynamical in the absence of a kinetic term for it. Thus we see the well-known result that the massless field equations do not determine the longitudinal field, A_i^L. What remains in (15b) is the statement of conservation of sources, $\partial_\mu j^\mu = 0$, seen here as a consequence of the longitudinal field equation.

In order to see more directly what might be the contributions to the S matrix, we could alternatively consider working with rescaled fields

$$\widetilde{A} \equiv (m/\mu)A \quad \text{and} \quad \widetilde{\pi} \equiv (\mu/m)\pi,$$

which give rise to a well-normalized propagator (with μ some independent mass scale). The resulting field equations derived from the Hamiltonian for the rescaled fields is

$$\mu(\Box - m^2)\widetilde{A} + \lambda m j = (\lambda/m)\partial_\mu j^\mu. \tag{16}$$

The S matrix is unchanged by such a field redefinition, but some of the implications of the longitudinal equations of motion are more explicit. In the limit $m \to 0$, $\widetilde{A}$ becomes a freely propagating field, decoupled from the sources, which *must* be conserved in order that the limit exists. Thus although the conservation condition may be apparent in different ways (depending upon the redefinition of fields), nevertheless the contribution to the S matrix of the longitudinal degree of freedom of a massive spin-1 field coupled to sources which are necessarily conserved in the limit $m \to 0$ is zero in the same limit.

E. Lagrangian formulation

Here we will reduce the Lagrangian and not the action so as to avoid any possible addition of a total derivative which would lead to a canonical change of variables. We will briefly reinvestigate the conclusions of Sec. II C from the Lagrangian. Start from the full Lagrangian

$$L(A_\mu, \dot{A}_\mu) \equiv \int d^3x\, \mathscr{L}^m,$$

where $\mathscr{L}^m$ is given by (6).

We have shown that A_0 is an unphysical degree of freedom. We therefore remove it from the full Lagrangian, using the χ_1 constraint, in order to construct the reduced Lagrangian

$$L| \equiv L(A_i, \dot{A}_i) = L(A_i, \dot{A}_i; \text{constraint} = 0).$$

The constraint, $\chi^1 = 0$, from Sec. II B is

$$\partial_i \pi^i - m^2 A^0 + \lambda j^0 = 0,$$

where π^i is given by (2)

$$\pi^i = -\partial^0 A^i + \partial^i A^0.$$

Thus we can express A^0 in terms of the physical variables A_i, by

$$A^0 = [1/(\nabla^2 - m^2)][\partial_i \dot{A}^i - \lambda j^0]. \tag{17}$$

The result is

$$\begin{aligned} L| = \int d^3x\Big[& -\frac{1}{2}\partial_\mu A_i \partial^\mu A^i + m^2 A_i A^i - \frac{1}{2}(\partial_\mu \partial_i A^i) \\ & \times \frac{1}{\nabla^2 - m^2}(\partial^\mu \partial^j A_j)\Big] - m^2(\partial_i A^i) \\ & \times \frac{1}{\nabla^2 - m^2}(\partial_j A^j) + \lambda(\partial_i \dot{A}^i) \\ & \times \frac{1}{\nabla^2 - m^2} j^0 + \lambda A_i j^i - \lambda^2 j_0 \frac{1}{\nabla^2 - m^2} j^0\Big], \end{aligned}$$

where we have performed several partial integrations, and dropped a surface integral. In terms of the transverse and longitudinal variables we can write

$$\begin{aligned} L| &\equiv L^T|(A^T, \dot{A}^T) + L^L|(A^L, \dot{A}^L) \\ &\equiv L^T| + L^L| + \text{"constants"} \end{aligned} \tag{18}$$

with

$$L^T| = \int d^3x\Big[-\frac{1}{2}(\partial_\mu A_i^T \partial^\mu A^{Ti} + m^2 A_i^T A^{Ti}) + \lambda A_{i}^{T} j^{Ti}\Big],$$

$$\begin{aligned} L^L| = \int d^3x\Big[& -\frac{m^2}{2}\Big(\partial_\mu A \frac{\nabla^2}{\nabla^2 - m^2} \\ & \times \partial^\mu A + m^2 A \frac{\nabla^2}{\nabla^2 - m^2} A\Big) \\ & + \lambda \partial_0 A \frac{m^2}{\nabla^2 - m^2} j^0 + \lambda \partial_\mu A j^\mu\Big] \end{aligned}$$

and the final "constants" representing boundary integrals and terms quadratic in the sources.

Field redefinition could be performed as in the Hamiltonian formulation but it will make no difference to physical quantities. We see that the S matrix will become independent of conserved sources in the massless limit. Note that in the derivation of $L|$, nowhere did we assume $m \neq 0$; so we can, in fact, set $m \equiv 0$ to recover completely the massless theory. In that case only the final term remains in $L|$ and source conservation is again a consequence of the longitudinal field equation.

III. SPIN-2

Here we shall follow the same steps as in the spin-1 case. We will find that the spin-2 problem is somewhat different and even more subtle than the spin-1 case.

A. The $m=0$ case

Again we will start from the most orthodox Lagrangian. However, we will have to dabble with it to put it in a form most suitable for our objective;

$$L' = \int d^3x\,\mathscr{L}'$$
$$= -\int d^3x\Big[\frac{1}{2}\partial_\lambda h_{\mu\nu}\partial^\lambda h^{\mu\nu} - \frac{1}{2}\partial_\lambda h^\mu_\mu \partial^\lambda h^\nu_\nu$$
$$+ \partial_\mu h^{\mu\nu}\partial_\nu h^\lambda_\lambda - \partial^\mu h_{\mu\lambda}\partial_\nu h^{\nu\lambda} - \kappa h_{\mu\nu}\Phi^{\mu\nu}\Big],$$

where $\Phi^{\mu\nu}$ is the stress-energy tensor for the external, non-dynamic, sources to which the spin-2 field $h_{\mu\nu}$ is coupled. The canonical conjugate to $h_{\mu\nu}$ according to L' is

$$\pi'^{\mu\nu} = \frac{\delta L'}{\delta(\partial_0 h_{\mu\nu})}.$$

Thus the various components of $\pi'^{\mu\nu}$ are the appropriate coefficients in a generic variation of L'. In this way we obtain

$$\pi'^{00} = -\partial_i h^{i0}, \quad \pi'^{i0} = \partial^i h^\mu_\mu - 2\partial_j h^{ij},$$
$$\pi'^{ij} = \partial_0 h^{ij} - \eta^{ij}\partial_0 h^k_k + \eta^{ij}\partial^k h_{0k}.$$

The expressions for π'^{00} and π'^{i0} are constraints, since they do not contain time derivatives. At this point we perform a canonical transformation $\pi' \to \pi$, $h' \to h$, by adding a total time derivative to $\mathscr{L}'$, to obtain

$$L = \int d^3x\,\mathscr{L} = \int d^3x\{\mathscr{L}'$$
$$+ \partial_0[h^{i0}(\partial_i h^\mu_\mu - 2\partial^j h_{ij})]\}. \qquad (19)$$

Upon making certain that the various spatial surface terms that appear in δL vanish, one has the new variables

$$\pi^{00} = 0, \quad \pi^{i0} = 0,$$

and

$$\pi^{ij} = \partial_0 h^{ij} - \eta^{ij}\partial_0 h^k_k + 2\eta^{ij}\partial^k h_{0k}$$
$$+ \partial^i h^{0j} + \partial^j h^{0i}. \qquad (20)$$

The expressions for $\pi^{0\mu}$ are the primary constraints (PC) and so we will use Dirac's weak equality

$$\psi^0 \equiv \pi^{00} \approx 0,$$
$$\psi^i \equiv \pi^{0i} \approx 0.$$

This is the form in which we would like the PC's to be. We shall therefore use L, and not L' as the Lagrangian describing the spin-2 field. Clearly, the equations of motion for $h_{\mu\nu}$ are unchanged, since the action is changed only by a "time" surface term.

With these variables the canonical Hamiltonian density is,

$$\mathscr{H}_c = \pi^{\mu\nu}\partial_0 h_{\mu\nu} - \mathscr{L}$$
$$= \tfrac{1}{2}\pi\cdot\pi - \tfrac{1}{4}(\pi^i_i)^2 + 2\pi^{ij}\partial_0 h_{ij} - \partial_i h^0_0 \partial^i h^j_j$$
$$+ \tfrac{1}{2}\partial_i h_{jk}\partial^i h^{jk} - \tfrac{1}{2}\partial_i h^j_j \partial^i h^k_k$$
$$- (\partial h)\cdot(\partial h) + \partial_i h^{ij}\partial_j h^\mu_\mu - \kappa h_{\mu\nu}\Phi^{\mu\nu}, \qquad (21)$$

where we have introduced a notation which we will use throughout, viz,

$$\pi\cdot\pi = \pi_{ij}\pi^{ij}$$

and

$$(\partial h)\cdot(\partial h) = (\partial_i h^{ij})(\partial^k h_{kj}).$$

The primary Hamiltonian is

$$H_p = \int d^3x\,\mathscr{H}_p = \int d^3x[\mathscr{H}_c + v_0(x)\psi^0(x)$$
$$+ v_i(x)\psi^i(x)].$$

With the multipliers $v_0(x)$ and $v_i(x)$, this is the time-evolution operator.

With some care one can take

$$\{h_{\mu\nu}(x), \pi^{\sigma\tau}(y)\} = \delta^\sigma_{(\mu}\delta^\tau_{\nu)}\delta^{(3)}(x-y)$$

as the fundamental Poisson bracket. The secondary constraints are

$$0 \approx \partial_0\psi^0 = \{\pi^{00}, H_p\} = \nabla^2 h^i_i - \partial_{ij}h^{ij} + \kappa\Phi^{00} \equiv \lambda^0,$$
$$0 \approx \partial_0\psi^i \equiv \{\pi^{0i}, H_p\} = \partial_j\pi^{ij} + \kappa\Phi^{0i} \equiv \lambda^i.$$

There are no more constraints on the fields, for the conservation in time of the secondary constraints simply gives the expression for conservation of the sources,

$$0 \approx \partial_0\lambda^0 = \{(\nabla^2 h^i_i - \partial_i\partial_j h^{ij}), H_p\}$$
$$+ \kappa\partial_0\Phi^{00} = \kappa\partial_\mu\Phi^{\mu 0}, \qquad (22a)$$
$$0 \approx \partial_0\lambda^i = \{\partial_j\pi^{ij}, H_p\} + \kappa\partial_0\Phi^{0i} = \kappa\partial_\mu\Phi^{\mu i}, \qquad (22b)$$

where in the first equation we have used $\lambda^i \approx 0$. Note that conservation of sources is a necessity in the $m=0$ case, just as in the spin-1 case. Due to the gauge invariance of the theory [reflected by the arbitrariness of the functions, $v_0(x)$ and $v_i(x)$], all eight constraints are first class in nature,

$$\{\psi^\mu, \psi^\nu\} = \{\psi^\mu, \lambda^\nu\} = \{\lambda^\mu, \lambda^\nu\} = 0.$$

The dimension of the reduced phase space (RPS), according to the discussion in Sec. II A, is $[20 - 2(8)] = 4 = 2 + 2$, corresponding to two physical h fields, and two π fields, consistent with the dimension of the irreducible part (i.e., traceless) of the $\square\!\square$ representation of the little group $SO(D-2) = SO(2)$.

As discussed in Sec. II A, we can isolate these DDF's without having to fix the gauge explicitly; they are the components of $h_{\mu\nu}$ and $\pi^{\mu\nu}$ which Poisson commute with all of the constraints. As is well known, h^{TT} and π^{TT} (see Appendix A) are the DDF's since

$$\{h^{TT}_{ij}, \psi^\mu\} = \{h^{TT}_{ij}, \lambda^\mu\} = \{\pi^{TT}_{ij}, \psi^\mu\} = \{\pi^{TT}_{ij}, \lambda^\mu\} = 0.$$

Of course, h^{TT} is identified with the helicity ± 2 states of the spin-2 field. Since the TT operator is the unit operator in the TT subspace of the full phase space,

$$\{h^{TT}_{ij}(x), \pi^{kl}_{TT}(y)\} = (TT)^{kl}_{ij}\delta^{(3)}(x-y),$$

means that h^{TT} and π^{TT} are a canonical pair.

As in the spin-1 case, the form of the reduced Hamiltonian would depend on the choice of the gauge; however, the second-order field equation for h^{TT} is quite independent of that, i.e.,

$$\Box h_{ij}^{TT} = -\kappa\Phi_{ij}^{TT}, \tag{23}$$

where $\Phi_{ij}^{TT} \equiv (TT)_{ij}^{kl}\Phi_{kl}$.

B. The $m \neq 0$ case

The choice of a mass term for the spin-2 field is even more diverse than for the free Lagrangian. A form sufficiently general to allow the study of a class of theories is

$$-\tfrac{1}{2}m^2[h_{\mu\nu}h^{\mu\nu} - \alpha h^{\mu}_{\mu}h^{\nu}_{\nu}]. \tag{24}$$

It is well-known that

$\alpha = 1 \Leftrightarrow$ pure spin-2 theory,

$\alpha \neq 1 \Leftrightarrow$ scalar spin-2 theory.

We shall discuss the physical contents of these two theories later. For now, let us say that the complete Lagrangian density for the massive spin-2 field is $\mathcal{L}'$ of (18) plus the mass term of (24). However, to put the constraints in their simplest form we take the canonically transformed Lagrangian of (19), i.e.,

$$\mathcal{L}^m = \mathcal{L} - \tfrac{1}{2}m^2[h_{\mu\nu}h^{\mu\nu} - \alpha h^{\mu}_{\mu}h^{\nu}_{\nu}]. \tag{25}$$

1. $\alpha = 1$ case

As we shall see below, this case corresponds to a free, pure, massive spin-2. The primary constraints are unaffected by the mass term. We have

$$\zeta^0 \equiv \pi^{00} \approx 0$$

$$\zeta^i \equiv \pi^{0i} \approx 0.$$

π^{ij}, too, is unaffected and is given in (20).

The primary Hamiltonian becomes

$$H_p^m = \int d^3x\, \mathcal{H}_p^m$$

$$= \int d^3x[\mathcal{H}_c^m + v_0(x)\zeta^0(x) + v_i(x)\zeta^i(x)], \tag{26}$$

where the canonical Hamiltonian density $\mathcal{H}_c^m$ is,

$$\mathcal{H}_c^m = \mathcal{H}_c + \tfrac{1}{2}m^2[h_{\mu\nu}h^{\mu\nu} - h^{\mu}_{\mu}h^{\nu}_{\nu}], \tag{27}$$

and $\mathcal{H}_c$ is given in (21).

Due to the lack of gauge invariance, the multipliers $v_0(x)$ and $v_i(x)$ are no longer arbitrary and they will be found below. The time evolution of the PC's according to H_p^m gives us four of the secondary constraints,

$$0 \approx \partial_0\zeta^0 = \{\pi^{00}, H_p^m\} = -(m^2 - \nabla^2)h^i_i - \partial_i\partial_j h^{ij} + \kappa\Phi^{00} \equiv \rho^0(x), \tag{28}$$

$$0 \approx \partial_0\zeta^i = \{\pi^{0i}, H_p^m\} = +\partial_j\pi^{ij} - m^2h^{0i} + \kappa\Phi^{0i} \equiv \rho^i(x). \tag{29}$$

But there are more secondary constraints,

$$0 \approx \partial_0\rho^0 = \{[-(m^2 - \nabla^2)h^i_i - \partial_i\partial_j h^{ij}], H_p^m\} + \kappa\partial_0\Phi^{00}$$

$$\approx m^2[\partial_i h^{0i} + \tfrac{1}{2}\pi^i_i] + \kappa\partial_\mu\Phi^{0\mu} \equiv \omega^0(x), \tag{30}$$

where we have used (29) to put this constraint in a form suitable for our objective (taking linear combinations of constraints leaves the DB's invariant, and the reduced theory unaffected). We also note that by an alternative use of $\rho^i \approx 0$, one can put (30) in a more "symmetric" form when compared with (28),

$$\omega'^0(x) \equiv \tfrac{1}{2}m^2\pi^i_i + \partial_{ij}\pi^{ij} + \kappa[\partial_\mu\Phi^{\mu 0} - \partial_i\Phi^{i0}] \approx 0. \tag{31}$$

The preservation of $\rho^i \approx 0$ in time does not give a secondary constraint; instead, it tells us what the multiplier $v_i(x)$ is

$$0 \approx \partial_0\rho^i = \tfrac{1}{2}m^2v^i(x) - m^2[\partial_j h^{ij} - \partial^i h^{\mu}_{\mu}] + \partial_\mu\Phi^{\mu i}.$$

Being a multiplier condition, this equation will be written as the strong equation

$$\tfrac{1}{2}m^2v^i(x) = m^2[\partial_j h^{ij} - \partial^i h^{\mu}_{\mu}] - \kappa\partial_\mu\Phi^{\mu i}. \tag{32}$$

The time evolution of this multiplier condition gives no constraints, only the time derivative of $v^i(x)$. At this point we note that

$$\partial_0 h^{0i} = \{h^{0i}(x), H_p^m\} = \tfrac{1}{2}v^i(x). \tag{33}$$

So in fact $v_i(x)$ is (two times) one of the velocities for which one could not solve, in the massless case, on account of gauge invariance. Here, however, it is determined by (32).

Going on, we find the last of the secondary constraints,

$$0 \approx \partial_0\omega^0 = \{m^2(\partial_i h^{0i} + \tfrac{1}{2}\pi^i_i), H_p^m\} + \kappa\partial_{0\mu}\Phi^{0\mu}$$

$$\approx \tfrac{3}{2}m^4h^{\mu}_{\mu} + \kappa[\tfrac{1}{2}m^2\Phi^{\mu}_{\mu} + \partial_\mu\partial_\nu\Phi^{\mu\nu}] \equiv \eta^0(x) \tag{34}$$

where we have used (32) and (28).

The preservation of (34) gives another multiplier condition,

$$0 \approx \partial_0\eta^0 = \{\tfrac{3}{2}m^4h^{\mu}_{\mu}, H_p^m\} + \tfrac{1}{2}\kappa m^2\partial_0\Phi^{\mu}_{\mu} + \kappa\partial_0\partial_\mu\partial_\nu\Phi^{\mu\nu}$$

$$\approx -\tfrac{3}{2}m^4v_0(x) + \tfrac{3}{4}m^4\pi^i_i + \tfrac{1}{2}\kappa m^2\partial_0\Phi^{\mu}_{\mu} + 3\kappa m^2\partial_\mu\Phi^{0\mu} + \kappa\partial_0\partial_\mu\partial_\nu\Phi^{\mu\nu},$$

in which we have used (30). And this we write as the strong equation

$$m^4v_0(x) = \tfrac{1}{2}m^4\pi^i_i + \kappa[\tfrac{1}{3}m^2\partial_0\Phi^{\mu}_{\mu} + 2m^2\partial_\mu\Phi^{0\mu} + \tfrac{2}{3}\partial_0\partial_\mu\partial_\nu\Phi^{\mu\nu}]. \tag{35}$$

Note that

$$\partial_0 h_{00}(x) = \{h_{00}(x), H_p^m\} = v_0(x),$$

so $v_0(x)$ is indeed the other velocity [see (33)] which in the massless case was undetermined, but now is completely fixed by (35).

This exhausts all the constraints and the multiplier conditions. Here we list them again for later reference. The constraints are

$$\zeta^0 \equiv \pi^{00}, \tag{36a}$$

$$\zeta^i \equiv \pi^{0i}, \tag{36b}$$

$$\rho^0 \equiv -[m^2 - v^2]h^i_i - \partial_{ij}h^{ij} + \kappa\Phi^{00}, \tag{36c}$$

$$\rho^i \equiv \partial_j \pi^{ij} - m^2 h^{0i} + \kappa \Phi^{0i}, \tag{36d}$$

$$\omega^0 \equiv m^2[\partial_i h^{0i} + \tfrac{1}{2}\pi^i_i] + \kappa \partial_\mu \Phi^{0\mu}, \tag{36e}$$

$$\eta^0 \equiv \tfrac{3}{2} m^4 h^\mu_\mu + \kappa[\tfrac{1}{2} m^2 \Phi^\mu_\mu + \partial_\mu \partial_\nu \Phi^{\mu\nu}], \tag{36f}$$

and the multipliers

$$m^2 v^i(x) \equiv 2m^2[\partial_j h^{ij} - \partial^i h^\mu_\mu] - 2\kappa \partial_\mu \Phi^{i\mu}, \tag{37a}$$

$$m^4 v_0(x) \equiv \tfrac{1}{2} m^4 \pi^i_i + \kappa[\tfrac{1}{3} m^2 \partial_0 \Phi^\mu_\mu + 2m^2 \partial_\mu \Phi^{0\mu} + \tfrac{2}{3}\partial_0 \partial_\mu \partial_\nu \Phi^{\mu\nu}]. \tag{37b}$$

A brief discussion of these constraints and the multiplier conditions is in order. First, not all of the constraints are of the same type, and in the reduction process they will not all be treated in the same way: some simply set equal to zero some unphysical degree of freedom (examples, ζ^0 and ζ^i), and others remove one degree of freedom in favor of another (example, ρ^0, ρ^i, ω^0). And for the massive spin-2 theory, one of the constraints (η^0) fits none of these categories. One might expect to use η^0 to remove h^{00} in favor of h^i_i (and sources). However, the coefficient of h^{00} in the action (and Hamiltonian) is exactly the constraint ρ^0; so h^{00} is the Lagrange multiplier whose variation enforces the ρ^0 constraint. Although one does not actively use the η^0 constraint to remove h^{00}, it would be a mistake not to count it as a constraint, for its presence affects the dimension of the reduced phase space.

According to the discussion in Sec. II A, the dimension of the reduced phase space is $[20 - 1(10)] = 5 + 5$, corresponding to five h fields and five π fields. We have used the fact that all ten constraints are second class [see (B2)]. As expected, five is the dimension of the irreducible (i.e., traceless) part of the $\square\square$ representation of the little group $SO(D-1) = SO(3)$, so we are, in fact, dealing with a pure spin-2 massive field.

We can already notice that the *implication* of conservation of sources, in the limit $m \to 0$, is in contrast to what we have observed for the spin-1 case. There, the two second class (SC) constraints of the massive theory became first class (FC) (the two constraints of the massless theory) in the limit $m \to 0$. And the multiplier condition of the massive theory became the expression for conservation of sources, which is essential in the massless theory. Here, in the spin-2 case, the situation is best described with the aid of a self-explanatory diagram (see Table I).

At this stage it is possible to begin solving the constraints strongly for the variables that one chooses to remove from the Hamiltonian or the action. As indicated above, at times one has no choice (for example, $\pi^{0\mu} = 0$), but in certain cases there is no unique variable that one may call unphysical. This can be seen in the ρ^0 constraint, which essentially relates two scalar degrees of freedom. At first sight it is tempting to use this constraint to remove h^i_i. Although allowed, this is by no means the only possibility. One can decompose h_{ij} in a variety of manners (orthogonal or not) and then solve the constraint for one scalar piece of h_{ij} in terms of another. Evidently, in the presence of constraints of the type ρ^0, there is no unique scalar piece of h_{ij} that one can call physical (or unphysical); in principle one has some freedom in the choice of the variable to be removed.

Simplicity of the algebra might be a strong guiding principle in the decision of which variable to remove. Another factor which may influence a decision is whether or not the prospective physical variables and their conjugates have simple commutation relations (of course, in the reduction of the Lagrangian where one is not at all interested in the conjugate variables this point is irrelevant). We also remark that, for some degree of freedom, the action may not reflect explicit Lorentz covariance.

Since, in the reduction process, one has to use the constraints as strong equations, what is meant by "commutation relations" is, of course, commutation relations in terms of *Dirac brackets* [on account of (7)], which we have computed in Appendix B. Thus, to proceed, we shall treat the constraints as strong equations and accordingly use DB's. Once we have the fully reduced Hamiltonian on the RPS, the Dirac brackets will, of course, be equivalent to Poisson brackets in the physical degrees of freedom.

2. Reduced phase space (RPS)

Thus we are looking for DDF's, q and p such that they obey the following schematic commutation relations:

$$\{q(x), p(y)\}_D = \mathbb{1}\delta^{(3)}(x-y),$$

$$\{q(x), q(y)\}_D = 0 = \{p(x), p(y)\}_D,$$

where $\mathbb{1}$ is an appropriate identity operator. It is clear that h^{00} and h^{0i} cannot be physical, on account of $\pi^{00} = 0$, $\pi^{0i} = 0$, and the DB of (B5). Let us then concentrate on the i,j components. We can rewrite (B4) as

$$\{h_{ij}(x), \pi^{kl}(y)\}_D = \{h_{ij}(x), \pi^{kl}(y)\} + (2/3m^4)[\nabla^2 L_{ij} + \tfrac{1}{2} m^2 \eta_{ij}] \times [\nabla^2 T^{kl} - m^2 \eta^{kl}]\delta(x-y), \tag{38}$$

where L_{ij} and T_{ij} have been discussed in Appendix A. Notice that

$$\{h_{ij}(x), \pi^{ij}(y)\}_D = 5\delta^{(3)}(x-y), \tag{39}$$

expressing the fact that, of the six degrees of freedom in h_{ij} (or π^{kl}), only five are independent (i.e., physical). This can of course be seen from the constraints, which can be massaged to resemble the operators appearing in (38):

$$[\nabla^2 T_{ij} - m^2 \eta_{ij}] h^{ij} + \kappa \Phi^{00} = 0, \tag{40a}$$

$$[\nabla^2 L_{ij} + \tfrac{1}{2} m^2 \eta_{ij}] \pi^{ij} + \kappa[\partial_\mu \Phi^{0\mu} + \partial_i \Phi^{0i}] = 0. \tag{40b}$$

We have already stated a number of considerations which will govern our determination of the RPS. We now

TABLE I. This table illustrates the $m \to 0$ behavior of the constraints and the multiplier conditions. (FC) and (SC) stand for first and second class, respectively.

$m \neq 0$		$m = 0$
eight (SC) constraints	$\Rightarrow$	eight (FC) constraints
one (SC) constraint	$\Rightarrow$	$\partial_\mu \Phi^{\mu 0} = 0$
one (SC) constraint	$\Rightarrow$	$\partial_\mu \partial_\nu \Phi^{\mu\nu} = 0$
v^i condition	$\Rightarrow$	$\partial_\mu \Phi^{\mu i} = 0$
v_0 condition	$\Rightarrow$	$0 = 0$

demand as an additional requirement, that the DDF's—q_{ij} and p^{kl}—satisfy

$$\{q_{ij}(x),p^{kl}(y)\}_D = 1^{ij}_{kl}\delta^{(3)}(x-y),$$
$$\{q_{ij}(x),q^{kl}(y)\}_D = 0 = \{p_{ij}(x),p^{kl}(y)\}_D, \tag{41}$$

where we have yet to decide what 1^{ij}_{kl} should be. From (39) we see that 1^{ij}_{kl} must have rank-5. But we would also like it to be a constant operator, i.e., no factors of m. A choice which obviously meets these conditions is

$$1^{kl}_{ij} = \tfrac{1}{2}(\delta^k_i\delta^l_j + \delta^l_i\delta^k_j) - \tfrac{1}{3}\eta_{ij}\eta^{kl}. \tag{42a}$$

Note that with this particular choice

$$1^{kl}_{ii} = 0 = 1^{kk}_{ij}. \tag{42b}$$

Recall that q_{ij} represents only five of the six degrees of freedom in h_{ij}. We would like to write q_{ij} as a projection of h_{ij}, and similarly for p^{ij} in terms of π^{ij}, i.e.,

$$q_{ij} = \theta^{kl}_{ij}h_{kl}, \quad p^{ij} = \tilde{\theta}^{ij}_{kl}\pi^{kl},$$

where θ and $\tilde{\theta}$ are projection operators of rank-5. With the choice (42a) we see from (42b) that q and p will be traceless,

$$q^i_i = 0, \quad p^i_i = 0.$$

Mindful of the considerations given above, we have been able to write down a variety of canonical pairs q_{ij}, p^{ij} representing the physical degrees of freedom on the reduced phase space. Perhaps the simplest which can easily be shown from (38) to satisfy all of our requirements are

$$q_{ij} = h_{ij} - L_{ij}h^k_{\ k} = [\tfrac{1}{2}(\delta^k_i\delta^l_j + \delta^l_i\delta^k_j) - L_{ij}\eta^{kl}]h_{kl}, \tag{43}$$

and

$$p^{ij} = \pi^{ij} - \tfrac{1}{2}T^{ij}\pi^k_k = [\tfrac{1}{2}(\delta^i_k\delta^j_l + \delta^i_l\delta^j_k) - \tfrac{1}{2}T^{ij}\eta_{kl}]\pi^{kl}. \tag{44}$$

We can decompose the q and p of (43), (44) [see (A.3)] as

$$q_{ij} = q^{TT}_{ij} + q^{TL}_{ij} + q^{T2L}_{ij}, \tag{45a}$$

corresponding to 2, 2, and 1 degree of freedom, respectively. Similarly

$$p_{ij} = p^{TT}_{ij} + p^{TL}_{ij} + p^{T2L}_{ij}. \tag{45b}$$

In terms of the various projections of h and π, we can write (see Appendix A)

$$\begin{aligned} q^{TT}_{ij} &= h^{TT}_{ij} \equiv (TT)^{kl}_{ij}h_{kl}, \\ q^{TL}_{ij} &= h^{TL}_{ij} \equiv (TL)^{kl}_{ij}h_{kl}, \\ q^{T2L}_{ij} &\equiv \tfrac{1}{2}[T_{ij} - 2L_{ij}]T^{kl}h_{kl} \neq (T2L)^{kl}_{ij}h_{kl}, \\ p^{TT}_{ij} &= \pi^{TT}_{ij} \equiv (TT)^{kl}_{ij}\pi_{kl}, \\ p^{TL}_{ij} &= \pi^{TL}_{ij} \equiv (TL)^{kl}_{ij}\pi_{kl}, \\ p^{T2L}_{ij} &\equiv -\tfrac{1}{2}[T_{ij} - 2L_{ij}]L^{kl}\pi_{kl} \neq (T2L)^{kl}_{ij}\pi_{kl}. \end{aligned} \tag{46}$$

Note the lack of symmetry in q^{T2L} and p^{T2L}, which is simply a consequence of the lack of symmetry between (40a) and (40b). We point out that the choice of degrees of freedom used here (of h and π) is equivalent to an initial choice made by previous authors.[3]

Before attending to the matter of reduction, we stop to elaborate on the *meaning* of a reduced Hamiltonian. In contrast with reduction of the Lagrangian, for the Hamiltonian the reduction can be subtle. First, we give a general discussion of this subtlety; and then as an application we specialize to the case at hand.

3. On Hamiltonian reduction. General considerations

The concern of this section is with the claim that a reduced Hamiltonian is not necessarily the Hamiltonian on the RPS. To elaborate, we start by more carefully defining each term.

A reduced Hamiltonian for the RPS variables q and p is the function which one obtains simply by setting the constraints equal to zero in the full Hamiltonian, i.e.,

$$H\,|\,(q,p) \equiv H(q,p;\ \text{constraint} = 0).$$

A reduced Lagrangian would be defined in a similar way

$$L\,|\,(q,\dot{q}) \equiv L(q,\dot{q};\ \text{constraint} = 0).$$

And, of course

$$H(q,p;\ \text{constraint}) \equiv p\dot{q} - L(q,\dot{q};\ \text{constraint}),$$

where $p = \delta L/\delta\dot{q}$. Hence

$$H\,|\,(q,p) = (p\dot{q})|_{\text{constraint}=0} - L(q,\dot{q}). \tag{47}$$

But for a function $H(q,p)$ to be a true Hamiltonian for the q,p variables, one must be able to write

$$H(q,p) = p\dot{q} - L(q,\dot{q}), \tag{48}$$

for only then will the extremization of the action $\int dt\, L(q,\dot{q})$ give rise to the Hamiltonian equations of motion

$$\dot{q} = \frac{\delta H(q,p)}{\delta p}, \quad \dot{p} = -\frac{\delta H(q,p)}{\delta q}.$$

Equation (47) defines a reduced Hamiltonian, and (48) is the definition of a Hamiltonian for the q,p variables (i.e., RPS variables). Now, in general

$$(p\dot{q})|_{\text{constraint}=0} \neq (p\dot{q}),$$

whereas, by definition $L(q,\dot{q}) = L\,|\,(q,\dot{q})$. Hence the statement mentioned above—that the reduced Hamiltonian may not be equal to the Hamiltonian on the RPS.

At this stage we might consider doing one of two things; we can find a "new" RPS, $\tilde{q}$ and $\tilde{p}$, with

$$q = q(\tilde{q},\tilde{p}), \quad p = p(\tilde{q},\tilde{p}),$$

such that

$$(p(\tilde{q},\tilde{p})\dot{q}(\tilde{q},\tilde{p}))|_{\text{constraint}=0} = \tilde{p}\dot{\tilde{q}},$$

modulo constant terms. Then, the reduced Hamiltonian *will be* the Hamiltonian for the new RPS variables $\tilde{q}$ and $\tilde{p}$, i.e.,

$$H(q = q(\tilde{q},\tilde{p}), p = (\tilde{q},\tilde{p})) = \tilde{p}\dot{\tilde{q}} - L(\tilde{q},\dot{\tilde{q}}).$$

On the other hand, we can simply compute $(p\dot{q})|_{\text{constraint}=0}$ as

$$(p\dot{q})|_{\text{constraint}=0} \equiv p\dot{q} + \Delta H(\dot{q},p). \tag{49}$$

Here, the Hamiltonian for the reduced variables q and p will be

$$H(q,p) = H\,|\,(q,p) - \Delta H(\dot{q}(q,p),p),$$

where $\Delta H(\dot{q}(q,p),p)$ is *defined* by (49) and $\dot{q} = \dot{q}(q,p)$ is obtained from the inversion of $p = \delta L(q,\dot{q})/\delta\dot{q}$.

This concludes our general discussion on Hamiltonian reduction. Next we illustrate these points with the examples which concern us in this paper.

Examples Spin-1 (see Sec. II): In this case we have

$$\int d^3x(\pi^i\dot{A}_i)|_{\text{constraint}=0} = \int d^3x(\pi^i\dot{A}_i)$$

$$\equiv \int d^3x(p^i\dot{q}_i).$$

So, in fact, here is a trivial case where the reduced Hamiltonian is equal to the Hamiltonian on the RPS spanned by the variables $q_i \equiv A_i$ and $p^i \equiv \pi^i$. This is a result which we have already used in our treatment of the spin-1 theory.

Spin-2 (see Sec. III B 1): Here the presence of the constraints (40) changes the situation. With (43) and (44) these constraints can be written as

$$(L\cdot q) + \frac{m^2}{\nabla^2} h^i_i - \frac{\kappa}{\nabla^2}\Phi^{00} = 0, \tag{50}$$

$$(L\cdot p) + \frac{1}{2}\frac{m^2}{\nabla^2}\pi^i_i + \frac{\kappa}{\nabla^2}(\partial_\mu\Phi^{0\mu} + \partial_i\Phi^{0i}) = 0, \tag{51}$$

where $L\cdot q$ stands for $L_{ij}q^{ij}$, etc. Again by using (43) and (44), we can write

$$\int d^3x(\pi\cdot\dot{h})|_{\text{constraint}=0}$$

$$= \int d^3x(p^{ij} + \frac{1}{2}T^{ij}\pi^k_k)\partial_0(q_{ij} + L_{ij}h^k_k)|_{\text{constraint}=0}$$

$$= \int d^3x\{(p\dot{q}) + \frac{\kappa}{m^2}[(L\cdot p)\partial_0\Phi^{00}$$

$$+ (L\cdot\dot{q})(\partial_\mu\Phi^{0\mu} + \partial_i\Phi^{0i})]\}, \tag{52}$$

where we have used the constraints (50) and (51) to remove the three-traces from the first equation. We can rewrite this equation as

$$= \int d^3x\left[p^{ij} - \frac{\kappa}{3m^2}(T-2L)^{ij}(\partial_\mu\Phi^{0\mu} + \partial_k\Phi^{0k})\right]$$

$$\times\partial_0\left[q_{ij} - \frac{\kappa}{3m^2}(T-2L)_{ij}\Phi^{00}\right] + \text{“constants,”}$$

where “constants” refers to a term of order $(\kappa)^2$ and a spatial boundary term resulting from a spatial partial integration.

Now, modulo constant terms, this is precisely of the form $\int d^3x(\tilde{p}\cdot\dot{\tilde{q}})$, with

$$\tilde{q}_{ij} \equiv q_{ij} - (\kappa/3m^2)(T-2L)_{ij}\Phi^{00},$$

$$\tilde{p}_{ij} \equiv p_{ij} - (\kappa/3m^2)(T-2L)_{ij}(\partial_\mu\Phi^{0\mu} + \partial_k\Phi^{0k}).$$

So the reduced Hamiltonian $H|(q=q(\tilde{q}),\, p=p(\tilde{p}))$ is the Hamiltonian on the RPS as defined by $\tilde{q}$ and $\tilde{p}$.

Rather than using these shifted variables we may also use the “old” variables q and p and then the Hamiltonian for these variables is

$$H(q,p) = H|(q,p) - \Delta H(\dot{q}(q,p),p),$$

where $H|(q,p)$ is the reduced Hamiltonian, and $\Delta H(\dot{q},p)$ can be read off from (52), i.e.,

$$\Delta H(\dot{q},p) = \int d^3x\frac{\kappa}{m^2}[(L\cdot p)(\partial_0\Phi^{00})$$

$$+ (L\cdot\dot{q})(\partial_\mu\Phi^{0\mu} + \partial_i\Phi^{0i})]. \tag{53}$$

To find $\dot{q} = \dot{q}(q,p)$ one could calculate $p = \delta L(q,\dot{q})/\delta\dot{q}$ and invert the resulting expression. However, since all we need in (53) is the relation between $(L\cdot\dot{q})$ and $(L\cdot p)$, we can take a shortcut. First, note that (20) implies

$$\pi^i_i = -2\partial_0 h^i_i + 4\partial^i h_{0i}.$$

Then by using (36e) it follows that

$$\partial_0 h^i_i = \tfrac{1}{2}\pi^i_i - (2\kappa/m^2)\partial_\mu\Phi^{0\mu}.$$

Taking ∂_0 of (50) and combining with (51) we find that

$$L\cdot\dot{q} = L\cdot p.$$

The final expression for the Hamiltonian in the q,p RPS becomes

$$H(q,p) = H|(q,p) - \int d^3x\frac{2\kappa}{m^2}(L\cdot p)(\partial_\mu\Phi^{0\mu}), \tag{54}$$

with $H|(q,p)$ being the reduced Hamiltonian, to be computed in the next section. Curiously, $\Delta H = 0$ if the sources are conserved.

4. Reduced Hamiltonian and equations of motion

Having demonstrated the significance for the complete theory of the reduced Hamiltonian, we now turn to its computation for the spin-2 case. We begin by consecutively imposing the constraints of (36) for a systematic reduction of the Hamiltonian in (26).

Imposing the constraints ζ^0 and ζ^i simply reduces the primary Hamiltonian back to the canonical Hamiltonian of (27), thereby eliminating the need for the multiplier conditions. We shall keep (36c) until the last stage of the reduction. ρ^i is used to remove all h^{0i} dependence. ω^0 combined with ρ^i give the “π analog” of the ρ^0 constraint which we have already listed in (31). This, too, we shall keep for later application. And, as we indicated in Sec. III B 1, the η^0 constraint which might be expected to remove h^{00} in favor of h^i_i (or eventually $L\cdot q$) and sources, is completely passive; the coefficient of h^{00} in the Hamiltonian (and Lagrangian) is precisely the ρ^0 constraint. The resulting partially reduced Hamiltonian is

$$\int d^3x\left[\frac{1}{2}\pi\cdot\pi - \frac{1}{4}(\pi^i_i)^2 + \frac{1}{m^2}(\partial\pi)\cdot(\partial\pi) + \frac{2\kappa}{m^2}\pi^{ij}\partial_i\Phi_{j0}\right.$$

$$+\frac{1}{2}h\cdot(m^2-\nabla^2)h + \frac{1}{2}h^i_i(m^2-\nabla^2)h^j_j$$

$$\left.-(\partial h)\cdot(\partial h) - \kappa h^i_i\Phi^{00} - \kappa h\cdot\Phi\right],$$

where a constant term of order $(\kappa)^2$, and boundary terms resulting from spatial partial integrations have been dropped.

In working towards a fully reduced Hamiltonian [i.e., $H|(q,p)$] we can now use the definition of (43) and (44) to write

$$h_{ij} = q_{ij} + L_{ij}h^k_k, \quad \pi_{ij} = p_{ij} + \tfrac{1}{2}T_{ij}\pi^k_k.$$

The result is

$$\int d^3x\left[\frac{1}{2}p\cdot p - \frac{1}{2}(L\cdot p)\pi^k_k + \frac{1}{m^2}(\partial p)\cdot(\partial p) + \frac{2\kappa}{m^2}p^{ij}\partial_i\Phi_{j0}\right.$$

$$+\frac{1}{2}q\cdot(m^2-\nabla^2)q + (L\cdot q)(m^2+\nabla^2)h^k_k - (\partial q)\cdot(\partial q)$$

$$\left.+ m^2(h^i_i)^2 - \kappa h^i_i(\Phi^{00} + L\cdot\Phi) - \kappa q\cdot\Phi\right],$$

where again spatial partial integrations have been performed at will, and we have used $q_i^i = T\cdot q + L\cdot q = 0$.The remaining two constraints, ρ^0 and ω'^0, which we have delayed using until now, are given in a suitable form in Eqs. (50) and (51). These can now be used to remove h_i^i and π_i^i in favor of $L\cdot q$ and $L\cdot p$, respectively. The resulting reduced Hamiltonian is

$$H|(q,p) = \int d^3x\Big\{\frac{1}{2}p\cdot p - (L\cdot p)\Big(\frac{\nabla^2}{m^2}\Big)(L\cdot p) + \Big(\frac{1}{m^2}\Big)(\partial p)\cdot(\partial p) + \Big(\frac{\kappa}{m^2}\Big)[(L\cdot p)(\partial_\mu\Phi^{0\mu} + \partial_i\Phi^{0i}) + 2p^{ij}\partial_i\Phi_{j0}]$$
$$+\frac{1}{2}q\cdot(m^2-\nabla^2)q - (L\cdot q)\nabla^2(L\cdot q) - (\partial q)\cdot(\partial q) + \kappa(L\cdot q)\Big[\Phi^{00} + \Big(\frac{\nabla^2}{m^2}\Big)L\cdot\Phi\Big] - \kappa q\cdot\Phi\Big\}.$$

As previously, terms of order $(\kappa)^2$ have been dropped in the above.

As discussed in Sec. III B 3 the true Hamiltonian for the RPS variables q and p is given by (54). Thus

$$H(q,p) = H|(q,p) - \int d^3x\frac{2\kappa}{m^2}(L\cdot p)(\partial_\mu\Phi^{0\mu})$$
$$= \int d^3x\Big\{\frac{1}{2}p\cdot p + (L\cdot p)\Big(\frac{\nabla^2}{m^2}\Big)(L\cdot p) + \Big(\frac{1}{m^2}\Big)(\partial p)\cdot(\partial p) + \Big(\frac{\kappa}{m^2}\Big)[-(L\cdot p)\partial_0\Phi^{00} + 2p^{ij}\partial_i\Phi_{j0}]$$
$$+\frac{1}{2}q\cdot(m^2-\nabla^2)q - (L\cdot q)\nabla^2(L\cdot q) - (\partial q)\cdot(\partial q) + \kappa\Big[(L\cdot q)\Big(\Phi^{00} + \Big(\frac{\nabla^2}{m^2}\Big)L\cdot\Phi\Big) - q\cdot\Phi\Big]\Big\}.$$

This expression becomes much more transparent if we decompose all of $H(q,p)$ according to the scheme used in (45); i.e.,

$$H(q,p) = H^{TT}(q^{TT},p^{TT}) + H^{TL}(q^{TL},p^{TL}) + H^{T2L}(q^{T2L},p^{T2L}),$$

where

$$H^{TT}(q^{TT},p^{TT}) \equiv H^{TT}| = \int d^3x\Big[\frac{1}{2}p^{TT}\cdot p^{TT} + \frac{1}{2}q^{TT}\cdot(m^2-\nabla^2)q^{TT} - \kappa q^{TT}\cdot\Phi^{TT}\Big], \tag{55a}$$

$$H^{TL}(q^{TL},p^{TL}) \equiv H^{TL}| = \int d^3x\Big[\frac{1}{2}p^{TL}\cdot\Big(1-\frac{\nabla^2}{m^2}\Big)p^{TL} + \frac{2\kappa}{m^2}p^{ij}_{TL}\,\partial_i\Phi_{j0} + \frac{1}{2}m^2q^{TL}\cdot q^{TL} - \kappa q^{TL}\cdot\Phi^{TL}\Big], \tag{55b}$$

$$H^{T2L}(q^{T2L},p^{T2L}) \equiv H^{T2L}| = \int d^3x\Big[\frac{1}{2}p^{T2L}\cdot p^{T2L} - \frac{\kappa}{m^2}(L\cdot p^{T2L})(\partial_0\Phi^{00} + 2\,\partial_i\Phi^{0i})$$
$$+\frac{1}{2}q^{T2L}\cdot(m^2-\nabla^2)q^{T2L} + \kappa(L\cdot q^{T2L})\Big(\Phi^{00} + \Big(\frac{\nabla^2}{m^2}\Big)L\cdot\Phi\Big) - \kappa q^{T2L}\cdot\Phi^{T2L}\Big]. \tag{55c}$$

Even though this is the most manifest form of the Hamiltonian on the RPS, it might be discomforting to see a $[m^2]^{-1}$ appearing in (55b) and (55c). There are several reasons for why this m dependence is not a serious problem. First, as we shall see, when the second-order equations of motion are constructed, $[m^2]^{-1}$ will often be the coefficient of a term containing some form of the expression for the conservation of sources. Then, for conserved sources the second order equations of motion will not contain $[m^2]^{-1}$. Second, the form of (55) is certainly dependent on what variables the constraints are solved for in the reduction process. But as we have already argued, it is only the invariant S-matrix elements which are physically relevant. The requirement of a nonpathological Hamiltonian (i.e., no $[m^2]^{-1}$) was not one of our criteria in constructing a reduced phase space. Nor is it necessary, for one can always redefine the fields in (55) to make the $[m^2]^{-1}$ disappear. With the freedom of field redefinition at our disposal, it is only the way the physical quantities scale under the scaling of the fields which really concerns us. We shall return to this point later.

The Hamiltonian equations of motion for the various components of q and p are given by

$$\dot q = \{q,H\}_D,\quad \dot p = \{p,H\}_D,$$

where H is as in (55). Due to the orthogonality of the decomposition (45), only H^{TT} contributes to $\dot q^{TT}$ and $\dot p^{TT}$. Similarly for the other components of q and p. We obtain

$$\dot q^{TT}_{ij} = p^{TT}_{ij}, \tag{56a}$$
$$\dot q^{TL}_{ij} = (1-\nabla^2/m^2)p^{TL}_{ij} + (2\kappa/m^2)(TL)^{kl}_{ij}\partial_k\Phi_{l0}, \tag{56b}$$
$$\dot q^{T2L}_{ij} = p^{T2L}_{ij} - (\kappa/m^2)(T2L)^{kl}_{ij}L_{kl}(\partial_0\Phi^{00} + 2\partial_i\Phi^{0i}), \tag{56c}$$

$$\dot p^{TT}_{ij} = -(m^2-\nabla^2)q^{TT}_{ij} + \kappa\Phi^{TT}_{ij}, \tag{57a}$$
$$\dot p^{TL}_{ij} = -m^2q^{TL}_{ij} + \kappa\Phi^{TL}_{ij}, \tag{57b}$$
$$\dot p^{T2L}_{ij} = -(m^2-\nabla^2)q^{T2L}_{ij} - \kappa(T2L)^{kl}_{ij}\times L_{kl}(\Phi^{00} + (\nabla^2/m^2)L\cdot\Phi) + \kappa\Phi^{T2L}_{ij}. \tag{57c}$$

5. Reduced configuration space and the $m\to 0$ limit

The second-order equations of motion for q^{TT} are quite straightforward to obtain. Solving (56a) for p^{TT}_{ij} and substituting in (57a) gives

$$(\Box - m^2)q^{TT}_{ij} = -\kappa\Phi^{TT}_{ij}.$$

In terms of h_{ij}^{TT}, (46) gives $(\Box - m^2)h_{ij}^{TT} = -\kappa\Phi_{ij}^{TT}$.

The discussion of the limit as $m\to 0$ of q^{TT} is similar to that of A^T in the spin-1 case. The equation of motion for q^{TT}, in the limit $m\to 0$, is equal to the equation of motion for h^{TT} [i.e., a massless spin-2, see (23)]. So we can say that the TT degrees of freedom of a massive spin-2 field smoothly go over to the TT degrees of freedom (helicity ± 2) of a massless spin-2 field.

The two degree of freedom in q^{TL} behave as the one degree of freedom A^L did in the spin-1 case, with certain subtle differences. Equations (56b) and (57b) resemble their spin-1 analog (11b) and (11d). Thus we shall treat q^{TL} in a similar fashion (see Sec. II C), i.e., solve (56b) for p^{TL}, take its time derivative and substitute the results into $(V^2 - m^2)$ times (57b). We get

$$m^2(\Box - m^2)q_{ij}^{TL} = -m^2\kappa\Phi_{ij}^{TL} + 2\kappa(TL)_{ij}^{kl}\partial_k\partial_\mu\Phi_l^\mu. \tag{58}$$

In terms of h_{ij}^{TL}, (46) gives

$$m^2(\Box - m^2)h_{ij}^{TL} = -m^2\kappa\Phi_{ij}^{TL} + 2\kappa(TL)_{ij}^{kl}\partial_k\partial_\mu\Phi_l^\mu.$$

On letting $m\to 0$ in (58) we find that q^{TL} becomes nondynamical (in the absence of a kinetic term), decouples from the TL part of the source, and leaves us with

$$(TL)_{ij}^{kl}\partial_k\partial_\mu\Phi_l^\mu = 0. \tag{59}$$

Already, there is one crucial difference compared with the spin-1 result. In spin-1, conservation of sources (being only one algebraic equation) could be derived from the $m\to 0$ limit of the equation of motion for A_i^L (being only one degree of freedom). In spin-2, conservation of sources, $\partial_\mu\Phi^{\mu\nu} = 0$, imposes four independent conditions. The $m\to 0$ limit of the q^{TL} (representing two degrees of freedom) equations of motion gives (59), which is effectively only two algebraic conditions placed on the sources, since the operator (TL) is of rank 2. Actually, the set of equations in (59) is essentially equivalent to two conservations, $T_{ik}\partial_\mu\Phi^{\mu k} = 0$. There is still one remaining degree of freedom q^{T2L} for which the limit $m\to 0$ must be considered. As we shall see, its equation of motion can place at most one algebraic condition on the sources. In recognition of the condition imposed by the q^{TL} equation, we might anticipate its longitudinal complement, $\partial_{\mu i}\Phi^{\mu i} = 0$. In any case, based on the fact $3\neq 4$, it cannot be possible for the $m\to 0$ limit, of the equations for the resulting three unphysical degrees of freedom, to imply all four of the conservation conditions. In fact, with the particular degrees of freedom which we have used for the massive theory, not even three (strict) conservation conditions can be implied in the limit $m\to 0$ [see (63)].

As in the spin-1 case, introducing a rescaled field which becomes freely propagating in the massless limit will not alter any physical consequences of the theory, even though the rescaling may be singular in the massless limit. Thus, regardless of such rescaling, one can see, in the limit $m\to 0$, that the contribution to the S matrix is zero for the TL part of a spin-2 field (two degrees of freedom) coupled to sources which are conserved [or even not conserved, but satisfying (59)].

Next we turn to the remaining degree of freedom, represented by q^{T2L}. Its equation of motion is

$$(\Box - m^2)q_{ij}^{T2L} = -\kappa\Phi_{ij}^{T2L} + \kappa(T2L)_{ij}^{kl}L_{kl}\,\Phi^{00} + \frac{\kappa}{m^2}(T2L)_{ij}^{kl}L_{kl}(\partial_\mu\partial_\nu\Phi^{\mu\nu}). \tag{60}$$

Before considering the behavior of (6) as $m\to 0$, let us turn it into a more covariant expression. The degree of freedom, q^{T2L}, represents a scalar. From (46) we can identify this scalar as

$$(T\cdot h) = \tfrac{1}{3}(T - 2L)\cdot q^{T2L}. \tag{61}$$

Hence multiplying (60) by $\frac{1}{3}(T-2L)^{ij}$, we have

$$(\Box - m^2)(T\cdot h) = -\kappa T\cdot\Phi + \tfrac{2}{3}\kappa\Phi_\mu^\mu - (2\kappa/3m^2)(\partial_\mu\partial_\nu\Phi^{\mu\nu}). \tag{62}$$

There is one important difference between (60) and Eq. (58) for q^{TL}, and that is the appearance of $[m^2]^{-1}$ in (60). This is exactly reminiscent of the difference between (15b), for the scalar in the spin-1 theory, and (16) for the rescaled field. As the latter case, there is here a condition, viz.,

$$\partial_\mu\partial_\nu\Phi^{\mu\nu} = 0, \tag{63}$$

which *must* be satisfied for the massless theory to exist. Although it is one algebraic equation and it certainly holds for conserved sources, Eq. (63) is weaker than a conservation condition.

Clearly, even for $\partial_\mu\partial_\nu\Phi^{\mu\nu} = 0$, $(T\cdot h)$ is coupled to $T\cdot\Phi$ *and* Φ_μ^μ. One can easily demonstrate a useful identity relating $\partial_\mu\partial_\nu\Phi^{\mu\nu}$, $T\cdot\Phi$, and Φ_μ^μ;

$$T\cdot\phi = \Phi_\mu^\mu - (1/\nabla^2)(\partial_\mu\partial_\nu\Phi^{\mu\nu} - 2\partial_0\partial_\mu\Phi^{0\mu} + (1/\nabla^2)\Box\Phi^{00}. \tag{64}$$

The presence of the term $\Box\Phi^{00}$ suggests the following field redefinition as a means of obtaining a more covariant field equation,

$$\tilde h \equiv T\cdot h + (\kappa/\nabla^2)\Phi^{00} = (m^2/\nabla^2)h_i^i.$$

Then, (62) becomes

$$(\Box - m^2)\tilde h = \kappa\left[-\frac{m^2}{\nabla^2}\Phi^{00} - \tfrac{1}{3}\Phi_\mu^\mu + \frac{3m^2 - 2\nabla^2}{3m^2\nabla^2} \times(\partial_\mu\partial_\nu\Phi^{\mu\nu}) - \frac{2}{\nabla^2}(\partial_0\partial_\mu\Phi^{0\mu})\right], \tag{65}$$

and it is the field equation for the rescaled trace. Again, the discussion of the $m\to 0$ limit is possible only if $\partial_{\mu\nu}\Phi^{\mu\nu}\to 0$. And even with totally conserved sources, we still have

$$\Box\tilde h = -(\kappa/3)\Phi_\mu^\mu, \quad \text{as} \quad m\to 0.$$

In contrast to the situation for h^{TL}, which became decoupled from conserved sources in the massless limit, the scalar degree of freedom remains coupled to Φ_μ^μ. It is the attractive force due to the exchange of such a scaler (coupled to sources) which is the cause of the discontinuity in the $m\to 0$ limit.

In the scattering of two slow-moving massive conserved sources $\Phi_{\mu\nu}$ and $\Phi'_{\mu\nu}$ (i.e., $\Phi_\mu^\mu \neq 0 \neq \Phi'^\mu_\mu$), the requirement that Newton's law should follow for both the massive and the massless exchange of a spin-2 field, gives the following relationship between the coupling constants [1,3]:

$$\kappa^2(m) = \tfrac{3}{4}\kappa^2(m=0). \tag{66}$$

The factor $\frac{3}{4}$ is simply a result of the exchange of the scaler $\tilde{h}$ [see (65)]. So even if one of the sources is a massless object satisfying $\Phi^{\mu}_{\mu} = 0$ (e.g., light), the result of the scattering will still be different due to (70). In this case we could insist on $\Phi^{\mu}_{\mu} = 0$ *and* the continuity of the scattering amplitude. But then, clearly, the massive and massless cases will give rise to different Newtonian limits. "A discontinuity" is inevitable.

Concerning conservation, we already know from Table I that, for consistency, the η^0 constraint requires the condition $\partial_\mu \partial_\nu \Phi^{\mu\nu} = 0$ for the massless theory. But it is also apparent from the table that the ω^0 constraint requires the conservation condition, $\partial_\mu \Phi^{\mu 0} = 0$, in the limit $m \to 0$. Together, these two conditions *do* give the longitudinal conservation equation $\partial_\mu \partial_i \Phi^{\mu i} = 0$ which we had expected to obtain from the limit of the field equation for the scalar degree of freedom. In this manner we see that the η^0 constraint *does* play a role, even though it was not used in the construction of the Hamiltonian for the RPS.

We summarize. Of the five degrees of freedom representing a massive spin-2 field ($TT + TL + T2L$), as $m \to 0$ the TT part (two degrees of freedom) smoothly goes over the TT degrees of freedom of a massless spin-2 coupled to the TT part of the sources. The part h^{TL} becomes nondynamical, decouples from the sources, and leaves behind the requirement that

$$(TL)^{kl}_{ij} \partial_k \partial_\mu \Phi^{\mu}_{l} = 0,$$

equivalent to two of the conservation conditions on the sources

$$T_{ik} \partial_\mu \Phi^{\mu k} = 0. \tag{67}$$

Similarly, the massless limit of the q^{T2L} ($\sim T \cdot h$) field equation requires the condition

$$\partial_\mu \partial_\nu \Phi^{\mu\nu} = 0, \tag{68}$$

but the field is still coupled to sources. A redefined field $\tilde{h}$ has a purely covariant coupling to sources provided the additional condition

$$\partial_0 \partial_\mu \Phi^{0\mu} = 0 \tag{69}$$

is imposed. The coupling gives a nonzero contribution to the S matrix except for sources which are traceless:

$$\Phi^{\mu}_{\mu} = 0, \tag{70}$$

but the massive and massless theories will still have different Newtonian limits. Letting $m \to 0$ in the constraints provides the remaining conservation conditions and ensures that Eqs. (68) and (69) are satisfied in the limit.

6. Lagrangian formulation

Some of the conclusions arrived at from the reduced Hamiltonian in the previous section are more transparent from the reduction of the Lagrangian. The process of the reduction of the action was described in Sec. II E.

The full action is

$$L(h_{\mu\nu}, \dot{h}_{\mu\nu}) = \int d^3x \, \mathscr{L}^m,$$

where $\mathscr{L}^m$ is given in (25). The constraint (36d)

$$\partial_j \pi^{ij} - m^2 h^{0i} + \kappa \Phi^{0i} = 0, \tag{71}$$

with π^{ij} given by (20) leads to

$$(\nabla^2 - m^2) h^{0i} - \partial^i \partial_j h^{0j} - \partial^0 \partial_j h^{ij} + \partial^{0i} h^j_j + \kappa \Phi^{0i} = 0.$$

The constraint (36e) with π^i_i found from (20) gives

$$\partial_i h^{0i} = \partial^0 h^i_i + \frac{\kappa}{m^2} \partial_\mu \Phi^{0\mu}.$$

Substituting in (71) gives,

$$h^{0i} = \frac{1}{m^2 - \nabla^2} \left(\partial_0 \partial_j h^{ij} + \kappa \Phi^{0i} + \frac{\kappa}{m^2} \partial^i \partial^\mu \Phi_{0\mu} \right).$$

This will be the only source of $[m^2]^{-1}$ in the fully reduced Lagrangian.

Before proceeding, we write down the partially reduced action, with h^{0i} removed, using the above, and h^{00} removed as the multiplier of the constraint ρ^0,

$$\int d^3x \Big\{ - \Big[\tfrac{1}{2} (\partial_\mu h_{ij} \partial^\mu h^{ij} + m^2 h \cdot h + \partial_\mu h^i_i \partial^\mu h^j_j + m^2 h^i_i h^j_j) + \partial_\mu \partial_i h^{ij} \frac{1}{\nabla^2 - m^2} \partial^\mu \partial^k h_{kj} + m^2 (\partial h) \cdot (\partial h) \Big]$$
$$+ \kappa \Big[h \cdot \Phi + h^i_i \Phi^{00} - 2\Phi_{0i} \frac{1}{\nabla^2 - m^2} \partial_0 \partial_j h^{ij} - \kappa^2 \Big[\Phi_{0i} \frac{1}{\nabla^2 - m^2} \Phi^{0i} - \partial_\mu \Phi^{\mu 0} \frac{1}{m^2 (\nabla^2 - m^2)} \partial_\nu \Phi^{\nu 0} \Big] \Big] \Big\}, \tag{72}$$

where, again, we have dropped the boundary terms. Note that the $[m^2]^{-1}$ dependence occurs only in the last term, which is independent of the fields and automatically gives the conservation condition $\partial_\mu \Phi^{\mu 0} = 0$ as a regularity requirement in the massless limit.

For the sake of exposition, we shall carry out the next stage of the reduction with a decomposition slightly different from that used in the Hamiltonian section. We write

$$h_{ij} = h^t_{ij} + h^{KX}_{ij},$$

where h^t_{ij} is the transverse part of h_{ij},

$$h^t_{ij} = h_{ij} - 2L^k_{(i} h_{j)k} + L_{ij} L \cdot h,$$

and is related to the previous decomposition by

$$h^t_{ij} = h^{TT}_{ij} + \tfrac{1}{2} T_{ij} T \cdot h.$$

The part h^{KX}_{ij} is longitudinal

$$h^{KX}_{ij} = \partial_{(i} X_{j)},$$

in which the vector X_i is given by

$$X_i = (\partial^j / \nabla^2)(h_{ij} - \tfrac{1}{2} L_{ij} L \cdot h).$$

Its transverse and longitudinal components are related to the previous decomposition by

$$2\partial_{(i} X^T_{j)} = h^{TL}_{ij}, \quad \text{and} \quad 2\, \partial_i X^i = L \cdot h.$$

For the independent physical degrees of freedom we use

$$Q_{ij} = h_{ij}^{TT} + h_{ij}^{KX},$$

expressed in terms of h_{ij}^{TT} and the vector, X_i. Then the ρ^0 constraint allows h_i^i to be removed via

$$h_i^i = [1/(\nabla^2 - m^2)](2\nabla^2\partial_i X^i - \kappa\Phi^{00}). \tag{73}$$

The fully reduced Lagrangian is, using a self-explanatory notation,

$$L(Q,\dot{Q}) = L^{TT} + L^{KX} + \text{"constants,"}$$

where

$$L^{TT} = -\int d^3x\Big\{\frac{1}{2}\partial_\mu Q_{ij}^{TT}\partial^\mu Q^{TTij} + m^2 Q^{TT}\cdot Q^{TT} - \kappa Q^{TT}\cdot\Phi^{TT}\Big\},$$

and

$$L^{KX} = -\int d^3x\Big\{m^2\Big[\partial_\mu X_i^T\frac{\nabla^2}{\nabla^2 - m^2}\partial^\mu X^{Ti} + m^2 X_i^T\frac{\nabla^2}{\nabla^2 - m^2}X^{Ti}\Big] + 3m^4\Big[\partial_\mu\partial_i X^i\frac{1}{\nabla^2 - m^2}\partial^\mu\partial^j X_j + m^2\partial_i X^i\Big(\frac{1}{\nabla^2 - m^2}\Big)^2\partial_j X^j\Big] - \kappa\Big[2\partial_\mu X_i\Phi^{\mu i} + 2\partial_0\partial_i X^i\frac{1}{\nabla^2 - m^2}\partial_\mu\Phi^{\mu 0} + m^2\Big(2\partial_0 X_i\frac{1}{\nabla^2 - m^2}\Phi^{0i} + \partial_i X^i\frac{1}{\nabla^2 - m^2}(\Phi_\mu^\mu - L\cdot\Phi) - 3\partial_0\partial_i X^i\frac{1}{\nabla^2 - m^2}\partial_0\Phi^{00}\Big)\Big]\Big\}.$$

Again, we collected boundary conditions and terms quadratic in the sources as constants. The kinetic term has been split into transverse and longitudinal components for ease of recognizing the contributions from the propagator and the vertex in the S matrix. Notice now that the remaining three conservation conditions $\partial_\mu\Phi^{\mu i} = 0$ follow immediately in the massless limit, once we realize that $\partial_\mu\Phi^{\mu 0} = 0$ is required, in the limit by the source terms which we have suppressed above. The coupling of the scalar degree of freedom to the sources will again give a nonzero contribution to the S matrix. As previously, covariant coupling to conserved sources can be studied, in the limit, in terms of the trace from (73) and its relation to $\tilde{h}$ in (65).

C. $\alpha \neq 1$ case

We end the massive spin-2 discussion by making some comments about an alternative mass term, one where $\alpha \neq 1$, which is interesting for several reasons. First, it is important because, with $\alpha \neq 1$, the discontinuity disappears. However, the theory is unphysical due to the presence of negative energies (ghosts). Second, in the case of $\alpha = \frac{1}{2}$ (which is included in $\alpha \neq 1$), the mass m has been thought of (loosely) as coming from a cosmological constant, for the Lagrangian (25) becomes the linearized form of the Einstein–Hilbert Lagrangian plus a cosmological constant, arbitrarily evaluated on a flat background. Finally, the structure of the constraints is sufficiently different to warrant a comparison with the $\alpha = 1$ case. We shall not attempt to give a complete analysis for $\alpha \neq 1$ since it *is* an unphysical theory, but we highlight some differences and similarities with the analysis given above.

For the Lagrangian we take (25), but now with $\alpha \neq 1$. The Hamiltonian (27) changes accordingly. The constraints are now

$$\pi^{00} \approx 0, \tag{74a}$$

$$\pi^{0i} \approx 0, \tag{74b}$$

$$-(\alpha m^2 - \nabla^2)h_i^i - \partial_{ij}h^{ij} - m^2(1-\alpha)h^{00} + \kappa\Phi^{00} \approx 0, \tag{74c}$$

$$\partial_j\pi^{ij} - m^2 h^{0i} + \kappa\Phi^{0i} \approx 0, \tag{74d}$$

with the multiplier conditions

$$m^2(1-\alpha)v_0 = \tfrac{1}{2}\alpha m^2\pi_i^i - m^2(1-2\alpha)\partial_i h^{0i} + \kappa\partial_\mu\Phi^{0\mu}, \tag{75a}$$

$$m^2 v^i = 2m^2(\partial_j h^{ij} - \alpha\partial_i h_\mu^\mu) - 2\kappa\partial_\mu\Phi^{\mu i}. \tag{75b}$$

At this point it may seem that there is a discontinuity in the $\alpha \to 1$ limit since v_0 becomes undetermined by (75a). As far as the constraints are concerned the $\alpha \to 1$ limit is in fact smooth; (74c) goes over to (36c); the v_0-multiplier condition becomes the *constraint* in (36e) and its evolution gives rise to the constraints (36e) and (36f). The remaining constraints and multiplier conditions are already the same. Note that, as in the spin-1 case, the *multiplier conditions* are the ones that in the $m \to 0$ limit precisely give rise to the statement of conservation, $\partial_\mu\Phi^{\mu\nu} = 0$.

All the eight constraints are second class. Therefore, the dimension of the reduced phase space is $[20 - 1(8)] = 12 = 6 + 6$, corresponding to six h's and six π's. Six is not the dimension of any irrep of the little group $SO(D-1) = SO(3)$. Thus we can not identify the field h as a pure spin-s field, but it can represent an admixture of spin-2 (five degrees of freedom) and a scalar.

In contrast to the $\alpha = 1$ case (where we did not have to remove h^{00} directly), here we have to use (74c) to remove h^{00} strongly. Here h^{0i} is removed by (74d), and $\pi^{0\mu}$ by (74a), (74b). So, we can identify *all* of h_{ij} with the six above-mentioned DDF's. This can also be seen from the structure of the DB's of this theory,

$$\{h_{ij}(x),\pi^{kl}(y)\}_D = \{h_{ij}(x),\pi^{kl}(y)\}_P = \delta_{(i}^k\delta_{j)}^l\,\delta^{(3)}(x-y), \tag{76}$$

in contrast to (B4).

This makes the situation exactly analogous to the spin-1 case [see (8)]. Even the subtleties of Sec. III B 3 are irrelevant here, since there is no constraint analogous to (36c). Given this similarity we can immediately see that again the TT part of h_{ij} goes over to a TT massless spin-2 field, and the remaining four degrees of freedom couple to the four conditions $\partial_\mu\Phi^{\mu\nu}$ and therefore decouple from conserved sources completely. Hence, for $\alpha \neq 1$, there is no discontinuity as $m \to 0$, but as the theory is plagued with a negative-energy scalar (which can be easily seen from the action; also see Ref. 3), we shall not pursue this investigation further. The structure of the DB in (76) is sufficiently simple for one to arrive at the conclusions we have given here without any explicit calculation.

ACKNOWLEDGMENTS

One of the authors (CM) would like to thank Charles Torre for teaching him the fundamentals and the intricacies of the Dirac formalism. We also acknowledge conversations with Paul Frampton and James York. We are especially grateful to Jack Ng and Yasuhiro Okada for numerous valuable discussions.

Financial support has been received from the National Science Foundation (BFW) and the Department of Energy (CM and HVD).

APPENDIX A: AN ORTHOGONAL DECOMPOSITION

In this section we will define our decomposition.

The identity element is orthogonally decomposed as,

$$\delta^i_j = T^j_i + L^j_i, \quad \text{i.e.,} \quad T^j_i L^k_j = 0.$$

Both T and L are idempotent,

$$T^j_i T^k_j = T^k_i, \quad L^j_i L^k_j = L^k_i.$$

Note the transverse property, $\partial_j T^j_i = 0$. A particular representation of these operators is

$$T^j_i = \delta^j_i - (1/\nabla^2)\partial_i\partial^j, \quad L^j_i = (1/\nabla^2)\partial_i\partial^j.$$

The ranks are

$$T^i_i \equiv T^j_i \delta^i_j = 2, \quad L^i_i \equiv L^j_i \delta^i_j = 1. \tag{A1}$$

A three-vector A_i is decomposed accordingly, i.e.,

$$A_i = \delta^j_i A_j = (T^j_i + L^j_i) A_j \equiv A^T_i + A^L_i,$$

where $A^T_i = T^j_i A_j$ and $A^L_i = L^j_i A_j$ are the transverse and longitudinal components of A_i. Equation (A1) implies that A^T_i has two, and A^L_i has one degree of freedom. The latter, being longitudinal, can be written as

$$A^L_i \equiv \partial_i A, \tag{A2}$$

where A is a scalar, representing the one degree of freedom in A^L_i. From (A2) we have $A = (1/\nabla^2)\partial_i A^i$.

We decompose a symmetric rank-2 three-tensor h_{ij} orthogonally as,

$$\begin{aligned} h_{ij} = \delta^k_{(i}\delta^l_{j)} h_{kl} &= [T^k_{(i} + L^k_{(i}][T^l_{j)} + L^l_{j)}] h_{kl} \\ &= h^{TT}_{ij} + h^{TL}_{ij} + h^{T2L}_{ij} + \tfrac{1}{3}\eta_{ij} h^k_k, \end{aligned}$$

where $h^{TT}_{ij} = (TT)^{kl}_{ij} h_{kl}$, etc. with

$$\begin{aligned} (TT)^{kl}_{ij} &= \tfrac{1}{2}(T^k_i T^l_j + T^l_i T^k_j - T_{ij} T^{kl}), \\ (TL)^{kl}_{ij} &= \tfrac{1}{2}(T^k_i L^l_j + T^l_i L^k_j + T^l_j L^k_i + T^k_j L^l_i), \\ (T2L)^{kl}_{ij} &= \tfrac{1}{6}(T_{ij} - 2L_{ij})(T^{kl} - 2L^{kl}); \end{aligned}$$

the ranks are 2, 2, and 1, respectively; and they are calculated as

$$(TT)^{kl}_{ij} \delta^i_{(k} \delta^j_{l)} = 2, \text{ etc.}$$

These ranks correspond to the degrees of freedom of h^{TT}, h^{TL}, and h^{T2L}, respectively. Of course, in this decomposition, the trace h^i_i carries the sixth degree of freedom.

The "identity operator" of (42), written as a matrix, decomposes orthogonally as

$$\mathbb{1} = (TT) + (TL) + (T2L),$$

which is why we chose the decomposition put forward in Sec. III B 2. This induces the following orthogonal decomposition for the q, p of (43) and (44):

$$\begin{aligned} q &= q^{TT} + q^{TL} + q^{T2L}, \\ p &= p^{TT} + p^{TL} + p^{T2L}, \end{aligned} \tag{A3}$$

where $q^{TT}_{ij} \equiv (TT)^{kl}_{ij} q_{kl}$, etc.

Here q^{TT}, q^{TL}, and q^{T2L} carry two, two, and one degrees of freedom, respectively, giving the five degrees of freedom of q. Similarly for p. In the main decomposition used in Sec. III, although $q^{TT} = h^{TT}$ and $q^{TL} = h^{TL}$, $q^{T2L} \neq h^{T2L}$. Similarly for p.

APPENDIX B: DIRAC BRACKETS

Here we shall construct the DB's of the $\alpha = 1$ massive spin-2 field. The procedure of constructing the DB's is as described in Sec. II B. Although, when one has many constraints, an iterative process can be used to construct the DB's, we find it more expeditious to compute the whole matrix of PB's of the constraints Δ and then invert it all at once. The details of the algebra are not important, except to note that a remarkable simplification occurs in the inversion of this matrix.

The matrix Δ is a matrix of differential operators acting on a Dirac delta function. Thus, in general, the equation

$$\int dz\, \Delta(x-z) \cdot \Delta^{-1}(z-y) = \mathbb{1}\delta(x-y), \tag{B1}$$

is a partial differential equation for the elements of the matrix Δ^{-1}, whose solution depends on appropriate boundary conditions. However, the wonderful feature of our Δ, given in (B2), is that all terms involving differential operators acting on elements of Δ^{-1} on the left-hand side of (B1) cancel, leaving us with an *algebraic equation* to solve for the elements of Δ^{-1} without any worry about the boundary conditions [this is why there does not appear any sign function in Δ^{-1} in (B3) as one might otherwise expect].

The constraints, listed in the order of (36), give for the matrix of their PB's

$$\Delta(x-y)=\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3m^2 \\ 0 & & & & & 1 & & & \partial_1 & 0 \\ 0 & & \mathbf{0} & & & & 1 & & \partial_2 & 0 \\ 0 & & & & & & & 1 & \partial_3 & 0 \\ 0 & & & & & -2\partial_1 & -2\partial_2 & -2\partial_3 & X & 0 \\ 0 & -1 & & & -2\partial_1 & & & & & 3m^2\partial_1 \\ 0 & & -1 & & -2\partial_2 & & \mathbf{0} & & & 3m^2\partial_2 \\ 0 & & & -1 & -2\partial_3 & & & & & 3m^2\partial_3 \\ 0 & \partial_1 & \partial_2 & \partial_3 & -X & & & & & \frac{9}{2}m^4 \\ 3m^2 & 0 & 0 & 0 & 0 & 3m^2\partial_1 & 3m^2\partial_2 & 3m^2\partial_3 & \frac{9}{2}m^4 & 0 \end{vmatrix}\left(\frac{-1}{2}\right)m^2\delta^{(3)}(x-y). \tag{B2}$$

The notation is $\partial_1\delta^{(3)}(x-y)=(\partial/\partial x)\delta^{(3)}(x-y)$, etc. Here $X\equiv 3m^2-2\nabla^2$, and all vacant slots are zero. Note that all ten constraints are second class in nature. And, according to (B1),

$$\Delta^{-1}(x-y)=\begin{vmatrix} 0 & -2\nabla^2\partial_1 & -2\nabla^2\partial_2 & -2\nabla^2\partial_3 & -Y & 0 & 0 & 0 & 0 & -1 \\ -2\nabla^2\partial_1 & & & & & Z_1 & -2\partial_{12} & -2\partial_{13} & -2\partial_1 & 0 \\ -2\nabla^2\partial_2 & & \mathbf{0} & & & -2\partial_{12} & Z_2 & -2\partial_{23} & -2\partial_2 & 0 \\ -2\nabla^2\partial_3 & & & & & -2\partial_{13} & -2\partial_{23} & Z_3 & -2\partial_3 & 0 \\ Y & & & & & \partial_1 & \partial_2 & \partial_3 & 1 & 0 \\ 0 & -Z_1 & 2\partial_{12} & 2\partial_{13} & \partial_1 & & & & & 0 \\ 0 & 2\partial_{12} & -Z_2 & 2\partial_{23} & \partial_2 & & & \mathbf{0} & & 0 \\ 0 & 2\partial_{13} & 2\partial_{23} & -Z_3 & \partial_3 & & & & & 0 \\ 0 & -2\partial_1 & -2\partial_2 & -2\partial_3 & -1 & & & & & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$\times\left(\frac{2}{3m^4}\right)\delta^{(3)}(x-y). \tag{B3}$$

The notation is

$$\partial_{ij}\equiv\partial_i\partial_j,\quad Z_i\equiv 3m^2-2\partial_i\partial_i\ \ \text{(no sum over } i\text{)}$$

and $Y\equiv\nabla^2+\frac{3}{2}m^2$. Using the definition of the DB (Sec. II B) we get

$$\begin{aligned}&\{h_{ij}(x),\pi^{kl}(y)\}_D\\ &=\{h_{ij}(x),\pi^{kl}(y)\}-(2/3m^4)\\ &\quad\times[\partial_i\partial_j+\tfrac{1}{2}m^2\eta_{ij}]\times[\partial^k\partial^l+(m^2-\nabla^2)\eta^{kl}]\\ &\quad\times\delta^{(3)}(x-y),\end{aligned} \tag{B4}$$

$$\begin{aligned}\{h_{ij}(x),h_{0k}(y)\}_D=(2/3m^4)[&\partial_i\partial_j\partial_k\\ &-(3m^2/4)(\eta_{ik}\partial_j+\eta_{jk}\partial_i)\\ &+\tfrac{1}{2}m^2\eta_{ij}\partial_k]\delta^{(3)}(x-y).\end{aligned} \tag{B5}$$

All other nonzero DB's can be derived from these by the use of the constraints (now treated as strong equations). For example,

$$\begin{aligned}\{h^0_0(x),\pi^{kl}(y)\}_D&=\{(-h^i_i+\text{sources}),\pi^{kl}\}_D\\ &=-\eta^{ij}\{h_{ij}(x),\pi^{kl}(y)\}_D.\end{aligned}$$

[1]H. Van Dam and M. Veltman, Nucl. Phys. B **22**, 297 (1970); V. I. Zaharov, JETP Lett. **22**, 312 (1970); H. Van Dam and M. Veltman, Gen. Relativ. Gravit. **3**, 215 (1973).

[2]N. Dombey, J. Phys. A: Math Gen. **9**, 1375 (1976); N. Dombey and C. E. Vayonakis, Math Gen. **9**, 1381 (1976).

[3]D. G. Boulware and S. Deser, Phys. Rev. D **6**, 3368 (1972).

[4]A. Higuchi, Nucl. Phys. B **282**, 397 (1987).

[5]S. Deser and R. I. Nepomeckie, Phys. Lett. B **132**, 321 (1983); Ann. Phys. (New York) **154**, 396 (1984).

[6]C. Fronsdal, Nucl. Phys. B **167**, 237 (1980).

[7]L. H. Ford and H. van Dam, Nucl. Phys. B **169**, 126 (1980).

[8]S. Deser, J. H. Kay, and K. S. Stelle, Phys. Rev. D **16**, 2448 (1977).

[9]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Belfer Graduate Student School of Science, Yeshiva University, NY, 1964); K. Sundermeyer, *Constrained Dynamics, Lecture Notes in Physics*, Vol. 169 (Springer, Berlin, 1982); E. C. G. Sudarshan and N. Mukunda, *Classical Mechanics in a Modern Perspective* (Wiley, New York, 1978); P. G. Bergmann, Phys. Rev. **75**, 680 (1949); Bergmann's contribution was begun in this work, and was further expounded in subsequent papers with several additional authors, for references to which, see J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951).

[10]D. G. Boulware and W. Gilbert, Phys. Rev. **126**, 1563 (1962); S. Deser, Phys. Rev. D **21**, 2436 (1980).

[11]J. S. Schwinger, *Particles, Sources, and Fields* (Addison-Wesley, Reading, MA, 1970); C. Itzykson and J-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).

[12]M. Fierz, Helv. Phys. Acta **12**, 3 (1939).

On the many-time formulation of classical particle dynamics

G. Longhi
Dipartimento di Fisica, Universitá di Firenze, Firenze, Italy

L. Lusanna
Istituto Nazionale di Fisica Nucleare, Sezione di Firenze, Firenze, Italy

J. M. Pons
Departament de Fisica Teorica, Universitat de Barcelona, Barcelona, Spain

(Received 14 January 1988; accepted for publication 15 February 1989)

Starting from the standard one-time dynamics of n nonrelativistic particles, the n-time equations of motion are inferred, and a variational principle is formulated. A suitable generalization of the classical Lie–König theorem is demonstrated, which allows the determination of all the associated presymplectic structures. The conditions under which the action of an invariance group is canonical are studied, and a corresponding Noether theorem is deduced. A formulation of the theory in terms of n first-class constraints is recovered by means of coisotropic imbeddings. The proposed approach also provides for a better understanding of the relativistic particle dynamics, since it shows that the different roles of the physical positions and the canonical variables is not peculiar to special relativity, but rather to any n-time approach: indeed a nonrelativistic no-interaction theorem is deduced.

I. INTRODUCTION

As it is well known, in the treatment of bound states in the framework of quantum field theory, both relativistic and nonrelativistic, the states of the bound system are described by a wave function for many particles, which will depend in a natural way on the times of each elementary field. In order to give a physical interpretation to this wave function, we must give a meaning to this many time description, or, what is the same, to have a consistent dynamical theory for systems of n particles, with n different times.

At the relativistic level the bound states are described by the Bethe–Salpeter equation,[1] with the connected problems of the normalization and interpretation of its solutions. To get a better understanding of it, Todorov[2] and then Komar[3] developed a manifestly covariant classical relativistic model for two particles, of an action-at-a-distance kind,[4] which describes in a covariant way the instantaneous approximations of the Bethe–Salpeter equation, restricted to the two particle sector.[5] In the Todorov–Komar model the dynamics is given in terms of two first-class constraints, and, therefore, the relative time problem is related to the existence of gauge transformations generated by the constraints.[6] An equivalent model was discovered by Droz–Vincent,[7] which was based on a two-time formulation of the classical relativistic dynamics. Here we have the first example of the connection between the first-class constraints formulation and the many-time dynamics.

The Todorov–Komar–Droz–Vincent model for two particles, in its first quantized version,[2,8–10] generates a bilocal wave function, which is a solution of two coupled integrable integro-differential wave equations. In Ref. 11 a complete analysis of these equations has been done, by giving the set of their solutions (where the relative time appears in a phase factor), the explicit expression of the Poincaré invariant scalar products (see also Ref. 9 and 10 and the last paper quoted in Ref. 8), and, by solving the initial data problem, a probabilistic interpretation of the wave function is proposed. The connection with the Bethe–Salpeter equation is studied in Ref. 5, and, for the nonrelativistic limit, in Ref. 12. Attempts toward a second quantization along these lines are given in Ref. 13.

However, due to the complications introduced by special relativity, a clear understanding of all the involved structures, and a clear physical interpretation of them, is still lacking. One of these complications is for instance the problem of the most suitable definition of the relativistic position coordinates, see Ref. 14. The Todorov–Komar–Droz–Vincent model is the output of the many efforts to formulate the theory of the classical relativistic particle mechanics (see Refs. 4, 12, and 15 for reviews, and also Ref. 4. See Ref. 16 for reviews on the predictive mechanics, and see also Refs. 7 and 17), overcoming the difficulties introduced by the no-interaction theorem,[18] which prevents the physical coordinates from being simultaneously covariant and canonical in the interacting case, in any of the forms of the dynamics introduced by Dirac.[19] From here it emerges the dualism between the physical covariant coordinates $\{q^\mu\}$ and the phase-space canonical coordinates $\{x^\mu\}$. The models which use the first-class constraints approach are expressed in terms of the coordinates $\{x^\mu\}$, in order to avoid the consequences of the no-interaction theorem, while the model formulated in the predictive approach are expressed in terms of the coordinates $\{q^\mu\}$. The work of Droz–Vincent, in particular Ref. 20, establishes a bridge between the two approaches, and provides a connection between the above mentioned dualism and the many times formulation.

The present paper was originated by the wish to clarify these problems, avoiding the complications due to special relativity. We start from the classical nonrelativistic Newton's equations for n particles, as a preliminary laboratory, deferring the quantum aspects as well as the physical interpretation, and the interpretation as a gauge theory to a future paper. The first step will be to get an n-time version of New-

 0022-2488/89/081893-20$02.50

ton's equations of motion, which will be the nonrelativistic counterpart of the relativistic predictive equations, then we will put them in a first-order form.

In order to gain a canonical formulation, suitable for the quantization, we will give a generalization to n-times (nonautonomous case) of the classical Lie–König theorem[21] for which the reader is also referred to Refs. 22 and 23. In this way we will find all the n-time local (symplectic) structures, or better, the Poisson structures, which can be associated to the given equations of motion, and we will immediately find the dualism between the physical position $q^i(t^i)$ of the ith particle and its canonical coordinate $x^i(t^1,t^2,...,t^n)$. While the former only depends on its own time, the latter depends on all the t^i. Moreover we will get the n-time generalization of the inverse problem in the calculus of variations, in the first-order formalism[24] (see Ref. 25 for a review), and, as a by-product, we will get a nonrelativistic no-interaction theorem, and it will be possible to demonstrate the nonexistence of a predictive Lagrangian, independent on the accelerations, in the interacting case; only a singular Lagrangian can be defined.

The study of the invariance transformations of the n-time Newton equations in the first-order form will provide for an n-time generalization of the Currie–Hill conditions,[26] for the Galilei algebra, as well as of the first Noether theorem, and for the conditions on the invariance transformations for being canonically implementable, with the chosen symplectic structure.

The final step will be to define an enlarged phase space $\overline{M}$, with the new canonical variables t^i and ϵ_i, times and energies, respectively, with n first-class constraints. It will be shown that this is a coisotropic embedding[27] in the phase space $\overline{M}$ of the original presymplectic manifold.[28] (See, also Ref. 29 for a set of first-class constraints describing n nonrelativistic free particles, and Refs. 12 and 30 for the case of two nonrelativistic interacting particles.) Finally, the Droz–Vincent method[20] will allow the recovery of the physical position coordinates from the canonical ones.

It is our hope that, at the end of this paper, it will be clear that many features of the n-time approach are not peculiar of a relativistic theory, but, rather, they simply are more complicated in the relativistic case, with the result of hiding their basic simplicity. To reveal this simplicity we need a reformulation of Newton's equation of motion, which is probably useless for the applications, but it is inescapable for the present kind of problems.

As a matter of fact, the present analysis is quite general, and it could in principle be applied to any dynamical system. It is only necessary to specify the kinematical group of the theory, that is, for instance, the Poincaré group instead of the Galilei group. The only difference, which in practice becomes a real difficulty, is that in the present analysis the constraints are energy constraints, that is they are linear in the energies of the particles. This means that, in a relativistic theory, where the constraints are usually given in a covariant form, we should solve them in terms of the energies, and, in general, several local solutions will be possible. The present analysis must be separately applied to each of these solutions.

In Sec. II we will discuss the n-time approach to the equations of motion. We will explicitly develop a very simple model for two particles, for which we will give the explicit expression of the two-time forces in Appendix A.

In the same section we will discuss the canonical formulation of the dynamics.

In Sec. III the generalization to many independent variables of the classical Lie–Köning theorem will be outlined in order to establish, on general grounds, the existence of a canonical formulation. Some of the details are given in Appendix B. In Sec. IV we will discuss the invariances of the theory.

Finally, in Sec. V, the problem of the position coordinates and of their correlation with the canonical coordinates will be discussed.

Some of the present material was already presented in Refs. 30 and 31.

II. THE n-TIME FORMULATIOIN OF THE DYNAMICS

Given the equations of motion of a system of particles in a nonrelativistic theory

$$m_i\ddot{\mathbf{q}}^i = \mathbf{F}^i(t,\mathbf{q}^j,\mathbf{v}^j), \quad i=1,2,...,n, \tag{2.1}$$

where $\mathbf{q}^i$ are the positions of the particles and t the time in a given inertial reference frame, it is always possible, in principle, to get an n-time formulation by eliminating the integration constants from their solutions and their first derivatives. Let us write the general solution of the system (2.1),

$$\mathbf{q}^i = \mathbf{g}^i(t,c_1,c_2,...,c_{6n}), \tag{2.2}$$

where $c_1,c_2,...c_{6n}$ are integration constants. In solution (2.2) we may choose a different time for each particle, that is

$$\mathbf{q}^i(t^i) = \mathbf{g}^i(t^i,c_1,c_2,...,c_{6n}). \tag{2.3}$$

If we eliminate the integration constants $c_1,c_2,...,c_{6n}$ from Eqs. (2.3) and their derivatives

$$\mathbf{v}^i(t^i) = \frac{d}{dt^i}\mathbf{q}^i(t^i) = \dot{\mathbf{g}}^i(t^i,c_1,c_2,...,c_{6n}), \tag{2.4}$$

and substitute in

$$\mathbf{a}^i(t^i) = \frac{d^2}{dt^{i2}}\mathbf{q}^i(t^i) = \ddot{\mathbf{g}}^i(t^i,c_1,c_2,...,c_{6n}), \tag{2.5}$$

we get the n-time equations of motion

$$m_i\mathbf{a}^i(t^i) = \mathscr{F}^i(t^j,\mathbf{q}^j,\mathbf{v}^j). \tag{2.6}$$

Since the ith lhs only depends on t^i, we must have

$$\frac{d}{dt^j}\mathscr{F}^i = \left(\frac{\partial}{\partial t^j} + \mathbf{v}^j\cdot\frac{\partial}{\partial \mathbf{q}^j} + \mathbf{a}^j\cdot\frac{\partial}{\partial \mathbf{v}^j}\right)\mathscr{F}^i$$
$$= \left(\frac{\partial}{\partial t^j} + \mathbf{v}^j\cdot\frac{\partial}{\partial \mathbf{q}^j} + \frac{\mathscr{F}^j}{m_j}\cdot\frac{\partial}{\partial \mathbf{v}^j}\right)\mathscr{F}^i = 0, \tag{2.7}$$

for $i\neq j$.

Equations (2.7) can be called predictivity conditions, and the n-times forces $\mathscr{F}^i$ predictive forces. This in order to agree with the literature on predictive mechanics quoted in the Introduction (see Refs. 7, 16, and 17). Putting $t^1=t^2=\cdots=t^n=t$ in Eq. (2.6), we must have

$$\mathscr{F}^i(t,\mathbf{q}^j,\mathbf{v}^j) = \mathbf{F}^i(t,\mathbf{q}^j,\mathbf{v}^j). \tag{2.8}$$

Let us now consider the system (2.6): It seems difficult to get it from an action principle, or, more simply, to get the forces $\mathscr{F}^i$ from some potential, or even if possible, as is apparent from the example of the Appendix A, it will be very complicated. As a matter of fact, we will show in Sec. V that a second-order Lagrangian for the equations of motion (2.6) does not exist. So we will look for a possible canonical formulation in terms of other variables.

In Sec. III we will give a generalization to many variables t^i ($i = 1,2,...,n$) of the classical Lie–König theorem, which asserts that, for any given set of first-order ordinary differential equations, it is always possible to find new variables, $\mathbf{x}^i$ and $\mathbf{p}^i$ (in place of the positions and velocities), and a function of them, $\mathbf{H}$, such that the set of equations is transformed to canonical form.

In our case the generalization of this theorem says that, given the set of first-order equations (now partial derivative equations)

$$\frac{\partial \mathbf{q}^i}{\partial t^j} = \mathbf{v}^i \delta^i_j \,, \qquad m_i \frac{\partial \mathbf{v}^i}{\partial t^j} = \delta^i_j \mathscr{F}^i(t^k, \mathbf{q}^k, \mathbf{v}^k) \,, \tag{2.9}$$

it is always possible to find new variables $\mathbf{x}^i = \mathbf{x}^i(t^k,\mathbf{q}^k,\mathbf{v}^k)$ and $\mathbf{p}_i = \mathbf{p}_i(t^k,\mathbf{q}^k,\mathbf{v}^k)$, and besides n functions $H_i = H_i(t^k,\mathbf{x}^k,\mathbf{p}_k)$ satisfying the integrability conditions

$$\frac{\partial H_i}{\partial t^j} - \frac{\partial H_j}{\partial t^i} + \{H_i,H_j\} = 0 \,, \tag{2.10}$$

where

$$\{A,B\} = \sum_{i=1}^{n} \left(\frac{\partial A}{\partial \mathbf{x}^i} \cdot \frac{\partial B}{\partial \mathbf{p}_i} - \frac{\partial B}{\partial \mathbf{x}^i} \cdot \frac{\partial A}{\partial \mathbf{p}_i} \right), \tag{2.11}$$

such that the system (2.9) can be written in canonical form:

$$\frac{\partial \mathbf{x}^i}{\partial t^j} = \{\mathbf{x}^i, H_j\} \,, \qquad \frac{\partial \mathbf{p}_i}{\partial t^j} = \{\mathbf{p}_i, H_j\} \,. \tag{2.12}$$

We leave to the next section the demonstration of this result. As a matter of fact, this theorem asserts the existence of at least one canonical formulation, but says very little for actual construction of the Hamiltonians H_i.

For the moment, we will assume the existence of n functions H_i, satisfying the conditions (2.10), in order to formulate the action principle in canonical form.

Let us consider the following line integral:

$$S = \int_{a(l)}^{b} (\mathbf{p}_i \cdot d\mathbf{x}^i - H_i \, dt^i) \,, \tag{2.13}$$

where a and b are two points in the $(t^1,t^2,...,t^n)$ space and l is a path connecting them.

S is a functional of the $\mathbf{p}_i$ and $\mathbf{x}^i$ defined on the $(t^1,t^2,...,t^n)$ space, for which the usual symplectic structure is assumed

$$\{\mathbf{x}^{im}, \mathbf{p}_j{}^n\} = \delta^{mn}\delta^i_j \,, \tag{2.14}$$

where $m,n = 1,2,3$ and $i,j = 1,2,...,n$, and the $H_i(t^k,\mathbf{x}^k,\mathbf{p}_k)$ satisfy the conditions (2.10).

If we require $\delta S = 0$ for an arbitrary variation of $\mathbf{p}_i = \mathbf{p}_i(t^1,t^2,...,t^n)$ and $\mathbf{x}^i = \mathbf{x}^i(t^1,t^2,...,t^b)$, which vanishes in a and b, and for *any* choice of the path l, we get the following equations of motion:

$$d\mathbf{x}^i = \{\mathbf{x}^i, H_j\} \, dt^j \,, \qquad d\mathbf{p}_i = \{\mathbf{p}_i, H_j\} \, dt^j \tag{2.15}$$

(a sum over repeated indices is assumed), which are integrable in view of the conditions (2.10).

Equations (2.15) can be obtained on any given path l, in which case $dt^j = t'^j(\tau)d\tau$, where τ is any parameter for the path l, and the functions $t'^j(\tau)$ depend on the choice of l. Since Eqs. (2.15) are integrable, the canonical coordinates $\mathbf{x}^i$ and $\mathbf{p}_i$ exist as functions on the space $(t^1,t^2,...,t^n)$ and not only on the path l. So they will coincide with the particular solution found on a given l, when the independent variables t^i are restricted on l.

When $\mathbf{x}^i$ and $\mathbf{p}_i$ are solutions of the equations of motion (2.15), the canonical action S does not depend on l. Indeed the one-form

$$\theta = \mathbf{p}_i \cdot \mathbf{x}^i - H_i \, dt^i \,, \tag{2.16}$$

is closed, when Eqs. (2.15) hold

$$d\theta = \frac{1}{2}\left(\frac{\partial H_i}{\partial t^j} - \frac{\partial H_j}{\partial t^i} + \{H_i,H_j\} \right) dt^i \wedge dt^j = 0 \,.$$

For any dynamical variable $\xi = \xi(t^k,\mathbf{x}^k,\mathbf{p}_k)$, the equation of motion reads

$$d\xi = \left(\frac{\partial \xi}{\partial t^i} + \{\xi, H_i\} \right) dt^i \,. \tag{2.17}$$

Let us notice that in general we have that

$$\mathbf{x}^i = \mathbf{x}^i(t^1,t^2,...,t^n) \,,$$

and

$$\mathbf{p}_i = \mathbf{p}_i(t^1,t^2,...,t^n) \,,$$

whereas the positions $\mathbf{q}^i$ are only functions of the corresponding t^i. The only case in which the $\mathbf{x}^i$ only depend on t^i happens when the H_i, for $i \neq j$, do not depend on the $\mathbf{p}_i$, for instance when the ith particle is free.

This fact already shows that in general the canonical variables $\mathbf{x}^i$ cannot coincide with the positions $\mathbf{q}^i$. Only when the interaction between the particles vanishes, we can freely make the choice $\mathbf{x}^i = \mathbf{q}^i$.

The fact that the canonical variables $\mathbf{x}^i$ and the physical positions $\mathbf{q}^i$ do not coincide, apart from the free case, is common in the relativistic theories, where this fact is essentially a consequence of the no-interaction theorem. But we now see that it is not peculiar of relativity, but rather of a multitime approach.

We will discuss the relation between the positions $\mathbf{q}^i$ and the canonical variables in Sec. V, where we will give the precise connection between the canonical formalism and the equations of motion (2.6).

As previously mentioned, the next section is devoted to the proof of the existence of a canonical formulation, once Eqs. (2.6) are given, that is of a generalization of the Lie–König theorem.

III. A GENERALIZATION OF THE LIE–KÖNIG THEOREM

The Lie–König theorem[21] was discussed and applied in connection with the relativistic dynamics of systems of particles by Hill[22] and Kerner.[32]

Here we will essentially follow the treatment given in Refs. 22 in order to look for a generalization of this theorem to many independent variables $\{t^i\}$, $i = 1,2,...,n$.

The equations of motion (2.6) can be written as a first-order system as in Eqs. (2.9):

$$\frac{\partial \mathbf{q}^i}{\partial t^j} = \mathbf{v}^i \delta^i_j ,$$
$$\frac{\partial \mathbf{v}^i}{\partial t^j} = \frac{1}{m_i} \delta^i_j \mathcal{F}^i(t^k,\mathbf{q}^k,\mathbf{v}^k) , \tag{2.9'}$$

where $i,j = 1,2,...,n$.

The conditions (2.7) will now be written in the shorthand notation

$$Y_i \mathcal{F}^j = 0 , \quad \text{for } i \neq j , \tag{3.1}$$

where

$$Y_i = \frac{\partial}{\partial t^i} + \mathbf{v}^i \cdot \frac{\partial}{\partial \mathbf{q}^i} + \frac{\mathcal{F}^i}{m_i} \cdot \frac{\partial}{\partial \mathbf{v}^i} . \tag{3.2}$$

Let us introduce a new notation,

$$\{y^a\} = \{\mathbf{q}^i\} , \quad \text{for } a = 1,2,...,N = 3n ,$$
$$\{y^a\} = \{\mathbf{v}^i\} , \quad \text{for } a = N+1, N+2,...,2N . \tag{3.3}$$

With this notation the equations of motion (2.9) can be written

$$\frac{\partial y^a}{\partial t^i} = h^a_i(y,t) , \tag{3.4}$$

where

$$\{h^a_i\} = \{\mathbf{h}^j_i\} , \quad \mathbf{h}^j_i = \delta^j_i \mathbf{v}^i , \quad \text{for } a = 1,2,...,N ,$$
$$\{h^a_i\} = \{\mathbf{h}^{n+j}_i\} , \quad \mathbf{h}^{n+j}_i = \delta^j_i (1/m_i) \mathcal{F}^i ,$$
$$\text{for } a = N+1,...,2N . \tag{3.5}$$

The conditions (3.1) become

$$Y_i h^a_j = 0 , \quad \text{for } i \neq j , \tag{3.6}$$

where

$$Y_i = \frac{\partial}{\partial t^i} + h^a_i \frac{\partial}{\partial y^a} . \tag{3.7}$$

The conditions (3.6) are a particular case of the integrability conditions for the system (3.4), which is a Mayer system[33]; the more general integrability conditions are

$$Y_i h^a_j = Y_j h^a_i . \tag{3.8}$$

From now on, if not otherwise specified, we will assume the more general conditions (3.8) in place of (3.6).

Let us now look for a variational principle giving the set of equations of motion (3.4) in the form

$$\delta S = 0 , \tag{3.9}$$

where

$$S = \int_{a(l)}^{b} \theta , \quad \theta = U_a \, dy^a - V_i \, dt^i , \tag{3.10}$$

and where l is a path in the space of the times $T = \{t^i\}$, $i = 1,2,...,n$, connecting $a = \{t^i_{(a)}\}$ to $b = \{t^i_{(b)}\}$.

The variational principle (3.9) will be required for all the variations of the functions

$$y^a(t^1,t^2,...,t^n) ,$$

which vanish at the end points of the path l, a and b, and *for all the variations* of the path l with fixed endpoints.

At the end points a and b the functions $\mathbf{q}^i(t^1,t^2,...,t^n)$ will be considered as given, and the functions $\mathbf{v}^i(t^1,t^2,...,t^n)$ fixed, but undetermined. It is worth recalling that, when the more restricted integrability conditions (3.6) will hold, the variables $\mathbf{q}^i$ will only depend on their own time t^i.

A variation of $y^a(t^1,t^2,...,t^n)$ *and of* l amounts in performing independent variations of y^a and t^i.

If we introduce another notation

$$\{y^\alpha\} = \{y^a, t^i\} , \quad \alpha = 1,2,...,2N+n , \tag{3.11}$$

and put

$$U_\alpha = U_a , \quad \text{for } \alpha = 1,2,...,2N ,$$
$$U_\alpha = -V_i , \quad \text{for } \alpha = 2N+i, \ i = 1,2,...,n , \tag{3.12}$$

the one-form θ can be written

$$\theta = U_\alpha \, dy^\alpha . \tag{3.13}$$

The variation of this form gives rise to

$$\delta\theta = \left(\frac{\partial U_\alpha}{\partial y^\beta} - \frac{\partial U_\beta}{\partial y^\alpha}\right) dy^\alpha \, \delta y^\beta + d(U_\alpha \, \delta y^\alpha) ,$$

so the equations of motion corresponding to $\delta S = 0$ will be

$$\Gamma_{\alpha\beta} \, dy^\alpha = 0 , \tag{3.14}$$

where we have defined

$$\Gamma_{\alpha\beta} = \frac{\partial U_\alpha}{\partial y^\beta} - \frac{\partial U_\beta}{\partial y^\alpha} . \tag{3.15}$$

Equations (3.14) must be identified with Eqs. (3.4). Since Eqs. (3.14), when explicitly written, are

$$\Gamma_{ab} \, dy^b + \Gamma_{ai} \, dt^i = 0 ,$$
$$\Gamma_{ia} \, dy^a + \Gamma_{ij} \, dt^j = 0 , \tag{3.16}$$

this requires

$$|\Gamma_{ab}| \neq 0 \quad (a,b = 1,2,...,2N) , \tag{3.17}$$

and

$$\Gamma_{ab} h^b_i + \Gamma_{ai} = 0 ,$$
$$\Gamma_{ia} h^a_j + \Gamma_{ij} = 0 , \tag{3.18}$$

or

$$\Gamma_{ab} h^b_i = -\Gamma_{ai} ,$$
$$\Gamma_{ab} h^a_i h^b_j = \Gamma_{ij} . \tag{3.18'}$$

The first of these equations gives the "forces" h^a_i in terms of Γ,

$$h^a_i = -(\Gamma^{-1})^{ab} \Gamma_{bi} . \tag{3.19}$$

The condition (3.17) implies that the two-form

$$\omega = d\theta = -\tfrac{1}{2} \Gamma_{\alpha\beta} \, dy^\alpha \wedge dy^\beta \tag{3.20}$$

has rank $2N$, because ω can also be written as

$$\omega = -\tfrac{1}{2} \Gamma_{ab} \theta^a \wedge \theta^b , \tag{3.21}$$

where

$$\theta^a = dy^a - h_i^a\, dt^i. \tag{3.22}$$

The system $\theta^a = 0$, which is the set of equations of motion (3.4), is completely integrable due to the integrability conditions (3.8). This can also be verified with the Frobenius theorem, since

$$d\theta^a = \left(\frac{\partial h_i^a}{\partial y^b}\, dt^i\right) \wedge \theta^b, \tag{3.23}$$

where Eq. (3.8) was used.

Since

$$i_{Y_i}\theta^a = 0, \tag{3.24}$$

where i_{Y_i} is the contraction with respect the vector field Y_i (see for instance Ref. 34) the dual form of the Eq. (3.23) is in terms of the characteristic vector fields Y_i, that is

$$[Y_i, Y_j] = 0. \tag{3.25}$$

The fields Y_i span the kernel distribution of ω since

$$i_{Y_i}\omega = 0, \tag{3.26}$$

so the kernel distribution of ω is n-dimensional, the Y_i being linearly independent.

From Eq. (3.21), we have that, on the motion,

$$\omega = d\theta = 0.$$

This has the important consequence that the action S, when evaluated on the motion, *does not depend on the path* l, but only on the end points a and b. Indeed, using Stokes' theorem, we have that for two paths l and l' with the same end points

$$\left(\int_{a(l')}^{b} - \int_{a(l)}^{b}\right)\theta = \int_{\partial D}\theta = \int_D d\theta = 0, \tag{3.27}$$

where D is the domain, in the space of times $T = \{t^i\}$, bounded by l' and l.

In the notation (3.11), by setting

$$\begin{aligned} h_i^\alpha &= h_i^a, \quad \text{for } \alpha = a, \\ h_i^\alpha &= \delta_i^j, \quad \text{for } \alpha = 2N + j, \end{aligned} \tag{3.28}$$

we may write

$$Y_i = h_i^\alpha \frac{\partial}{\partial y^\alpha} \tag{3.29}$$

and the integrability conditions (3.8) can now be written

$$Y_i h_j^\alpha = Y_j h_i^\alpha, \tag{3.30}$$

since, for $\alpha = 2N + 1, \ldots, 2N + n$, they become identities. In the case of the restricted integrability conditions (3.1), the right-hand side of Eq. (3.30) is zero. Equations (3.4) can be written

$$dy^\alpha = h_i^\alpha\, dt^i, \tag{3.31}$$

and the equations for $\Gamma_{\alpha\beta}$ can be summarized as follows:

$$\Gamma_{\alpha\beta} = -\Gamma_{\beta\alpha}, \tag{3.32}$$

$$\frac{\partial \Gamma_{\alpha\beta}}{\partial y^\gamma} + \text{cyclic} = 0, \tag{3.33}$$

$$\Gamma_{\alpha\beta} h_i^\beta = 0, \tag{3.34}$$

where the first and the second equations are a consequence of Eq. (3.15), and the last collects Eqs. (3.18).

Together with these equations we must remember the requirement

$$|\Gamma_{ab}| \neq 0, \quad \text{for } a,b = 1,2,\ldots,2N. \tag{3.35}$$

Equations (3.32) and (3.33) are merely identities: this follows from the definition of $\Gamma_{\alpha\beta}$ in Eq. (3.16). Nevertheless they can be assumed as defining equations for $\Gamma_{\alpha\beta}$. Observe that from these equations and Eq. (3.34), and using the integrability conditions (3.8), the components Γ_{ai} and Γ_{ij} can be eliminated, obtaining two equations for Γ_{ab} solely

$$Y_i \Gamma_{bc} = (\Gamma_{ab}\partial_c - \Gamma_{ac}\partial_b) h_i^a, \tag{3.36}$$

$$\frac{\partial \Gamma_{ab}}{\partial y^c} + \text{cyclic} = 0. \tag{3.37}$$

Equations (3.33), when written explicitly for the components Γ_{ab}, Γ_{ai}, and Γ_{ij}, reveal the self-adjoint nature of Eqs. (3.14). Indeed these equations split in four sets of equations, of which two are Eqs. (3.37) and

$$\frac{\partial \Gamma_{ab}}{\partial t^i} + \frac{\partial \Gamma_{ia}}{\partial y^b} + \frac{\partial \Gamma_{bi}}{\partial y^a} = 0, \tag{3.38}$$

while the two other sets are identically satisfied, due to Eqs. (3.36) and (3.8).

The reader is referred to Ref. 35 for the definition of self-adjoint systems in the one-time case.

Let us notice that, when the restricted form (3.6) of the integrability conditions holds, to each solution Γ_{ab} of Eqs. (3.36) and (3.37), such that its restriction at equal times has the form

$$\|\Gamma_{ab}|_{t^1, t^2, \ldots, t^n = t}\| = \begin{pmatrix} \beta & -\alpha \\ \alpha & 0 \end{pmatrix}, \tag{3.39}$$

where α and β are $N \times N$ matrices, and $\alpha = \alpha^T$, $\beta = -\beta^T$, and where the zero sector refers to

$$\frac{\partial U^m_{i+n}}{\partial v^{kj}} - \frac{\partial U^k_{j+n}}{\partial v^{mi}} \quad (m,k = 1,2,3),$$

a solution of the inverse problem of the calculus of variations is associated, that is there exists a Lagrangian for Eqs. (2.1), and α and β can be expressed as

$$\alpha_{im,jk} = \frac{\partial^2 L}{\partial \dot q^{im}\, \partial \dot q^{jk}},$$

$$\beta_{im,jk} = \frac{\partial^2 L}{\partial q^{im}\, \partial \dot q^{jk}} - \frac{\partial^2 L}{\partial q^{ik}\, \partial \dot q^{im}},$$

where $\dot q^{im} = v^{im}$.

As shown in Ref. 24 this condition is equivalent to the Helmholtz conditions in the Douglas form, as are expressed in the first-order formalism (see Ref. 36). The same results are also given in Ref. 37, and, in a more geometric way, in Ref. 38 by using a variant of the one-forms (3.22). As there are inequivalent classes of solutions for Γ_{ab} (see in the following of this section), the system (2.1) will have no one, one or several inequivalent (or s-equivalent) Lagrangians, according to how many solutions Γ_{ab} admit the form (3.39) (see Ref. 25).

Coming back to the general discussion, with the integrability conditions in the general form (3.30), and before introducing local canonical variables, let us discuss the structure of the system (3.18) or (3.34), for U_a and V_i, for given forces $h_i^a(y,t)$.

First of all let us notice that Eq. (3.34) is invariant under the transformation

$$U_a \to U_a + \frac{\partial \Phi}{\partial y^a}, \tag{3.40}$$

where Φ is any function of the $\{y^\alpha\} \equiv \{y^a, t^i\}$.

Under this transformation θ transforms as

$$\theta \to \theta + d\Phi, \tag{3.41}$$

which does not modify the equations of motion. This transformation will be called a canonical transformation.

To be more general, we have to observe that even the transformation

$$\theta \to \theta + g_i(t^1, t^2, \ldots, t^n)\, dt^i, \tag{3.42}$$

does not modify the equations of motion, corresponding to the first of Eqs. (3.16), since there the $U_i = -V_i$ appear differentiated with respect to the y^a only, but it breaks the independence on the path l of the action, expressed by the second of Eqs. (3.16) since it is this equation that appears in $\delta\theta$ as a coefficient of δt^i. So we will not consider any more such a transformation as a true invariance of the theory.

Coming back to the transformation (3.40), we observe that a solution of Eqs. (3.34), recalling the definitiion (3.15), can be written

$$U_a = U_a^{(0)} + \frac{\partial \Phi}{\partial y^a},$$

$$V_i = -U_{2N+i} = U_a^{(0)} h_i^a - \frac{\partial \Phi}{\partial t^i}, \tag{3.43}$$

with $U_a^{(0)}$ a solution of the homogeneous equation

$$Y_i U_a^{(0)} + U_b^{(0)} \frac{\partial h_i^b}{\partial y^a} = 0, \tag{3.44}$$

satisfying the condition

$$\left| \frac{\partial U_a^{(0)}}{\partial y^b} \right| \neq 0. \tag{3.45}$$

This is a consequence of the following facts, which are demonstrated in Appendix B: if we put

$$V_i = -U_{2N+i} = U_a h_i^a + \lambda_i, \tag{3.46}$$

it follows for λ_i,

$$Y_i \lambda_j = Y_j \lambda_i. \tag{3.47}$$

This in turn implies that λ_i must be of the form

$$\lambda_i = -Y_i \Phi, \tag{3.48}$$

with Φ any arbitrary function. Finally, it is shown in Appendix B that

$$U_a - \frac{\partial \Phi}{\partial y^a}$$

satisfies Eq. (3.44), from which we have

$$U_a = U_a^{(0)} + \frac{\partial \Phi}{\partial y^a},$$

and thus

$$V_i = U_a^{(0)} h_i^a - \frac{\partial \Phi}{\partial t^i}.$$

Equation (3.44) is discussed in Appendix B.

With the solution (3.43), the one-form θ can be written as

$$\theta = \theta^{(0)} + d\Phi, \tag{3.49}$$

where

$$\theta^{(0)} = U_a^{(0)}\, dy^a - V_i^{(0)}\, dt^i = U_a^{(0)} \theta^a, \tag{3.50}$$

because

$$V_i^{(0)} = U_a^{(0)} h_i^a. \tag{3.51}$$

The solution (3.43) shows that any independent solution $U_a^{(0)}$ of Eq. (3.44) determines a set of solutions connected by canonical transformations (3.40).

Clearly Eqs. (3.43) define an equivalence relation, and inequivalent choices of Γ (with $\|\Gamma_{ab}\|$ not singular) will correspond to disjoint sets of solutions for U_α, each element of one of these disjoint sets being connected to another of the same set by a canonical transformation.

The situation described here is completely analogous to that of the one-time case, analyzed in Ref. 22.

Now we can demonstrate that each disjoint set has a different symplectic structure, giving to our equations of motion (3.4), or (2.9), the form of canonical equations of motion, which is the main task of the present section. The crucial result is that the two-form $\omega = d\theta$ has rank $2N$, as stressed after Eq. (3.20). On the other hand ω is closed, so from the (generalized) Darboux theorem,[34] we know that local coordinates exist such that

$$\omega = d\bar{\mathbf{p}}_i \wedge d\bar{\mathbf{x}}^i \quad (i = 1,2,\ldots,n), \tag{3.52}$$

from this we get

$$d(\theta - \bar{\mathbf{p}}_i\, d\bar{\mathbf{x}}^i) = 0,$$

or

$$\theta = \bar{\mathbf{p}}_i\, d\bar{\mathbf{x}}^i + d\phi, \tag{3.53}$$

where ϕ is a function of $\bar{\mathbf{x}}^i$, $\bar{\mathbf{p}}_i$, and t^i.

Equation (3.53) shows that, for each inequivalent choice of $\Gamma_{\alpha\beta}$, we have a (different) symplectic structure. The coordinates $\bar{\mathbf{p}}_i$ and $\bar{\mathbf{x}}^i$ are a choice of the Jacobi coordinates (initial data), for which the Hamiltonians are vanishing. Let us now go back to a general choice of canonical variables $\mathbf{x}^i$ and $\mathbf{p}_i$, by performing a (backward) Jacobi transformation.

If we take ϕ to be any function of $\bar{\mathbf{x}}^i$ and t^i, and of a set of N new variables $\mathbf{x}^i$, $\phi = \phi(\mathbf{x}^i, \bar{\mathbf{x}}^i, t^i)$, with the only requirement that

$$\left\| \frac{\partial^2 \phi}{\partial \mathbf{x}^i\, \partial \bar{\mathbf{x}}^j} \right\| \neq 0, \tag{3.54}$$

we may put

$$\frac{\partial \phi}{\partial \bar{\mathbf{x}}^i} = -\bar{\mathbf{p}}_i, \tag{3.55}$$

and

$$\frac{\partial\phi}{\partial t^i} = -\tilde{H}_i(\mathbf{x},\bar{\mathbf{x}},t),$$

$$\frac{\partial\phi}{\partial \mathbf{x}^i} = \mathbf{p}_i, \tag{3.56}$$

Condition (3.54) allows to invert the last equation (3.56), to get $\bar{\mathbf{x}}^i$ as a function of $\mathbf{x}^i$, $\mathbf{p}_i$, and t_i

$$\bar{\mathbf{x}}^i = \mathbf{g}^i(\mathbf{x},\mathbf{p},t). \tag{3.57}$$

Inserting this expression in $\tilde{H}_i$ we get n new functions of $\mathbf{x}^i$, $\mathbf{p}_i$, and t^i

$$H_i(\mathbf{x},\mathbf{p},t) = \tilde{H}_i(\mathbf{x},\mathbf{g}(\mathbf{x},\mathbf{p},t),t), \tag{3.58}$$

which will be our Hamiltonians. The first equation (3.56) can now be written

$$\frac{\partial\phi}{\partial t^i} + H_i\left(\mathbf{x}, \frac{\partial\phi}{\partial\mathbf{x}}, t\right) = 0. \tag{3.59}$$

It is easily verified that the Hamiltonians so determined satisfy the following integrability conditions:

$$\frac{\partial H_i}{\partial t^j} - \frac{\partial H_j}{\partial t^i} + \{H_i,H_j\} = 0, \tag{3.60}$$

where the Poisson bracket is defined with respect to the new canonical coordinates $\mathbf{x}$ and $\mathbf{p}$:

$$\{A,B\} = \sum_{i=1}^{n}\left(\frac{\partial A}{\partial \mathbf{x}^i}\cdot\frac{\partial B}{\partial \mathbf{p}_i} - \frac{\partial B}{\partial \mathbf{x}^i}\cdot\frac{\partial A}{\partial \mathbf{p}_i}\right). \tag{3.61}$$

Indeed, from Eqs. (3.59) we get

$$\begin{aligned}\frac{\partial^2\phi}{\partial t^j\,\partial t^i} &= -\frac{\tilde{H}_i(\mathbf{x},\bar{\mathbf{x}},t)}{\partial t^j}\\ &= -\frac{\partial H_i}{\partial t^j} - \frac{\partial H_i}{\partial \mathbf{p}_k}\cdot\frac{\partial}{\partial t^j}\frac{\partial\phi}{\partial\mathbf{x}^k}\\ &= -\frac{\partial H_i}{\partial t^j} + \frac{\partial H_i}{\partial \mathbf{p}_k}\cdot\left(\frac{\partial H_j}{\partial \mathbf{x}^k} + \frac{\partial H_j}{\partial \mathbf{p}_l}\cdot\frac{\partial^2\phi}{\partial\mathbf{x}^k\,\partial\mathbf{x}^l}\right),\end{aligned}$$

from which

$$\frac{\partial^2\phi}{\partial t^j\,\partial t^i} - \frac{\partial^2\phi}{\partial t^i\,\partial t^j} = -\frac{\partial H_i}{\partial t^j} + \frac{\partial H_j}{\partial t^i} - \{H_i,H_j\} = 0.$$

Let us stress that in this way we may generate sets of n functions H_i satisfying Eqs. (3.60), by simply choosing a function $\phi(\mathbf{x},\bar{\mathbf{x}},t)$, satisfying the condition (3.54).

The one-form θ becomes

$$\begin{aligned}\theta &= \bar{\mathbf{p}}_i\,d\bar{\mathbf{x}}^i + d\phi\\ &= \bar{\mathbf{p}}_i\,d\bar{\mathbf{x}}^i - \bar{\mathbf{p}}_i\,d\bar{\mathbf{x}}^i + \mathbf{p}_i\,d\mathbf{x}^i - H_i\,dt^i,\end{aligned}$$

that is

$$\theta = \mathbf{p}_i\,d\mathbf{x}^i - H_i(\mathbf{x},\mathbf{p},t)\,dt^i, \tag{3.62}$$

which is the local expression for θ we were looking for. This is the generalization to n times of the Poincaré–Cartan one-form. The one-time evolution vector field Y, which is the solution of the equations

$$i_Y\omega = 0, \quad i_Y\,dt = 1 \quad (\omega = d\theta),$$

is now replaced by the n vector fields Y_i, satisfying Eqs. (3.26), and

$$i_{Y_i}\,dt^j = \delta_i^j.$$

Clearly the inverse procedure, that is, given the functions H_i, satisfying the integrability conditions (3.60), to determine the function $\phi(\mathbf{x},\mathbf{p},t)$, amounts in determining a complete integral of the n (integrable) Jacobi equations (3.59).

In Appendix B the relation between the coordinates $\bar{\mathbf{x}}$ and $\bar{\mathbf{p}}$ and the integration constants of the original equations of motion (3.4) is discussed.

If we compare the expression of $d\theta$ as given by Eq. (3.62) with that given by Eq. (3.10), we get the following equality:

$$\Gamma_{ab} = \frac{\partial U_a}{\partial y^b} - \frac{\partial U_b}{\partial y^a} = [y^a,y^b], \tag{3.63}$$

where [,] is the Lagrange bracket

$$[y^a,y^b] = \frac{\partial\mathbf{x}^i}{\partial y^a}\cdot\frac{\partial\mathbf{p}_i}{\partial y^b} - \frac{\partial\mathbf{x}^i}{\partial y^b}\cdot\frac{\partial\mathbf{p}_i}{\partial y^a}, \tag{3.64}$$

where $\mathbf{x}^i$ and $\mathbf{p}_i$ are considered functions of the variables y^a and t^i.

The first equation (3.63) implies

$$\{y^a,y^b\} = -(\Gamma^{-1})^{ab}, \tag{3.65}$$

where we remember the condition (3.17), that is the hypothesis that $\|\Gamma_{ab}\|$ be nonsingular. Equation (3.65) follows from

$$[y^a,y^b]\{y^c,y^b\} = \delta^{ac}. \tag{3.66}$$

The first of Eqs. (3.18) gives

$$h_i^a = -(\Gamma^{-1})^{ab}\Gamma_{bi}, \tag{3.67}$$

where we have taken into account that

$$\{y^a,y^b\}\frac{\partial F}{\partial y^b} = \{y^a,F\}.$$

If we substitute the expression (3.67) into the expression of Γ_{ai} of Eq. (3.63), we get the forces in canonical form

$$h_i^a = \frac{\partial y^a}{\partial t^i} + \{y^a,H_i\}, \tag{3.68}$$

which can also be written

$$h_i^\alpha = Y_i y^\alpha = \left(\frac{\partial}{\partial t^i} + \{\cdot,H_i\}\right)y^\alpha, \tag{3.69}$$

from which we get the action of Y_i on any function of $\{\mathbf{x}^i,\mathbf{p}_i,t^i\}$:

$$Y_i = h_i^\alpha\frac{\partial}{\partial y^\alpha} = \frac{\partial}{\partial t^i} + \{\cdot,H_i\}. \tag{3.70}$$

This equation gives the canonical form of the vector fields, which span the kernel distribution of the two-form ω.

Equation (3.60) is recovered from Γ_{ij}, using Eq. (3.18′).

We see that Γ_{ab} determines the symplectic structure of the theory, as already observed, and that Γ_{ab} is a canonical invariant, which determines the canonical sector on which the original n-time theory is represented.

The equations of motion can now be obtained as the equations of the integral curves of the vector fields Y_i. Indeed, by solving (see, for instance, Ref. 34)

$$i_X\omega = 0, \tag{3.71}$$

we find for X

$$X = c^i Y_i = c^i\left(\frac{\partial}{\partial t^i} + \{\cdot, H_i\}\right), \tag{3.72}$$

for any choice of the functions c^i.

The integral curves of the vector fields Y_i have the equations

$$\begin{aligned} d\mathbf{x}^i &= \{\mathbf{x}^i, H_j\}\, dt^j, \\ d\mathbf{p}_i &= \{\mathbf{p}_i, H_j\}\, dt^j, \end{aligned} \tag{3.73}$$

which are the n-time canonical equations of motion. For any canonical observable $F(\mathbf{x},\mathbf{p},t)$ we get

$$dF = \left(\frac{\partial F}{\partial t^i} + \{F, H_i\}\right) dt^i. \tag{3.74}$$

Equations (3.60) now become the integrability conditioins for Eqs. (3.73).

Observe that the coordinates $\mathbf{x}^i$ and $\mathbf{p}_i$ are now, on the motion, functions of all the times $t^1, t^2, ..., t^n$, contrary to the original positions $\mathbf{q}^i$, which are functions of their own time t^i only.

If we restrict the $\mathbf{x}^i$ and the $\mathbf{p}_i$ of the given symplectic structure $\{\,,\,\}$ associated to a U_α, to $t^1 = t^2 = \cdots = t^n = t$, Eq. (3.73) give the usual one-time Hamilton's equations of motion, with

$$H = \sum_{i=1}^{n} H_i. \tag{3.75}$$

If we consider the case of the restricted integrability conditions (3.6), when, to the given canonical sector $U_\alpha^{(0)}$, a Γ_{ab} is associated, which at equal times satisfies Eq. (3.39), that is

$$\|\Gamma_{ab}|_{t^1, t^2, ..., t^n = t}\| = \begin{pmatrix} \beta & -\alpha \\ \alpha & 0 \end{pmatrix}, \tag{3.76}$$

we have a Lagrangian for Eq. (2.1), and its Hessian is given by the matrix α, so that $\det \alpha \neq 0$. Then

$$(\Gamma^{-1})^{ab} = \begin{pmatrix} 0 & \alpha^{-1} \\ -\alpha^{-1} & \alpha^{-1}\beta\alpha^{-1} \end{pmatrix},$$

which implies

$$\{q^{mi}, q^{kj}\} = 0, \quad \text{at } t^1 = t^2 = \cdots = t^n = t.$$

We can then make the identification

$$\mathbf{q}^i = \mathbf{x}^i, \quad \text{at } t^1 = t^2 = \cdots = t^n = t, \tag{3.77}$$

and, via the Legendre transformation, we may recover the standard one-time Hamiltonian formalism, with

$$H = \sum_{i=1}^{n} H_i.$$

As $\mathbf{q}^i = \mathbf{q}^i(t^i)$, it follows that

$$\{q^{mi}(t^i), q^{ki}(t^i)\}$$

does not depend upon the time t^j, for $j \neq i$: therefore its vanishing at equal times implies

$$\{q^{mi}(t^i), q^{ki}(t^i)\} = 0,$$

also in the n-time case.

Vice versa

$$\{q^{mi}, q^{kj}\} = 0 \quad \text{at } t^1 = t^2 = \cdots = t^n = t,$$

implies, considering Eq. (3.64),

$$(\Gamma^{-1})^{ab} = \begin{pmatrix} 0 & a \\ b & c \end{pmatrix} \Rightarrow \|\Gamma_{ab}|_{t^1 = t^2 = \cdots t^n = t}\| = \begin{pmatrix} -b^{-1}ca^{-1} & b^{-1} \\ a^{-1} & 0 \end{pmatrix},$$

for those symplectic structures for which $\det a \neq 0$, $\det b \neq 0$; but, from

$$\Gamma_{ab} = -\Gamma_{ba}$$

we get

$$a^{-1} = -b^{-1},$$

so we recover Eq. (3.39) with

$$\alpha = b^{-1} \quad \text{and} \quad \beta = b^{-1}cb^{-1}.$$

Instead, for a canonical sector $U_\alpha^{(0)}$, whose Γ_{ab} does not satisfy Eq. (3.39), there is no Lagrangian for Eq. (2.1), giving rise to this symplectic structure, restricted to equal times, via the Legendre transformation. In general, we now have

$$\{q^{mi}, q^{kj}\} \neq 0$$

even at equal times, and it is always $\mathbf{q}^i \neq \mathbf{x}^i$. The only way to define a Hamiltonian is by using Eq. (3.75).

This is a constructive way to get the symplectic structures and the Hamiltonians for the original system of equations of motion, which do not admit a Lagrangian.

Let us conclude this section by showing that the n-times formalism becomes the usual one in the free case, and for equal times, with a suitable choice of the solution of the equations for U_a and V_i.

In the free case we have

$$\{h_i^a\} \equiv \{\mathbf{h}_i^j\}, \quad \mathbf{h}_i^j = \delta_i^j \mathbf{v}^i \quad \text{for } a = 1,2,...,N,$$

and

$$\{h_i^a\} = 0, \quad \text{for } a = N+1, N+2, ..., 2N.$$

It is easily seen that a particular solution of Eqs. (3.17) and (3.18) for U_a and V_i is

$$\begin{aligned} \{U_a\} &= \{m_i \mathbf{v}^i\}, \quad \text{for } a = 1,2,...,N, \\ \{U_a\} &= 0, \quad \text{for } a = N+1, N+2, ..., 2N, \end{aligned} \tag{3.78}$$

and

$$V_i = \tfrac{1}{2} m_i (\mathbf{v}^i)^2. \tag{3.79}$$

The one-form θ becomes

$$\theta = \sum_{i=1}^{n} \left(m_i \mathbf{v}^i\, d\mathbf{q}^i - \frac{1}{2} m_i (\mathbf{v}^i)^2\, dt^i\right). \tag{3.80}$$

The choice (3.78) can also be used in the interacting case, when equal times are chosen $t^1 = t^2 = \cdots = t^n = t$, if the interaction Hamiltonians only depend on the positions, and not on the velocities. In this case we must linearly combine the equations of motion with equal coefficients, in order to restrict to the chosen path $t^i = t$, in the space of the times. We get

$$V = \sum_{i=1}^{n} V_i = \frac{1}{2} \sum_{i=1}^{n} m_i (\mathbf{v}^i)^2 + W(\mathbf{q}^i, t). \tag{3.81}$$

The one-form θ becomes

$$\theta = \sum_{i=1}^{n} m_i \mathbf{v}^i d\mathbf{q}^i - V\,dt, \tag{3.82}$$

which is the usual one.

IV. INVARIANCE TRANSFORMATIONS

In this section the infinitesimal transformations, which leave the equations of motion invariant, are considered. The invariance conditions for any infinitesimal transformation of the coordinates and the conditions for the existence of the corresponding canonical generators are established. These canonical generators are constant of the motion (Noether theorem), and they close a Poisson algebra, providing a canonical realization of the Lie algebra of the group of transformation under consideration, under the condition of a vanishing two-cohomology group.

All this is well known in the one-time case (see for instance Refs. 34 and 39).

The equations of motion (3.31)

$$dy^\alpha = h_i^\alpha\, dt^i, \tag{4.1}$$

where $\alpha = 1,2,...,2N+n$ and $N = 3n$, as already observed, are the characteristic system for the set of equations in the unknown $f(y)$

$$Y_i f(y) = 0. \tag{4.2}$$

An infinitesimal transformation

$$y^\alpha \to y'^\alpha = y^\alpha + \epsilon g^\alpha(y) \tag{4.3}$$

will be an invariance transformation of the equations of motion (4.1) if

$$[L_g, Y_i] = \lambda_{ij} Y_j, \tag{4.4}$$

where L_g is the vector field which performs the transformation (4.3) on the arguments of any function:

$$L_g = g^\alpha \frac{\partial}{\partial y^\alpha}. \tag{4.5}$$

Now, the lhs of Eq. (4.4) is

$$[L_g, Y_i] = (L_g h_i^\alpha - Y_i g^\alpha)\frac{\partial}{\partial y^\alpha}$$
$$= (L_g h_i^\alpha - Y_i g^\alpha + (Y_i g^j) h_j^\alpha)\frac{\partial}{\partial y^\alpha} - (Y_i g^j) Y_j,$$

and we get

$$[L_g, Y_i] + (Y_i g^j) Y_j$$
$$= (L_g h_i^\alpha - Y_i g^\alpha + (Y_i g^j) h_j^\alpha)\frac{\partial}{\partial y^\alpha}$$
$$= (L_g h_i^\alpha - Y_i g^\alpha + (Y_i g^j) h_j^\alpha)\frac{\partial}{\partial y^\alpha}.$$

Therefore, Eqs. (4.4) become

$$[L_g, Y_i] = -(Y_i g^j) Y_j, \tag{4.6}$$

and it implies

$$L_g h_i^\alpha - Y_i g^\alpha + h_j^\alpha Y_i g^j = 0. \tag{4.7}$$

These conditions could be called the *n-time Currie–Hill conditions*, as they are the generalization of the Currie–Hill conditions for the invariance under the Lorentz group in the one-time case.[26]

The same conditions (4.7) can be obtained by requiring that the Lie derivative, with respect to L_g, of the one-forms θ^a, defined in Eqs. (3.22), be proportional to the θ^a themselves.

If the transformations (4.3) form a Lie group G, the corresponding vector fields L_g will close a Lie algebra $\mathcal{G}$. If $L_1, L_2,...,$ is a basis of $\mathcal{G}$, we will have

$$[L_m, L_k] = c^l_{mk} L_l, \tag{4.8}$$

where c^l_{mk} are the structure constants of $\mathcal{G}$.

An infinitesimal transformation is a canonical transformation if it transforms the one-form θ of Eq. (3.13) as

$$\theta \to \theta' = \theta + \epsilon\, d\Omega. \tag{4.9}$$

Therefore, the transformations generated by the vector fields L_g will be canonical if

$$\mathcal{L}_{L_g}\theta = d\Omega_g, \tag{4.10}$$

where $\mathcal{L}_{L_g}$ is the Lie derivative with respect to the vector fields L_g ($\mathcal{L}_{L_g} = i_{L_g} d + d i_{L_g}$). From Eq. (4.10) we get

$$d(\mathcal{L}_{L_g}\theta) = 0,$$

or

$$\mathcal{L}_{L_g}\omega = d(i_{L_g}\omega) = 0, \tag{4.11}$$

where $\omega = d\theta$ (therefore L_g is a $d\theta$ symmetry[40]).

This last equation is the condition to which $\Gamma_{\alpha\beta}$ must satisfy, in order that the action of the group G be canonical. It is the condition for the existence of Ω_g in Eq. (4.9), and, using the expression (3.20) for ω, it gives

$$d(\Gamma_{\alpha\beta} g^\beta dy^\alpha) = 0,$$

or, with the use of Eq. (3.33), we can write

$$L_g \Gamma_{\alpha\beta} + \Gamma_{\alpha\gamma}\frac{\partial g^\gamma}{\partial y^\beta} + \frac{\partial g^\gamma}{\partial y^\alpha}\Gamma_{\gamma\beta} = 0. \tag{4.12}$$

Between all possible inequivalent choices of canonical formulations of the dynamics provided by the Lie–König theorem of Sec. III, only those satisfying the condition (4.12) will give rise to a canonical action of the group G. It may nevertheless appear that a subgroup of G could only be canonically represented; it will be called G_c.[41]

It may be seen that the condition (4.12) implies the condition (4.7), but not vice versa. Indeed, if we multiply Eq. (4.12) by h_i^α, since $h_i^\alpha \Gamma_{\alpha\beta} = 0$ [see Eq. (3.34)], we have, after some rearrangement,

$$\Gamma_{ab}(L_g h_i^b - Y_i g^b + h_j^b Y_i g^j) = 0,$$

which implies Eq. (4.7), since $\det\|\Gamma_{ab}\| \neq 0$.

If the condition (4.12) is satisfied, and since by hypothesis the manifold on which the original equations of motion are defined is simply connected since it is R^{2N+n}, with a global chart of coordinates $\{y^\alpha\} = \{\mathbf{y}^a, t^i\}$, from Eq. (4.11) we get

$$i_{L_g}\omega = dh_g, \tag{4.13}$$

that is the vector fields L_g are globally Hamiltonian,[39] with $h_g = \Omega_g - g^\alpha U_\alpha$. Using Eq. (3.20) for ω, this can be written

$$\frac{\partial h_g}{\partial y^\alpha} = \Gamma_{\alpha\beta} g^\beta, \tag{4.14}$$

which shows that the h_g are determined by $\Gamma_{\alpha\beta}$, apart from a set of integration constants, a change of which does not modify the cohomology class, to which the h_g belong (see in the following of this section). The difference with respect the one-time formulation lies in the fact that ω is here n-fold degenerate, as it is expected by Eq. (3.26). It follows that the correspondence between L_g and h_g given by Eq. (4.13) is not 1–1. L_g will be determined by h_g, apart from elements of the kernel distribution of ω, spanned by the vector fields Y_i.

If we evaluate Eq. (4.13) on the vector field Y_i, we get

$$(i_{L_g}\omega)(Y_i) = Y_i h_g ,$$

and, with Eq. (3.26)

$$Y_i h_g = 0 , \tag{4.15}$$

that is the Hamiltonians h_g are constants of motion. This is the (n-time) Noether theorem in the present context. Moreover, by again operating with i_{L_g} on Eq. (4.13), we get

$$L_g h_g = \mathscr{L}_{L_g} h_g = 0 ,$$

i.e., h_g is invariant under the action of L_g.

In order to clarify the role of the vector fields Y_i, let us consider those infinitesimal transformations (4.3), which leave invariant not only the equations of motion, but even the solutions themselves. They are of the form

$$y^\alpha \to y'^\alpha = y^\alpha + \epsilon F^i(y) h_i^\alpha , \tag{4.16}$$

where the $F^i(y)$ are n arbitrary functions of all the variables. Indeed, if we require that a particular solution of the equations of motion

$$y^\alpha = f^\alpha(t^i) ,$$

where the integration constants belong to the functional form of f^α, be left invariant under a transformation (4.3), that is

$$y'^\alpha = f^\alpha(t'^i) ,$$

we get for g^α

$$g^\alpha = h_i^\alpha g^i ,$$

with g^i arbitrary functions of y^α.

The transformations (4.16) are canonical transformations, since they satisfy the Eq. (4.13), with $h_g = 0$, so they belong to the subgroup G_c, and form a subgroup K of G_c. The Lie algebra of K is clearly spanned by the vector fields Y_i, since the corresponding generators L_g are given by

$$L_{Fh} = F^i h_i^\alpha \frac{\partial}{\partial y^\alpha} = F^i Y_i . \tag{4.17}$$

It easily seen that K is a normal subgroup, due to Eq. (4.4).

The present situation is again parallel to that discussed in Ref. 22.

Since K is normal, we may consider the factor group G_c/K, so getting a decomposition of G_c (or G) in cosets, each element of a given coset having the same effect on the solutions of the equations of motion of the others. As in Ref. 22, in each coset of this decomposition we have only one transformation leaving the times t^i fixed. Indeed, if g_1^α is a given transformation of a coset, by performing a transformation of the same coset such that

$$F^i(y) = -g_1^i(y) , \tag{4.18}$$

we get, from Eq. (4.3), with $\alpha = i = 1,2,...,n$,

$$t^i \to (t')^i = t^i + \epsilon(g_1^i + F^i) = t^i .$$

The transformation (4.18) is uniquely determined, so we have in each coset one and only one transformation, which leave all the times fixed. With this choice of the representative of each coset, we have a faithful representation of the group G_c/K (or G/K).

Observe that Eq. (4.12) is satisfied for any L_{Fh}.

Coming back to Eq. (4.13), let us see what it implies for L_g, in terms of the canonical variables. As already stressed, the correspondence between L_g and h_g is not 1–1, due to the degeneracy of the two-form ω. Equation (4.14) determines h_g, once the $\Gamma_{\alpha\beta}$ is given. Vice versa, from Eq. (4.14), with $\alpha = a$, and saturated with

$$\{y^c, y^a\} = -(\Gamma^{-1})^{ca} ,$$

[see Eq. (3.65)], and with Eq. (3.67), we get

$$\{h_g, y^c\} = g^c - g^i h_i^c . \tag{4.19}$$

Therefore we have the following expression for L_g:

$$L_g = g^\alpha \frac{\partial}{\partial y^\alpha} = (g^\alpha - g^i h_i^\alpha) \frac{\partial}{\partial y^\alpha} + g^i Y_i ,$$

that is

$$L_g = \{h_g, \cdot\} + g^i Y_i , \tag{4.20}$$

where the g^α are such that $i_{L_g}\omega = dh_g$, for a given h_g. Let us stress that $g^i + (g')^i$ is still solution of this equation, therefore: for given h_g, the g^i remain arbitrary. In Eq. (4.20) the $\{\,,\}$ is defined as in Eq. (3.61).

We see from this equation that L_g is determined by h_g, apart from elements of the algebra of the normal subgroup K, as

$$g^i Y_i = (g^i h_i^\alpha) \frac{\partial}{\partial y^\alpha} .$$

As a by-product of Eq. (4.13) we also get

$$i_{Y_i}(i_{L_g}\omega) = i_{Y_i} dh_g = Y_i h_g ,$$

$$-i_{L_g}(i_{Y_i}\omega) = Y_i h_g ,$$

that is, by using Eq. (3.26)

$$Y_i h_g = 0 ,$$

which is another way to get the n-time Noether theorem.

We see from Eq. (4.20) that L_g is a true canonical generator for those transformations which leave the time fixed, that is when $g^i = 0$. These transformations are the ones which belong to the faithful representation mentioned before.

From Eq. (4.20) we get an important result. If we consider the commutator

$$[L_m - g_m^i Y_i, L_n - g_n^j Y_j] ,$$

where L_m, L_n are the elements of a basis of the Lie algebra of G_c, with structure constants c_{mn}^k, as in Eq. (4.8), if we put

$$L_m = g_m^\alpha \frac{\partial}{\partial y^\alpha} , \tag{4.21}$$

we find

$$[L_m - g^i_m Y_i, L_n - g^j_n Y_j] = c^k_{mn}(L_k - g^i_k Y_i)\,, \quad (4.22)$$

where we have used Eq. (4.6).

The result (4.22) implies for h_g:

$$\{h_m, h_n\} = c^k_{mn} h_k + d_{mn}\,, \quad (4.23)$$

where the quantities d_{mn} are constants. They are constant in t^i even, as the h_m are constants of motion, and their Poisson bracket too. They must also satisfy the following relations:

$$d_{mn} = -d_{nm}\,, \quad (4.24)$$

$$c^k_{mn} d_{kl} + c^k_{nl} d_{km} + c^k_{lm} d_{kn} = 0\,, \quad (4.25)$$

that is the d_{mn} are the components of a two-cocycle defined on the Lie algebra $\mathscr{G}_c$ of G_c.

Let us once again stress that the explicit expressions for the h_g are determined by the choice of the $\Gamma_{\alpha\beta}$, as shown by Eq. (4.14), apart from the choice of some integration constants (in all the variables y^a and t^i). Different choices of these integration constants will give different Hamiltonians h_g, belonging to the same two-cocycle class of the two-cohomology group $H^2(\mathscr{G})$ of G.[42] Only if $H^2(\mathscr{G}) = 0$, will in general be possible to make a choice that eliminates the quantities d_{mn}, that is when the d_{mn} are of the form $c^k_{mn} c_k$.[43]

Another point which is worth mentioning, connected with the previous discussion, is the existence of transformations which leave the action invariant,[44] that is such that

$$\theta \xrightarrow{L_g} \theta' = \theta\,, \quad d\Omega_g = 0\,. \quad (4.26)$$

Again following Ref. 22 we will look for a canonical transformation

$$\theta \to \theta' = \theta + d\Lambda\,, \quad (4.27)$$

such that, combined with the transformation (4.9) generated by G, it will give a transformation leaving θ invariant, that is such that

$$\mathscr{L}_{L_g}\theta' = \mathscr{L}_{L_g}(\theta + d\Lambda) = d(\Omega_g + L_g(\Lambda)) = 0\,.$$

If we can choose

$$L_g(\Lambda) = -\Omega_g\,,$$

we will have θ invariant.

The condition for the existence of a function Λ such that

$$\mathscr{L}_{L_g}(\theta + d\Lambda) = 0 \quad (4.28)$$

can be obtained as follows. From Eq. (4.28) we get

$$d(h_g + i_{L_g}\theta + \mathscr{L}_{L_g}\Lambda) = 0\,,$$

where we used Eq. (4.13), that is

$$\mathscr{L}_{L_g}\Lambda = -(h_g + i_{L_g}\theta)\,, \quad (4.29)$$

apart from a constant, which can be reabsorbed in h_g.

If $\{L_m\}$ is a basis of the Lie algebra $\mathscr{G}$, as in Eq. (4.8), we have from Eq. (4.29)

$$[\mathscr{L}_{L_m}, \mathscr{L}_{L_n}]\Lambda = -\mathscr{L}_{L_m}(h_n + i_{L_n}\theta) + \mathscr{L}_{L_n}(h_m + i_{L_m}\theta)\,,$$

or, since

$$[\mathscr{L}_{L_m}, \mathscr{L}_{L_n}] = \mathscr{L}_{[L_m, L_n]} = c^k_{mn}\mathscr{L}_{L_k}\,,$$

$$c^k_{mn}(-h_k - i_{L_k}\theta) = -\mathscr{L}_{L_m}(h_n + i_{L_n}\theta) + \mathscr{L}_{L_n}(h_m + i_{L_m}\theta)\,. \quad (4.30)$$

Now we have

$$\begin{aligned} c^k_{mn} h_k - \mathscr{L}_{L_m} h_n + \mathscr{L}_{L_n} h_m &= c^k_{mn} h_k - L_m(h_n) + L_n(h_m) \\ &= c^k_{mn} h_k - \{h_m, h_n\} + \{h_n, h_m\} \\ &= c^k_{mn} h_k - 2\{h_m, h_n\}\,, \end{aligned}$$

since, from Eqs. (4.20) and (4.15), we have

$$L_m(h_n) = \{h_m, h_n\}\,. \quad (4.31)$$

On the other hand, since

$$i_{[X,Y]} = \mathscr{L}_X i_Y - i_Y \mathscr{L}_X\,,$$

we have

$$\begin{aligned} (\mathscr{L}_{L_m} i_{L_n} - \mathscr{L}_{L_n} i_{L_m})\theta &= i_{[L_m, L_n]}\theta - i_{L_m} i_{L_n}\omega \\ &= c^k_{mn} i_{L_k}\theta - \{h_m, h_n\}\,, \end{aligned}$$

where we used the fact that for any one-form θ it holds

$$\mathscr{L}_X i_Y \theta = i_X \mathscr{L}_Y \theta - i_X i_Y\, d\theta\,, \quad (4.32)$$

and that

$$i_{L_m} i_{L_n}\omega = i_{L_m} dh_n = L_m(h_n) = \{h_m, h_n\}\,.$$

Collecting these results we get that the condition (4.30) for the existence of Λ becomes

$$c^k_{mn} h_k - \{h_m, h_n\} = 0\,. \quad (4.33)$$

Λ will exist when this relation holds for the h_m, or for the h_m with some constant c_m added, since in Eq. (4.29) we disposed of an additive constant for each h_m. This means that the h_m must belong to a two-cohomology class equivalent to zero, or that it must be $H^2(\mathscr{G}) = 0$.

When the condition (4.33) holds, the group G is canonically realized and the h_m are comoments[34]; we know that this is always possible for the inhomogeneous Lorentz group, but not for the Galilei group.

In the n-time canonical formalism the dynamics is described in terms of variables which are functions of n independent times, that is the $\{\mathbf{x}^i\}$ and the $\{\mathbf{p}_i\}$; at equal times we should recover the usual canonical formalism. This implies that we should require that the canonical coordinates $\{\mathbf{x}^i\}$ and the position coordinates $\{\mathbf{q}^i\}$ be coinciding at equal times [see Eq. (3.75)]. It is just the condition which defines the position coordinates in a unique way, when we follow the reverse procedure. This definition of the position coordinates is due to Droz–Vincent,[20] who gave it in a relativistic context. The reverse procedure was used by him in order to build explicit models for two relativistic bodies.

In Sec. II and III we started from a set of equations of motion, written in a form best suited for the nonrelativistic dynamics. However, all that we have said is quite general, and not restricted to the nonrelativistic case.

In the next section we will develop the before mentioned reverse procedure, and we will discuss the position coordinates problem.

V. EXTENDED PHASE SPACE AND CONSTRAINED DYNAMICS

In order to develop an *a priori* canonical theory, from which to recover the position coordinates and their equations of motion, and besides to make contact with the constraint's theory,[6] it is useful to introduce an extended phase space, by defining a set of n new coordinates conjugated to the n times t^i, which will be called the energies ϵ_i, with

$$\{\epsilon_i, t^j\} = \delta_i^j .$$

The new Poisson brackets will be

$$\{A,B\}' = \{A,B\} + \frac{\partial A}{\partial \epsilon_i}\frac{\partial B}{\partial t^i} - \frac{\partial B}{\partial \epsilon_i}\frac{\partial A}{\partial t^i}. \tag{5.1}$$

The space of the variables $\{\mathbf{x}^i, \mathbf{p}_i, t^i\}$ of the previous sections is recovered by requiring the constraints

$$\psi_i = 0, \tag{5.2}$$

with

$$\psi_i = \epsilon_i - H_i(\mathbf{x},\mathbf{p},t) . \tag{5.3}$$

Since the H_i satisfy the integrability conditions (3.60), we have that the functions ψ_i are first class constraints[6]

$$\{\psi_i, \psi_j\}' = 0 . \tag{5.4}$$

The manifold M, defined by Eq. (5.2), which is the original space of the previous section, now becomes a submanifold of the new phase space $\overline{M} = R^{2(N+n)}$ $(N = 3n)$, since the rank of the matrix $\|\partial\psi_i/\partial(x,p,t,\epsilon)\|$ is clearly n over all M, and this guarantees that M is a $2N+n$ submanifold of $\overline{M}$.[45]

The manifold $\overline{M}$ can be defined as a symplectic manifold $(\overline{M}, \overline{\omega})$, with a two-form $\overline{\omega}$ defined by

$$\overline{\omega} = d\overline{\theta},$$
$$\overline{\theta} = \mathbf{p}_i\, d\mathbf{x}^i - \epsilon_i\, dt^i, \tag{5.5}$$

which is clearly a closed nondegenerate two-form.

The submanifold M with the two-form $\omega = d\theta$, defined in Eq. (3.62), which is closed but degenerate [see Eq. (3.26)], is a presymplectic manifold, in which a preferred Poisson structure has been introduced with Eq. (3.61).[28] The definition of M as a submanifold of $\overline{M}$ is an example of a theorem demonstrated by Gotay,[27] which asserts that every presymplectic manifold may be coisotropically embedded in a symplectic manifold. This means that M is a coisotropic submanifold of $\overline{M}$, which, in the language of the constraint's theory, means that it is defined by a set of first class constraints, exactly as we have seen with Eq. (5.4).

The conditions that must be satisfied for a coisotropic embedding are expressed in the following way. Let us define the mapping

$$j: M \to \overline{M},$$

as specified by the conditions (5.2). The conditions are now

(i) $\omega = j^*\overline{\omega}$,

(ii) $TM^\perp \subseteq T_j(TM)$,

where TM $(T\overline{M})$ is the set of tangent vectors on $M(\overline{M})$, $T_j(TM)$ is the set of tangent vectors on $\overline{M}$, which are tangent to the submanifold M, and $TM^\perp$ is defined as the set

$$TM^\perp = \{\overline{X} \in T\overline{M}, \overline{\omega}(\overline{X},\overline{Y}) = 0, \ \forall \overline{Y} \in T_j(TM)\}, \tag{5.6}$$

called the $\overline{\omega}$-orthogonal complement of M.[34]

We now verify that the conditions (i) and (ii) are satisfied, and, in doing that, we will get the precise connection between the extended phase space $\overline{M}$ and the original space M.

It is immediately seen from the definition of $\overline{\omega}$ in Eq. (5.5) and of ω in Eqs. (3.62) and (3.20), that the condition (i) is satisfied. Indeed we have that the mapping j is defined by the transformation

$$j: (\mathbf{x}'^i, \mathbf{p}'_i, t'^i) \to (\mathbf{x}^i, \mathbf{p}_i, t^i, \epsilon_i),$$

with

$$\mathbf{x}^i = \mathbf{x}'^i, \quad \mathbf{p}_i = \mathbf{p}'_i, \quad t^i = t'^i,$$
$$\epsilon_i = H_i(\mathbf{x}^i, \mathbf{p}_i, t^i), \tag{5.7}$$

so we have

$$j^*\overline{\omega} = j^*(d\mathbf{p}_i \wedge d\mathbf{x}^i - d\epsilon_i \wedge dt^i)$$
$$= d\mathbf{p}_i \wedge d\mathbf{x}^i - dH_i(\mathbf{x},\mathbf{p},t) \wedge dt^i = \omega . \tag{5.8}$$

On the other hand, given any vector field on M

$$X = \mathbf{a}^i \cdot \frac{\partial}{\partial \mathbf{x}^i} + \mathbf{b}_i \cdot \frac{\partial}{\partial \mathbf{p}_i} + c^i \frac{\partial}{\partial t^i}, \tag{5.9}$$

we have on $\overline{M}$

$$\overline{X} = T_j(X) = \mathbf{a}^i \cdot \left(\frac{\partial}{\partial \mathbf{x}^i} + \frac{\partial H_j}{\partial \mathbf{x}^i}\frac{\partial}{\partial \epsilon_j}\right) + \mathbf{b}_i \cdot \left(\frac{\partial}{\partial \mathbf{p}_i} + \frac{\partial H_j}{\partial \mathbf{p}_i}\frac{\partial}{\partial \epsilon_j}\right) + c^i\left(\frac{\partial}{\partial t^i} + \frac{\partial H_j}{\partial t^i}\frac{\partial}{\partial \epsilon_j}\right), \tag{5.10}$$

where the functions $\mathbf{a}^i$, $\mathbf{b}^i$, and c^i are restricted on M; this is the form of a vector field on $\overline{M}$ tangent to M. A vector field $\overline{X}$ of $\overline{M}$ belonging to $TM^\perp$ is such that

$$\overline{\omega} = (\overline{X},\overline{Y}) = 0, \quad \text{for any} \quad \overline{Y} \in T_j(TM),$$

and if $\overline{X}$ is given in components as

$$\overline{X} = \alpha^i \cdot \frac{\partial}{\partial \mathbf{x}^i} + \beta_i \cdot \frac{\partial}{\partial \mathbf{p}_i} + \gamma^i \frac{\partial}{\partial t^i} + \delta_i \frac{\partial}{\partial \epsilon_i}, \tag{5.11}$$

this means

$$\alpha^i = \frac{\partial H_j}{\partial \mathbf{p}_i}\gamma^j, \quad \beta_i = -\frac{\partial H_j}{\partial \mathbf{x}^i}\gamma^i,$$
$$\delta_i = -\frac{\partial H_j}{\partial t^i}\gamma^j. \tag{5.12}$$

Substituting in Eq. (5.11) we get

$$\overline{X} = \gamma^j \overline{Y}_j, \tag{5.13}$$

where

$$\overline{Y}_j = \{\psi_j, \cdot\}' \quad (j = 1,2,..,n) . \tag{5.14}$$

These n vector fields are precisely the images under j of the fields Y_i, annihilating the two-form ω

$$i_{Y_i}\omega = 0, \tag{3.26'}$$

that is

$$T_j(Y_i) = \overline{Y}_i + \left(\frac{\partial H_j}{\partial t^i} - \frac{\partial H_i}{\partial t^j} + \{H_j,H_i\}\right)\frac{\partial}{\partial \epsilon_j} = \overline{Y}_i , \tag{5.15}$$

due to the integrability conditions for the H_i.

So we have that $TM^\perp$ is spanned by these vector fields $\overline{Y}_i$, which span the kernel distribution of ω, as found in Sec. III.

The coincidence of the kernel distribution of ω and $TM^\perp$ is characteristic of the coisotropic case; in particular, the vector fields $\overline{Y}_i$ generate the functions ψ_i, which define the submanifold M

$$i_{\overline{Y}_i}\overline{\omega} = d\psi_i , \tag{5.16}$$

which corresponds to Eq. (5.13).

It is now easily verified that even the condition (ii) is satisfied. Indeed any $\overline{Y} \in T_j(TM)$ has the form (5.10), and, with the choice

$$\mathbf{a}^i = \frac{\partial H_j}{\partial \mathbf{p}_i}, \quad \mathbf{b}_i = -\frac{\partial H_j}{\partial \mathbf{x}^i}, \quad c^i = \delta^i_j , \tag{5.17}$$

for any given j, and making again use of the integrability conditions for the H_j, we get that Eq. (5.10) becomes Eq. (5.14), which shows that $TM^\perp \subseteq T_j(TM)$.

In conclusion we have that the two conditions for a coisotropic embedding are satisfied. The relation between $\overline{M}$ and the submanifold M is given by Eqs. (5.8) and (5.10). The equations of motion, which were given by Eq. (3.71), are now given by Eq. (5.16), with the vector fields $\overline{Y}_i$ given by Eqs. (5.14), to which we have to add the algebraic equations $\psi_i = 0$. It is clear from Eq. (5.4) that the ψ_i are constants of motion.

From Eq. (5.16) we get explicitly the equations of motion, as the equations for the integral curves of the vector fields $\overline{Y}_i$

$$\begin{aligned} d\mathbf{x}^i &= \{\mathbf{x}^i,\psi_j\}'\, d\tau^j , \\ d\mathbf{p}_i &= \{\mathbf{p}_i,\psi_j\}'\, d\tau^j , \\ dt^i &= \{t^i,\psi_j\}'\, d\tau^j , \\ d\epsilon_i &= \{\epsilon_i,\psi_j\}'\, d\tau^j , \end{aligned} \tag{5.18}$$

where $j = 1,2,...,n$.

Due to the particular form of the functions ψ_j, the equations for the times are simply

$$dt^i = -d\tau^i , \tag{5.19}$$

which determine the parameters τ^j in terms of physical coordinates. In this way, and using the constraints (5.2), we recover the equations of motion (3.73).

Since the manifold $(\overline{M},\overline{\omega})$ is symplectic, that is the closed two-form $\overline{\omega}$ is now not degenerate, the correspondence between the functions ψ_j and the vector fields $\overline{Y}_j$ given by Eqs. (5.14) or (5.16) does not have the ambiguity of the analogous corespondence in the space M, manifested by the appearance of the undetermined components g^i in Eq. (4.19).

This is also true for the Hamiltonian vector fields generating the transformations of the invariance group G of Sec. IV. We have indeed, using Eq. (5.10), that the vector fields on $\overline{M}$ corresponding to the vector fields L_g of Sec. IV in Eq. (4.12), are

$$\overline{L}_g = T_j(L_g) = \{h_g, \cdot\}' + g^i\{\psi_i, \cdot\}' . \tag{5.20}$$

In getting this result we have to use the fact that the h_g are constants of motion [see Eq. (4.15)], which can now be written

$$\{h_g,\psi_i\}' = 0 , \tag{5.21}$$

and the first class character of the ψ_i.

If we choose a basis $\{h_m\}$ for the Hamiltonians h_g, such that

$$\overline{L}_m = \{h_m, \cdot\}' , \tag{5.22}$$

where $\overline{L}_m = T_j(L_m)$, and where the L_m satisfy the commutation relations (4.8), we see, from Eq. (5.20), that any $\overline{L}_g$ can be written as

$$\overline{L}_g = \sum_m a_m \overline{L}_m + \sum_{i=1}^n g^i \overline{Y}_i , \tag{5.23}$$

which agrees with

$$i_{\overline{L}_g}\overline{\omega} = d\overline{h}_g , \tag{5.24}$$

or

$$\overline{L}_g = \{\overline{h}_g, \cdot\}' , \tag{5.25}$$

where

$$d\overline{h}_g = \sum_m a_m\, dh_m + \sum_{i=1}^n g^i\, d\psi_i . \tag{5.26}$$

Equation (5.24) is easily verified

$$\begin{aligned} i_{\overline{L}_g}\overline{\omega} &= \sum_m a_m i_{\overline{L}_m}\overline{\omega} + \sum_{i=1}^n g^i i_{\overline{Y}_i}\overline{\omega} \\ &= \sum_m a_m\, dh_m + \sum_{i=1}^n g^i\, d\psi_i = d\overline{h}_g . \end{aligned}$$

From Eq. (5.24) we also get

$$\overline{L}_g = \{\overline{h}_g, \cdot\}' = \sum_m a_m \{h_m, \cdot\}' + \sum_{i=1}^n g^i\{\psi_i, \cdot\}' . \tag{5.27}$$

The vector fields $\overline{L}_m$ and $\overline{Y}_i$, tangent to the submanifold M, leave invariant the constraints hypersurface $\psi_i = 0$, indeed

$$\begin{aligned} \overline{L}_m \psi_i &= \{h_m,\psi_i\}' = 0 , \\ \overline{Y}_j \psi_i &= \{\psi_j,\psi_i\}' = 0 , \end{aligned} \tag{5.28}$$

where the first of these equations comes from the constancy of the Hamiltonians h_m [see Eq. (5.21)], and the second corresponds to Eq. (5.4).

In other words, Eqs. (5.28) show that under the action of the algebra $\mathcal{G}$ of the invariance group G, of which the vector fields $\overline{Y}_i$ are an Abelian subalgebra, the new energy canonical coordinates ϵ_i are defined in such a way to transform like the Hamiltonians H_i

$$\begin{aligned} \{h_m,\epsilon_i\}' &= \{h_m,H_i\}' , \\ \{\psi_j,\epsilon_i\}' &= \{\psi_j,H_i\}' . \end{aligned} \tag{5.29}$$

The meaning of the subalgebra spanned by the vector fields $\overline{Y}_i$ is that it generates the transformations of reparametrization $\tau^i \to \tau'^i = \phi^i(\tau)$ of the equations of motion (5.18), and are invariances of the dynamics.

Observe however that the canonical coordinates $\{\mathbf{x}^i,\mathbf{p}_i,t^i,\epsilon_i\}$ are not observables, in the usual meaning in which this word is used in the constraints theory[46]; the observables, that is the canonical quantities with zero Poisson brackets with the constraints ψ_i are more precisely the constants of motion (that is the independent initial data of the dynamical problem). This is a slight departure from the usual gauge theories.

The systems under consideration have vanishing canonical Hamiltonian, $H_c = 0$, and are described by n first-class constraints $\psi_i = 0$, which are in strong involution [see Eq. (5.4)], and which are generators of the gauge transformations in $\overline{M}$.

In the standard approach[6] one introduces the Dirac Hamiltonian

$$H_D = \sum_{i=1}^{n} \lambda^i(\tau)\psi_i , \tag{5.30}$$

and one writes the following Hamiltonian equations:

$$\frac{dA(\tau)}{d\tau} = \{A(\tau),H_D\}' \approx \sum_i \lambda^i(\tau)\{A(\tau),\psi_i\}' , \tag{5.31}$$

where A is any of the canonical variables of the extended phase space, function of a scalar parameter τ. These equations describe the most general gauge transformations in $\overline{M}$, due to the arbitrariness of the Lagrange multipliers $\lambda^i(\tau)$. Equations (5.18) are recovered by introducing n independent parameters τ^i in the following way:

$$d\tau^i = \lambda^i(\tau)\, d\tau , \tag{5.32}$$

and by a redefinition of the quantity A as function of the n τ^i.

We now have all the ingredients for a discussion of the formalism in the extended phase space.

Here we are most interested in discussing the problem of the physical positions $\mathbf{q}^i$, when we start with an *a priori* canonical dynamics. Following Ref. 20, the position coordinates can be defined as the solutions of the following partial differential equations of the first order:

$$\overline{Y}_j\mathbf{q}^i = -\{\mathbf{q}^i,\psi_j\}' = 0 , \quad \text{for } i\neq j , \tag{5.33}$$

where $i,j = 1,2,\ldots,n$, and with

$$\mathbf{q}^i = \mathbf{x}^i , \quad \text{at } t^i = t , \tag{5.34}$$

as Cauchy surface for the definition of the initial conditions.

Observe that the hypersurface $t^i = t$ in the phase space is not invariant under the action of the vector fields $\overline{Y}_i$. This is the required condition for the Lemma (V9) of Ref. 20 to hold. The $\mathbf{q}^i$ determined by Eqs. (5.33) and (5.34) are unique. They are a function of the τ^i, but, in view of Eq. (5.19), they can be thought of as a function of the times t^i.

From Eq. (5.33) it follows

$$\mathbf{v}^i = -\{\mathbf{q}^i,\psi_i\}' = \overline{Y}_i\mathbf{q}^i , \tag{5.35}$$

$$\mathbf{a}^i = \{\{\mathbf{q}^i,\psi_i\}',\psi_i\}' = \overline{Y}_i\mathbf{v}^i , \tag{5.36}$$

and

$$\overline{Y}_j\mathbf{v}^i = -\{\mathbf{v}^i,\psi_j\}' = 0 , \quad \text{for } i\neq j , \tag{5.37}$$

$$\overline{Y}_j\mathbf{a}^i = -\{\mathbf{a}^i,\psi_j\}' = 0 , \quad \text{for } i\neq j . \tag{5.38}$$

Observe that Eqs. (5.21) ensure the invariance of the solutions of Eq. (5.33) under the invariance group G.

In the spirit of the present approach, we have to look for position coordinates $\mathbf{q}^i$ not depending on the energies ϵ_i, since the $\mathbf{q}^i$ must live in the original phase space. So we can reinforce the requirement (5.33) with

$$\{\mathbf{q}^i;t^j\}' = 0 , \quad \text{for any } j. \tag{5.39}$$

From these equations it follows:

$$\{\mathbf{v}^i,t^j\}' = 0 , \tag{5.40}$$

$$\{\mathbf{a}^i,t^j\}' = 0 . \tag{5.41}$$

From Eqs. (5.33), (5.39), (5.40), and (5.41) it follows:

$$\begin{aligned} \mathbf{v}^i &= \overline{Y}_i\mathbf{q}^i = Y_i\mathbf{q}^i , \\ \mathbf{a}^i &= \overline{Y}_i\mathbf{v}^i = Y_i\mathbf{v}^i , \end{aligned} \tag{5.42}$$

where Y_i are the vector fields on M defined in Eq. (3.70).

In the case in which the restricted integrability conditions (3.1) hold, the $\mathbf{q}^i$ and the $\mathbf{v}^i$ coincide with the original y^a of Eq. (3.3). Actually, we have recovered Eqs. (2.9), starting from a phase-space approach. As a by-product, the variables $\mathbf{q}^i$ so obtained have the same covariance properties of the original n-time physical coordinates, and, like them, are not canonical variables. Of course, they become canonical variables when at least one Lagrangian does exist for the Newtonian system [see Eq. (3.76)].

Also in the n-time formulation of the nonrelativistic theory one of the forms of the no-interaction theorem of the relativistic dynamics[18] has been obtained: the identification $\mathbf{q}^i = \mathbf{x}^i$, i.e., both covariant and canonical, is allowed in the free case only, because only then it is $\mathbf{x}^i(t^1,t^2,\ldots,t_n) = \mathbf{x}^i(t^i) = \mathbf{q}^i(t^i)$ (see Ref. 47 for a nonrelativistic no-interaction theorem).

We now give a formal demonstration of the nonrelativistic n-time no-interaction theorem (which can however be also applied to a relativistic system described by a set of first-class constraints). In this demonstration no use is done of the canonical kinematical Galilei (Poincaré) algebra (see the review paper quoted in Ref. 18 for a comparison).

Our hypotheses are that the system is described by the first-class constraints (5.2) in $\overline{M}$; the physical coordinates $\mathbf{q}^i$ are obtained from Eqs. (5.33) and (5.34), with the $\mathbf{v}^i$ and $\mathbf{a}^i$ given by Eqs. (5.35) and (5.36), and moreover Eq. (5.34) still holds when the times are different, so that

$$\{q^{i,m},q^{j,k}\}' = \{x^{i,m},x^{j,k}\}' = 0$$

[this is the crucial hypothesis, since it implies $\mathbf{x}^i = \mathbf{x}^i(t^i)$].

A final hypothesis, which is also needed in the standard relativistic formulation, is that

$$\det\left(\frac{\partial v^{i,m}}{\partial p_{i,k}}\right)\neq 0 . \tag{5.43}$$

From Eq. (5.33) and from $\{q^{i,m},q^{j,k}\}' = 0$ we get

$$\{q^{i,m},v^{j,k}\}' = 0 , \quad \text{for } i\neq j , \tag{5.44}$$

which implies

$$\frac{\partial v^{j,k}}{\partial p_{i,m}} = 0 , \quad \text{for } i\neq j .$$

Then we get

$$\det\left(\frac{\partial v^{j,k}}{\partial p_{i,m}}\right) = \prod_i \det\left(\frac{\partial v^{i,k}}{\partial p_{i,m}}\right) \neq 0\,, \tag{5.45}$$

due to Eq. (5.43).

From Eqs. (5.4), (5.33), and (5.44) we obtain

$$\begin{aligned}
&\{v^{i,m},\psi_j\}' = 0\,, \quad \text{for } i\neq j\,,\\
&\{a^{i,m},\psi_j\}' = 0\,, \quad \text{for } i\neq j\,,\\
&\{v^{i,m},v^{j,k}\}' = 0\,, \quad \text{for } i\neq j\,,\\
&\{q^{i,m},a^{j,k}\}' = 0\,, \quad \text{for } i\neq j\,.
\end{aligned} \tag{5.46}$$

From the last of these equations, and with the use of Eq. (5.44), we get

$$\frac{\partial a^{j,k}}{\partial p_{i,m}} = \sum_l \frac{\partial \mathbf{v}^l}{\partial p_{i,m}}\cdot\frac{\partial a^{j,k}}{\partial \mathbf{v}^l} = \frac{\partial \mathbf{v}^i}{\partial p_{i,m}}\cdot\frac{\partial a^{j,k}}{\partial \mathbf{v}^i} = 0\,, \quad \text{for } i\neq j\,.$$

Equation (5.43) then implies

$$\frac{\partial a^{j,k}}{\partial v^{i,m}} = 0\,, \quad \text{for } i\neq j\,. \tag{5.47}$$

Finally, Eqs. (5.37), (5.44) and the third of Eq. (5.46) imply

$$\begin{aligned}
\{v^{i,m},a^{j,k}\}' &= -\sum_l \frac{\partial v^{i,m}}{\partial \mathbf{p}_l}\cdot\frac{\partial a^{j,k}}{\partial \mathbf{q}^l}\\
&= -\frac{\partial v^{i,m}}{\partial \mathbf{p}_i}\cdot\frac{\partial a^{j,k}}{\partial \mathbf{q}^i} = 0\,, \quad \text{for } i\neq j\,,
\end{aligned} \tag{5.48}$$

and, from Eq. (5.43)

$$\frac{\partial a^{j,k}}{\partial q^{i,m}} = 0\,, \quad \text{for } i\neq j\,. \tag{5.49}$$

Equations (5.47) and (5.49) give the result

$$\mathbf{a}^i = \mathbf{a}^i(\mathbf{q}^i,\mathbf{v}^i,t^i)\,.$$

Space-time invariance then implies

$$\mathbf{a}^i = \mathbf{a}^i(\mathbf{v}^i)\,,$$

and this equation describes a free motion only.

In the case of the restricted integrability conditions (3.1), and by looking at the one-form θ [see Eqs. (3.10) and (3.62)]

$$\begin{aligned}
S = \int_{a(l)}^{b} \theta &= \int_{a(l)}^{b} (\mathbf{U}_i\cdot d\mathbf{q}^i + \mathbf{U}_{i+n}\cdot d\mathbf{v}^i - V_i\, dt^i)\\
&= \int_{a(l)}^{b} (\mathbf{p}_i\cdot d\mathbf{x}^i - H_i\, dt^i)\,,
\end{aligned} \tag{5.50}$$

we see that the requirement $\mathbf{q}^i = \mathbf{x}^i$ at different times, leading to the no-interaction theorem, implies

$$\begin{aligned}
&\mathbf{U}_i = \mathbf{p}_i\,,\\
&\mathbf{U}_{i+n} = \frac{\partial\phi(\mathbf{v}^1,\mathbf{v}^2,\ldots,\mathbf{v}^n)}{\partial \mathbf{v}^i}\,,\\
&V_i = H_i\,.
\end{aligned} \tag{5.51}$$

Therefore, $\mathbf{U}_{i+n}\cdot d\mathbf{v}^i = d\phi(\mathbf{v}^1,\mathbf{v}^2,\ldots,\mathbf{v}^n)$, and this surface term can be eliminated from the action (3.10). Moreover, the form of $\mathbf{U}_{i+n}$ implies

$$\Gamma_{i+n,m;j+n,k} = 0 \Rightarrow \Gamma_{ab} = \begin{pmatrix} a & b\\ -b & 0\end{pmatrix}, \tag{5.52}$$

at different times. Vice versa Eq. (5.52) implies the second of Eqs. (5.51), and Eq. (5.50) gives

$$\mathbf{U}_i\cdot d\mathbf{q}^i - V_i\, dt^i + d\phi = \mathbf{p}_i\cdot d\mathbf{x}^i - H_i\, dt^i\,.$$

This last equation can be rewritten

$$\begin{aligned}
U_i\cdot d\mathbf{q}^i - V_i\, dt^i &= \mathbf{p}_i\cdot d\mathbf{x}^i - H_i\, dt^i - d\phi\\
&= \mathbf{p}'_i\cdot d\mathbf{x}'^i - H'_i\, dt^i\,,
\end{aligned} \tag{5.53}$$

where $\mathbf{x}'^i$, $\mathbf{p}'_i$, H'_i are defined by the canonical transformation associated to ϕ. Therefore Eq. (5.42) imply

$$\mathbf{q}^i = \mathbf{x}'^i\,, \quad \mathbf{U}_i = \mathbf{p}'_i\,, \quad V_i = H'_i\,,$$

at different times.

This is a new formulation of the no-interaction theorem: if Γ_{ab} satisfies Eq. (5.52) at different times, a free motion only is allowed.

Let us return to the line action (3.10) for the first-order system (2.9). A necessary and sufficient condition[48] to get from Eq. (3.10) an action for a second-order system, independent of the accelerations, is

$$\mathbf{U}_{i+n} = \frac{\partial\phi(\mathbf{v}^1,\mathbf{v}^2,\ldots,\mathbf{v}^n)}{\partial \mathbf{v}^i}\,,$$

where

$$\mathbf{v}^i = \frac{d\mathbf{q}^i}{dt}\,.$$

But, as we have just seen, this condition implies the no-interaction theorem. Therefore we get the result that an acceleration independent action for the n-time second-order equation (2.6) (which could be called a predictive action) does not exist, except in the free case. As shown in Ref. 49 a predictive action, giving rise to a canonical realization of the kinematical algebra at the Galilei or Poincaré groups, can be only obtained with a Fokker-like action, depending on the accelerations of every order. Instead, if we restrict the line action (3.10) to the path $t^1 = t^2 = \cdots = t^n$, the previous condition on $\mathbf{U}_{i+n}$ implies Eq. (5.52) at equal times only, and this is Eq. (3.76), which was shown to be the condition for the existence of a Lagrangian associated to the canonical structure Γ_{ab}. Let us remark that, if the canonical structure Γ_{ab} does not admit a Lagrangian, the corresponding $\mathbf{U}_{i+n}$ is not the gradient of a function $\phi(\mathbf{v}^1,\mathbf{v}^2,\ldots\mathbf{v}^n)$.

Let us now consider a set of Newton equations (2.1) admitting a Lagrangian $L(\mathbf{q}^i,\dot{\mathbf{q}}^i,t)$ (or many s-equivalent Lagrangians L_ρ). Its action S_L can be put in a form invariant under a reparametrization, by enlarging the configuration space from $\{\mathbf{q}^i\}$ to $\{\mathbf{q}^i(\tau),t(\tau)\}$

$$S_L = \int d\tau\, \dot{t}(\tau) L\left(\mathbf{q}^i(\tau), \frac{\dot{\mathbf{q}}^i(\tau)}{\dot{t}(\tau)}, t(\tau)\right). \tag{5.54}$$

In the n-time approach, a canonical structure Γ_{ab}, satisfying Eq. (3.76), corresponds to L, and, therefore, there is a set of first-class constraints (5.2) in the enlarged phase space $\bar{M}$. The associated Dirac Hamiltonian H_D [see Eq. (5.30)], gives a set of Hamilton equations, where $\mathbf{x}^i(\tau) = \mathbf{x}^i(t^1(\tau),t^2(\tau),\ldots,t^n(\tau))$. The first half of these equations, implemented with the constraints (5.2), is

$$t^i(\tau) = -\lambda^i(\tau),$$
$$\dot{\mathbf{x}}^i(\tau) = \sum_j \lambda^j(\tau)\{\mathbf{x}^i(\tau),\psi_j\}', \quad (5.55)$$
$$\psi_i = \epsilon_i - H_i(\mathbf{x},\mathbf{p},t) = 0.$$

When these equations can be inverted to get ϵ_i, $\mathbf{p}_i$, λ^i in terms of $\mathbf{x}^i$, $\dot{\mathbf{x}}^i$, t^i, $\dot{t}^i$, the following inverse Legendre transformation generates a singular Lagrangian L_1, depending on $\mathbf{x}^i$ and *not* on $\mathbf{q}^i$

$$L_1(\mathbf{x}^i(\tau),t^i(\tau),\dot{\mathbf{x}}^i(\tau)) = \mathbf{p}_i\cdot\dot{\mathbf{x}}^i - \epsilon_i\dot{t}^i - H_D. \quad (5.56)$$

Its Hessian has n null eigenvectors, implying the existence of the n first-class constraints (5.2), and of n gauge invariances of L_1, of which one is the τ-reparametrization invariance, which guarantees $H_c = 0$.

The L_1 corresponding to the case of a harmonic oscillator has been calculated in Ref. 12, and it is given in Appendix A. It turns out that

$$S_1 = \int d\tau\, L_1 \neq \int (\mathbf{f}^i\cdot d\mathbf{x}^i + g_i\, dt^i)$$

(that is, it is associated to a generalization of Finsler geometry[50]), showing once again that a predictive action does not exist. However, as shown in Ref. 12, S_1 restricted at equal times reproduces Eq. (5.54), since at equal times it is $\mathbf{q}^i = \mathbf{x}^i$. (See Ref. 51 for other forms of the no-interaction theorem in the relativistic case, based on singular Lagrangians.)

Let us remark that the systems we are considering have constraints ψ_i linear in the ϵ_i. Therefore, the first two of Eqs. (5.55) do not depend on ϵ_i, and the λ^i are always equal to the $-\dot{t}^i$. This fact prevents the use of the so-called Dirac Lagrangian L_D, often used for the path integral of the relativistic systems, whose ψ_i are at least quadratics in the $\epsilon_i = p_i^0$

$$L_D(x^{i,\mu},\dot{x}^{i,\mu},\lambda^i) = p_{i,\mu}\dot{x}^{i,\mu} - H_D,$$

where the $p_{i,\mu}$ are obtained from the first half of the Hamilton equations [they are analogous to the first two of Eq. (5.55)], but the equations $\psi_i = 0$ are not required. In L_D the λ^i are Lagrange multipliers, which are considered as einbeins[52] to recover the constraints $\psi_i = 0$ (they appear as secondary constraints from the primary ones

$$\pi_i = \frac{\partial L_D}{\partial\dot{\lambda}^i} = 0).$$

For instance, for the free relativistic particle in the massless limit, the corresponding equations (5.55) cannot be solved for the λ^i, and only L_D is available.

When the Newton equations (2.1) do not admit a Lagrangian L, there are still the first-class constraints for each allowed canonical structure Γ_{ab}, not necessarily satisfying Eqs. (3.76). In this case, either L_1 does not exist [Eqs. (5.55) are not invertible in the $\mathbf{p}_i$], or L_1 does not admit an equal time limit.

VI. CONCLUSIONS

To conclude let us summarize the results obtained. Starting from the Newton equations (2.1), the n-time second-order equations (2.6) were obtained, and then they were transformed in the first-order system (2.9). To each solution U_α^0 of Eqs. (3.44) [modulo a canonical transformation of the kind of Eq. (3.40)] a symplectic structure $(\Gamma^{-1})^{ab}$ is associated. To each of these symplectic structures there is an associated set of first-class constraints (5.2) in the enlarged phase space $\overline{M}$. We have also seen how to recover the original equations (2.9) from each canonical structure, by means of the Droz–Vincent equations (5.30) and (5.31).

Let us remark that what we have constructed are only *local* canonical structure $(\Gamma^{-1})^{ab}$: whether some or all of them can be globalized will depend on the given system; it is also possible that no global canonical structure will exist.

The open problem is which canonical structure is more relevant from the physical point of view; different Hamiltonians will generate different classical and quantum theories. (See Ref. 25 for a review of these ambiguities, and for a rich bibliography on the argument.)

When the Newton equations (2.1) admit one Lagrangian [modulo the trivial transformation $L\to L + dF(q,t)/dt$], there is only one canonical structure with a Γ, which satisfies Eq. (3.76), thus allowing to recover the Lagrangian L. This is the preferred canonical structure, and at equal times we recover the standard Hamiltonian formalism, with $\mathbf{q}^i = \mathbf{x}^i$ and

$$H = \sum_i H_i.$$

When the Newton equations (2.1) admit K (with possibly $K = \infty$) s-equivalent Lagrangians L_ρ, with $\rho = 1,2,\ldots,K$, there are K canonical structures with $\Gamma_{ab}^{(\rho)}$ satisfying Eqs. (3.76). Some of them will be excluded because of the lack of a canonical realization of the kinematical algebra (of the Galilei group, in this case) (see Ref. 53 for a noncanonical realization of the Lorentz algebra). Thus only say $L_{\rho'}$ and $\Gamma_{ab}^{(\rho')}$, with $\rho' = 1,2,\ldots,h\leqslant K$, are left, and, as shown in Ref. 41, each of them canonically realizes a different subgroup of the dynamical symmetries of the equations (2.1). When the system has no bound states, one selects the unique $\Gamma_{ab}^{(\rho')}$, which satisfies the separability condition.[54] The other restrictions which can be imposed are that: (1) $H = \Sigma_i H_i$ must be interpretable as the energy of the system; (2) if we add a perturbation, the perturbed Newton equations must still allow the existence of at least one Lagrangian[55]: this is a very stringent condition, which usually singles out a unique L, and a unique canonical structure; (3) the equal time action $S_{(\rho')} = \int dt\, L_{(\rho')}$ becomes the phase of the wave function at the quantum level: therefore, at least in principle, an interference experiment could discriminate among the various $S_{(\rho')}$.[56] When the Newton equations (2.1) do not allow any Lagrangian, a canonical structure with Γ_{ab} satisfying Eq. (3.76) does not exist. In general, even at equal times, we have $\{q^{i,m},q^{j,k}\}\neq 0$, and, since we do not have a good definition of the energy, we cannot give any particular significance to $H = \Sigma_i H_i$. The other two restrictions are the requirement of a canonical realization of the kinematical group, and the requirement (3) above. Only if the dynamics is separable, a unique canonical structure is singled out.

Once a canonical structure has been chosen, we obtain a well defined set of first-class constraints associated to the original Newton equations, even when they are coupled to external fields. In general the coupling to external fields will be not minimal; in the constraint's approach this implies that, when there is a well defined one-time theory underlying the first class constraints, like in this case, not every coupling with external fields will be allowed, but only those which preserve the first-class character of the constraints. See for a comparison the restrictions on the external supersymmetric fields, when coupled to matter supermultiplets in Ref. 57.

Let us remark that the chosen canonical structure in the presymplectic manifold M has to be identified with the class of Dirac brackets, which can be defined starting from the constraint's theory in the enlarged phase space $\bar{M}$, with the gauge fixings $t_1 - t_2 = \text{const}$. Indeed, in our construction, there is the underlying hypothesis that the time variables are globally defined on M.

As already noticed the present approach can be applied to relativistic particle systems described by a set of first-class constraints. One has to solve the mass-shell constraints in the energies and to apply the present analysis to each determination of the energy spectrum. In this way the manifest covariance is lost and, at the quantum level, this would correspond to the not manifestly covariant Hamiltonian approach of Feshback–Villars[58] after the Foldy–Wouthuysen transformation has been performed.[59]

APPENDIX A: AN EXAMPLE

In order to give an explicit example of the procedure sketched in Sec. II, let us consider the very simple case of a two particles system, with a harmonic mutual force. The one-time equations of motion of this system are

$$\begin{aligned} m_1\ddot{\mathbf{q}}^1 &= -k(\mathbf{q}^1-\mathbf{q}^2), \\ m_2\ddot{\mathbf{q}}^2 &= +k(\mathbf{q}^1-\mathbf{q}^2), \end{aligned} \tag{A1}$$

where k is the elastic constant. In this simple case it is possible to get the following explicit expression of the two-time forces:

$$\begin{aligned} \mathcal{F}^1 &= +(m_1\omega^2/\Delta)\{-\omega(1+\alpha\cos(\omega\tau))(\mathbf{q}^1-\mathbf{q}^2) \\ &\quad +\alpha(\omega\tau\cos(\omega\tau)-\sin(\omega\tau))\mathbf{v}^1 \\ &\quad +(\omega\tau+\alpha\sin(\omega\tau))\mathbf{v}^2\}, \\ \mathcal{F}^2 &= -\alpha\,(m_2\omega^2/\Delta)\{-\omega(\alpha+\cos(\omega\tau))(\mathbf{q}^1-\mathbf{q}^2) \\ &\quad +(\alpha\omega\tau+\sin(\omega\tau))\mathbf{v}^1 \\ &\quad +(\omega\tau\cos(\omega\tau)-\sin(\omega\tau))\mathbf{v}^2\}, \end{aligned} \tag{A2}$$

where

$$\begin{aligned} &\alpha = m_1/m_2, \quad \omega=\sqrt{k/\mu}, \\ &\mu = m_1m_2/m, \quad m = m_1+m_2, \end{aligned} \tag{A3}$$

and

$$\begin{aligned} &\tau = t^1 - t^2, \\ &\Delta = \omega[1+\alpha^2+2\alpha\cos(\omega\tau)+\alpha\omega\tau\sin(\omega\tau)]. \end{aligned} \tag{A4}$$

It is easily verified that, when $\tau = 0$, we get the forces (A1). Moreover, we may verify that the integrability conditions

$$\frac{d\mathcal{F}^1}{dt^2} = 0, \quad \frac{d\mathcal{F}^2}{dt^1} = 0, \tag{A5}$$

are satisfied.

In Eqs. (A2) the $\mathbf{q}^i$ and the $\mathbf{v}^i$ ($i = 1,2$) must be understood as functions of their own time t^i:

$$\mathbf{q}^1 = \mathbf{q}^1(t^1), \quad \mathbf{q}^2 = \mathbf{q}^2(t^2),$$

and so forth.

The general solution of the two-time equations of motion

$$m_i\mathbf{a}^i(t^i) = \mathcal{F}^i(\tau,\mathbf{q}^1-\mathbf{q}^2,\mathbf{v}^i), \tag{A6}$$

can be written

$$\begin{aligned} \mathbf{q}^1(t^1) &= \mathbf{a}+\mathbf{b}t^1+\mathbf{c}\cos(\omega t^1)+\mathbf{d}\sin(\omega t^1), \\ \mathbf{q}^2(t^2) &= \mathbf{a}+\mathbf{b}t^2-\alpha[\mathbf{c}\cos(\omega t^2)+\mathbf{d}\sin(\omega t^2)], \end{aligned}$$

where **a**, **b**, **c**, and **d** are 12 constraints of integration.

Observe that $\mathcal{F}^1 + \mathcal{F}^2$ is in general different from zero, except when $\tau = 0$. That is, the action-reaction law is satisfied at $\tau = 0$ only.

Also observe that, since the original set of equations (A1) is autonomous, the forces (A2) depend on $\tau = t^1 - t^2$ only.

We give here without demonstrations an example of two first class constraints, describing the dynamics of two particles of masses m_1 and m_2, respectively,

$$\psi_i = \epsilon_i - H_i \approx 0 \quad (i = 1,2), \tag{A7}$$

with

$$H_i = (1/2m_i)[\mathbf{p}_i^{\,2} + V(\rho^2)], \tag{A8}$$

where

$$\rho = \mathbf{r} - (\tau/m)\mathbf{p},$$

and

$$\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2, \quad \mathbf{r} = \mathbf{x}^1 - \mathbf{x}^2.$$

Here $\mathbf{x}^i$ are the canonical variables for the two particles. When we consider the equal time dynamics, we have to add the two constraints and to put $t^1 = t^2$. In this case we have $\mathbf{x}^i = \mathbf{q}^i$, and the sum of the two constraints becomes the usual conservation of the total energy, with a potential given by

$$(1/2\mu)V(\mathbf{r}^2),$$

μ being the reduced mass.

In the case of a harmonic oscillator we have to choose

$$V(\rho^2) = \mu^2\omega^2\rho^2.$$

In this last case it is possible to give the explicit expression of the singular Lagrangian, mentioned in Sec. V, and given in Ref. 12. Its expression is the following:

$$L_1 = \left[1 + \kappa\frac{m_1m_2}{\dot{t}^1\dot{t}^2}\left(\frac{\dot{t}^1}{m_1} - \frac{\dot{t}^2}{m^2}\right)^2\right]^{-1} \cdot \left[\frac{m_1(v^1)^2}{2\dot{t}^1} + \frac{m_2(v^2)^2}{2\dot{t}^2} + \kappa\frac{m_1m_2}{\dot{t}^1\dot{t}^2}\left(\frac{\dot{t}^1}{m_1} + \frac{\dot{t}^2}{m_2}\right)(v^1-v^2)^2\right] - \frac{\mu^2\omega^2}{2}\left(\frac{\dot{t}^1}{m_1} + \frac{\dot{t}^2}{m_2}\right)\mathbf{r}^2,$$

where

$$\kappa = \mu^2\omega^2\tau^2/m^2 ,$$

and

$$v^i = \dot{\mathbf{x}}^i + \frac{\mu^2\omega^2\tau}{m}\left(\frac{\dot{t}^1}{m_1} + \frac{\dot{t}^2}{m_2}\right)\mathbf{r} .$$

APPENDIX B: DISCUSSION OF EQ. (3.44)

In this appendix we want to show that the equations for U_a and V_i

$$\Gamma_{ab} h_i^b = \Gamma_{ai} , \quad \Gamma_{ab} h_i^a h_j^b = \Gamma_{ij} \tag{3.18$'$}$$

(where we recall that $a,b = 1,2,...,2N$; and $i,j = 1,2,...,n$) imply for the λ_i defined by the equation

$$V_i = U_a h_i^a + \lambda_i , \tag{3.46}$$

the condition

$$Y_j \lambda_i = Y_i \lambda_j , \tag{3.47}$$

which has for general solution

$$\lambda_i = -Y_i \Phi , \tag{3.48}$$

with Φ an arbitrary function of $\{y^a\}$ and $\{t^i\}$.

If we perform the substitution (3.46) in the expression for the Γ_{ai}, and use Eqs. (3.18$'$), we get

$$\begin{aligned}\Gamma_{ai} &= \frac{\partial U_a}{\partial t^i} + \frac{\partial}{\partial y^a}(U_b h_i^b + \lambda_i) \\ &= \frac{\partial U_a}{\partial t^i} - \Gamma_{ab} h_i^b + h_i^b \frac{\partial U_a}{\partial y^b} \\ &\quad + U_b \frac{\partial h_i^b}{\partial y^a} + \frac{\partial \lambda_i}{\partial y^a} \\ &= Y_i U_a + U_b \frac{\partial h_i^b}{\partial y^a} + \frac{\partial \lambda_i}{\partial y^a} + \Gamma_{ai} ,\end{aligned}$$

from which we get the following equation for the $\{U_a\}$, once the λ_i are given

$$Y_i U_a = -U_b \frac{\partial h_i^b}{\partial y^a} - \frac{\partial \lambda_i}{\partial y^a} . \tag{B1}$$

The second of Eqs. (3.18$'$) can be written

$$\begin{aligned}&h_i^a Y_j U_a - h_j^a Y_i U_a \\ &= -U_a\left(\frac{\partial h_i^a}{\partial t^j} - \frac{\partial h_j^a}{\partial t^i}\right) - \left(\frac{\partial \lambda_i}{\partial t^j} - \frac{\partial \lambda_j}{\partial t^i}\right),\end{aligned}$$

which, using Eq. (B1), gives

$$Y_j \lambda_i - Y_i \lambda_j + U_a (Y_j h_i^a - Y_i h_j^a) = 0 .$$

But the h_i^a satisfy the integrability conditions (3.8), so we get

$$Y_j \lambda_i = Y_i \lambda_j .$$

The general solution of these equations in λ_i is given by Eq. (3.48), since they are the integrability conditions in order that Eq. (3.48) could be integrated in Φ, for given λ_i.

With Eq. (3.48) the functions V_i can be written

$$V_i = U_a h_i^a - Y_i \Phi ,$$

so that Eq. (B1) for the U_a becomes

$$Y_i U_a + U_b \frac{\partial h_i^b}{\partial y^a} = Y_i \frac{\partial \Phi}{\partial y^a} + \frac{\partial \Phi}{\partial y^b}\frac{\partial h_i^b}{\partial y^a} ,$$

since

$$\left[\frac{\partial}{\partial y^a}, Y_i\right] = \frac{\partial h_i^b}{\partial y^a}\frac{\partial}{\partial y^b} , \tag{B2}$$

or

$$Y_i\left(U_a - \frac{\partial \Phi}{\partial y^a}\right) + \frac{\partial h_i^b}{\partial y^a}\left(U_b - \frac{\partial \Phi}{\partial y^b}\right) = 0 , \tag{B3}$$

which is Eq. (3.44). So we finally have

$$U_a = U_a^0 + \frac{\partial \Phi}{\partial y^a} ,$$

and

$$V_i = U_a^0 h_i^a - \frac{\partial \Phi}{\partial t^i} ,$$

with U_a^0 satisfying Eq. (3.44), which is the wanted result.

It is now easily verified that Eq. (3.44) for the U_a^0 are integrable. Indeed the integrability conditions of a system like

$$Y_i U_a = A_{ai}(y,t,U) , \tag{B4}$$

are

$$Y_j A_{ai} = Y_i A_{aj} , \tag{B5}$$

as it is easily verified using Eq. (3.25).

For the Eq. (3.44) the condition (B5) becomes

$$Y_j\left(U_b^0 \frac{\partial h_i^b}{\partial y^a}\right) = (i \leftrightarrow j) ,$$

that is

$$U_b^0 Y_j \frac{\partial h_i^b}{\partial y^a} - U_c^0 \frac{\partial h_j^c}{\partial y^b}\frac{\partial h_i^b}{\partial y^a} = (i \leftrightarrow j) ,$$

or, using Eq. (B2)

$$U_b^0 \frac{\partial}{\partial y^a} Y_j h_i^b - U_c^0\left(\frac{\partial h_j^c}{\partial y^b}\frac{\partial h_i^b}{\partial y^a} + \frac{\partial h_j^b}{\partial y^a}\frac{\partial h_i^c}{\partial y^b}\right) = (i \leftrightarrow j) ,$$

and finally

$$U_b^0 \frac{\partial}{\partial y^a}(Y_j h_i^b - Y_i h_j^b) = 0 , \tag{B6}$$

which is satisfied. So the equation for U_a^0 are integrable.

Another way to get the result (3.48) and the last result, or better to see how to integrate the equations for U_a^0, once the general solution of the original equations of motion is known, is the following.

Let us introduce the new variables

$$z^a = \tilde{z}^a(y,t) , \tag{B7}$$

defined as the solutions of the system

$$Y_i \tilde{z}^a = \left(\frac{\partial}{\partial t^i} + h_i^a \frac{\partial}{\partial y^a}\right)\tilde{z}^a = 0 . \tag{B8}$$

This system is completely integrable, since the vector fields Y_i are commuting. It follows that it has $2N$ independent solutions, which we call $\tilde{z}^a(y,t)$. For a general function $f(y,t)$ let us put

$$f(y,t) = \tilde{f}(z,t) \,. \tag{B9}$$

From Eq. (B8) we have

$$Y_i f(y,t) = \frac{\partial}{\partial t^i} \tilde{f}(z,t) \,, \tag{B10}$$

and this means that Eqs. (3.47) can be written

$$\frac{\partial}{\partial t^j} \tilde{\lambda}_i(z,t) = \frac{\partial}{\partial t^i} \tilde{\lambda}_j(z,t) \,, \tag{B11}$$

which has the general solution

$$\tilde{\lambda}_i(z,t) = - \frac{\partial}{\partial t^i} \tilde{\Phi}(z,t) = - Y_i \Phi(y,t) \,. \tag{B12}$$

Transforming the equation for the U^0_a to the new variables, that is Eq. (3.44), we get

$$\frac{\partial}{\partial t^i} \tilde{U}^0_a(z,t) = - \tilde{U}^0_b(z,t) \frac{\partial \tilde{z}^c}{\partial y^a} \frac{\partial}{\partial z^c}\left(\frac{\partial \tilde{y}^b}{\partial t^i}\right),$$

where $\tilde{y}^b(z,t)$ are the old variables in terms of the new ones. Since

$$\frac{\partial \tilde{z}^a}{\partial y^b} \frac{\partial \tilde{y}^b}{\partial z^c} = \delta^a_c \,, \tag{B13}$$

for the required independence of the $\tilde{z}^a$, we get

$$\frac{\partial \tilde{y}^a}{\partial z^c} \frac{\partial \tilde{U}^0_a}{\partial t^i} = - \tilde{U}^0_b \frac{\partial}{\partial t^i} \frac{\partial \tilde{y}^b}{\partial z^c} \,,$$

or

$$\frac{\partial}{\partial t^i}\left(\frac{\partial \tilde{y}^a}{\partial z^b} \tilde{U}^0_a\right) = 0 \,. \tag{B14}$$

These equations have the general solution for $\tilde{U}^0_a$

$$\tilde{U}^0_a = K_a(\tilde{z}(y,t)) \,, \tag{B15}$$

where the functions K_a are $2N$ arbitrary functions.

Clearly, the variables z^a are nothing more than the complete set of the constants of motion of the original system of equations of motion. Their knowledge determines the U^0_a, with the arbitrariness expressed by Eq. (B14), and in turn they determine the whole class of solutions for the U_a and V_i.

The one-form θ^0 of Eq. (3.50) can now be written

$$\theta^0 = K_a \frac{\partial \tilde{y}^a}{\partial z^b} dz^b + K_a\left(\frac{\partial \tilde{y}^a}{\partial t^i} - \tilde{h}^a_i\right) dt^i \,;$$

but, from the definition of the variables $z^a = \tilde{z}^a(y,t)$, and from their assumed invertibility, we get

$$\frac{\partial \tilde{z}^a}{\partial y^b} \frac{\partial \tilde{y}^b}{\partial t^i} + \frac{\partial \tilde{z}^a}{\partial t^i} = 0 \,, \tag{B16}$$

which, beside Eq. (B8), tells us that

$$\frac{\partial \tilde{y}^b}{\partial t^i} = \tilde{h}^b_i \,. \tag{B17}$$

So we get

$$\theta^0 = K_a \frac{\partial \tilde{y}^a}{\partial z^b} dz^b \,. \tag{B18}$$

Clearly, we may choose

$$\begin{aligned} K_a \frac{\partial \tilde{y}^a}{\partial z^b} &= z^{b+N}, \quad \text{for } b = 1,2,\dots,N \,, \\ K_a \frac{\partial \tilde{y}^a}{\partial z^b} &= 0 \,, \quad \text{for } b = N+1, N+2,\dots,2N \,, \end{aligned} \tag{B19}$$

which is a solution which satisfies the condition (3.17). Indeed, we may write Γ_{ab} in the following way:

$$\Gamma_{ab} = \frac{\partial \tilde{z}^c}{\partial y^a} W_{cd} \frac{\partial \tilde{z}^d}{\partial y^b} \,, \tag{B20}$$

with

$$W_{cd} = \frac{\partial}{\partial z^d}\left(\frac{\partial \tilde{y}^a}{\partial z^c} \tilde{U}^0_a\right) - \frac{\partial}{\partial z^c}\left(\frac{\partial \tilde{y}^a}{\partial z^d} \tilde{U}^0_a\right), \tag{B21}$$

and it is only necessary to verify that W_{cd} is nonsingular. But with the choice (B19), we see that W_{cd} is the symplectic metric:

$$W_{cd} = \delta_{c+N,d} - \delta_{d+N,c} \,, \tag{B22}$$

which has determinant equal to $+1$.

With the choice (B19) we have

$$\theta^0 = \sum_{b=1}^{N} z^{b+N} dz^b \,, \tag{B23}$$

which is in symplectic form, and shows the connection of the variables $\bar{p}_i$ and $\bar{x}^i$ ($i = 1,2,\dots,N$), with the constants of motion z^a ($a = 1,2,\dots,2N$) [see Eq. (3.53) and (3.49)].

[1]E. E. Salpeter and H. Bethe, Phys. Rev. **84**, 1232 (1951); M. Gell-Mann and F. Low, *ibid.* **84**, 350 (1951); G. C. Wick, *ibid.* **96**, 1124 (1954); M. Nakanishi, Suppl. Prog. Theor. Phys. **43**, 1 (1969); A. A. Arkhipov and I. V. Savrin, Serpukhov report No. 83-50, 1983 (unpublished).

[2]I. T. Todorov, Report No. Comm. JINRE2-10125, Dubna, 1976 (unpublished); Ann. Inst. H. Poincaré A **28**, 207 (1978).

[3]A. Komar, Phys. Rev. D **18**, 1881, 1887, 3617 (1978); **19**, 2908 (1979).

[4](a) *Theory of Action-at-a-Distance in Relativistic Particle Mechanics*, edited by E. H. Kerner (Gordon and Breach, New York, 1972); (b) *Relativistic Action-at-a-Distance: Classical and Quantum Aspects*, edited by J. A. Llosa, *Lectures Notes in Physics*, Vol. 162 (Springer, Berlin, 1982); (c) *Constraint's Theory and Relativistic Dynamics*, Firenze, 1986, edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987); (d) E. C. G. Sudarshan and N. Mukunda, *Classical Mechanics: A Modern Perspective* (Wiley, New York, 1974).

[5]H. W. Crater and P. Van Alstine, Phys. Rev. D **30**, 2585 (1984); H. Sazdjian, *ibid.* **33**, 3401, 3425, 3435 (1986); Phys. Lett. B **156**, 381 (1985).

[6]P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); *Lectures on Quantum Mechanics*, Belfer Graduate School of Science, Monographs Series, Yeshiva University, New York, 1964; J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951); P. G. Bergmann and J. Goldberg, *ibid.* **98**, 531 (1955); E. C. G. Sudarshan and N. Mukunda, *Classical Mechanics: A Modern Perspective*, (Wiley, New York, 1974); A. Lichnerowicz, C. R. Acad. Sci. Paris A **280**, 523 (1975); A. J. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems*, Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche, Fisiche e loro Applicazioni, no. 22, Accademia Nazionale dei Lincei, Roma, 1975; K. Sundermeyer, *Constrained Dynamics with Applications to Yang–Mills Theory, General Relativity, Classical Spin, Dual String Model, Lecture Notes in Physics*, Vol. 169 (Springer, Berlin, 1982).

[7]Ph. Droz-Vincent, Lett. Nuovo Cimento **1**, 839 (1969); Phys. Scr. **2**, 129 (1970); Nuovo Cimento B **12**, 1 (1972); Ann. Inst. H. Poincaré A **20**, 269 (1974); A **27**, 407 (1977); A **32**, 377 (1980); Rep. Math. Phys. **8**, 79 (1975).

[8]Ph. Droz-Vincent, Lett. Nuovo Cimento **1**, 800 (1971); Phys. Rev. D **19**, 702 (1979); Nuovo Cimento A **58**, 355 (1980); Lett. Nuovo Cimento, **30**, 375 (1981); **33**, 383 (1982); **35**, 83 (1982); **38**, 177 (1983); **40**, 165 (1984); **44**, 199 (1985); Phys. Rev. D **29**, 687 (1984).

[9]V. A. Rizov, H. Sazdjian, and I. T. Todorov, Ann. Phys. (NY) **165**, 59 (1985).
[10]V. Iranzo, J. Llosa, F. Marqués, and A. Molina, Ann. Inst. H. Poincaré, A **40**, 1 (1984).
[11]G. Longhi and L. Lusanna, Phys. Rev. D **34**, 3707 (1986).
[12]L. Lusanna, in *Proceedings of the VII Seminar on Problems of High Energy Physics and Quantum field Theory*, (Protvino, USSR, 1984), Vol. I, p. 123; in *Proceedings of the IV M. Grossmann Meeting*, Rome, 1985, edited by R. Ruffini (Elsevier, Amsterdam, 1986), p. 1163.
[13]Ph. Droz-Vincent, Lett. Nuovo Cimento **23**, 184 (1978); **28**, 166 (1980); in Ref. 4(b) and 4(c); J. Geom. Phys. **2**, 101 (1985).
[14]M. H. L. Pryce, Proc. R. Soc. London A **195**, 62 (1948); T. D Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949); L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950); T. F. Jordan and N. Mukunda, *ibid.* **132**, 1842 (1953); E. H. Kerner, J. Math. Phys. **6**, 1218 (1965); A. J. Kalnay, "The Localization Problem" in *Studies in Foundations, Methodology and Philosophy of Science*, edited by M. Bunge (Springer, Berlin, 1971), Vol. 4; M. Lorente and P. Roman, J. Math. Phys. **15**, 70 (1974); J. Niederle, "Localizability of Particles" in *Hadron Constituents and Symmetries*, Smolenice, 1976, edited by J. Lonik and M. Noga (Veda, Bratislavia, 1978), Vol. 3; H. Sazdjian, Nucl. Phys. B **161**, 469 (1979).
[15]I. T. Todorov, Sissa report, Trieste, 1980 (unpublished); in *Quantum Theory, Groups, Fields and Particles*, edited by A. O. Barut (Reidel, Dordrecht, 1983); T. Takabayasi, Suppl. Prog. Theor. Phys. **67**, 1 (1979); L. Lusanna, Nuovo Cimento B **65**, 135 (1981); G. Longhi, in Ref. 4(b).
[16]L. Bel, *Mecanica Relativista Predictiva*, Report UAB FT-34 de la Universidad Autonoma de Barcelona, 1977 (unpublished); in *Differential Geometry and Relativity*, edited by M. Cahen and M. Flato (Reidel, Dordrecht, 1976), p. 197; R. N. Hill, in Ref. 4(b); V. Iranzo, J. A. Llosa, and A. Molina, Ann. Phys. (NY) **160**, 114 (1983); J. M. Pons, *ibid.* **148**, 192 (1983); D. G. Currie and T. F. Jordan, in *Quantum Theory and Statistical Physics*, Boulder, 1967, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. XA, p. 91.
[17]L. Bel, Ann. Inst. H. Poincaré A **12**, 307 (1970); **14**, 189 (1971); **18**, 57 (1973), and in Ref. 4(b); R. Arens, Arch. Rat. Mech. Anal. **47**, 255 (1972), and in Ref. 4(b); L. Bel and J. Martin, Ann. Inst. H. Poincaré A **22**, 173 (1975); **33**, 409 (1980); **34**, 231 (1981); D. Hirondel, J. Math. Phys. **15**, 1689 (1974); **20**, 104 (1979); L. Bel and X. Fustero, Ann. Inst. H. Poincaré A **25**, 411 (1976).
[18]D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963); H. Leutwyler, Nuovo Cimento **37**, 556 (1965); E. Kerner, J. Math. Phys. **6**, 1218 (1965); L. Bel. Ann. Inst. H. Poincaré A **14**, 189 (1971); Ph. Droz-Vincent, Nuovo Cimento **12**, 1 (1972); Rep. Math. Phys. **8**, 79 (1975); A. F. Kracklauer, J. Math. Phys. **17**, 693 (1976); A. Salas, *ibid.* **20**, 2459 (1979); V. Iranzo, J. Llosa, F. Marqués, and A. Molina, *ibid.* **24**, 1665 (1983); F. Marqués, V. Iranzo, A. Molina, A. Montoto, and J. Llosa, Phys. Rev. D **31**, 314 (1985); X. Jaen, A. Molina, and V. Iranzo, J. Math. Phys. **27**, 512 (1986); J. Pons, Ann. Phys. (NY) **148**, 192 (1983). For a review, see S. Chelkowski, J. Nietendel, and R. Suchanek, Acta Phys. Polon. B **11**, 809 (1980).
[19]P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[20]Ph. Droz-Vincent, Rep. Math. Phys. **8**, 79 (1975).
[21]E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U. P., New York, 1960), p. 275.
[22]R. N. Hill, J. Math. Phys. **8**, 1756 (1966).
[23]W. Sarlet and F. Cantrijn, Hadronic **J. 1**, 101, 1497 (1978).
[24]W. Sarlet, J. Phys. A **15**, 1503 (1982).
[25]G. M. Morandi, C. Ferrario, G. Lo Vecchio, G. Marmo, and C. R. Rubano, "The inverse problem in the calculus of variations and the geometry of the tangent bundle," preprint Univ. Napoli, 1987.
[26]D. G. Currie, Phys. Rev. **142**, 817 (1966); R. N. Hill, J. Math. Phys. **8**, 201 (1967).
[27]M. J. Gotay, Proc. Am. Math. Soc. **84**, 111 (1982).
[28]J. M. Souriau, *Structure Des Systèmes Dynamiques* (Dunod, Paris, 1970). See also Ref. 39.
[29]A. Komar, Phys. Rev. D **18**, 1881 (1978); **19**, 2908 (1979); C. Piron, in *Quantum Theory, Groups, Fields and Particles*, edited by A. O. Barut (Reidel, Dordrecht, 1983).
[30]G. Longhi, "Multitime approach to nonrelativistic and relativistic quantum mechanics," in Actas de los Encuentros Relativistas '84, St. Ander, 1984, edited by the Universidad de Valencia, 1984, p. 131.
[31]G. Longhi, L. Lusanna, and J. M. Pons, in Ref. 4(c), p. 252.
[32]E. H. Kerner, J. Math. Phys. **6**, 1218 (1965).
[33]See for instance L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).
[34]R. Abraham and J. E. Mardsen, *Foundations of Mechanics*, second ed. (Benjamin/Cummings, London, 1978).
[35]R. M. Santilli, *Foundations of Theoretical Mechanics I, the Inverse Problem in Newtonian Mechanics* (Springer, New York, 1978).
[36]W. Sarlet, Hadronic J. **2**, 407 (1979); **3**, 765 (1980); J. Phys. A **14**, 2227 (1981); W. Sarlet and F. Cantrijn, Hadronic J. **1**, 101 (1978).
[37]M. Crampin, J. Phys. A **14**, 2567 (1981); M. Henneaux, *ibid.* **15**, L93 (1982); Ann. Phys. (NY) **140**, 45 (1982); P. De Ritis, G. Marmo, G. Platania, and P. Scudellaro, Int. J. Theor. Phys. **22**, 931 (1983).
[38]M. Crampin, G. E. Prince, and G. Thomson, J. Phys. A **17**, 1437 (1984).
[39]N. Woodhouse, *Geometric Quantization* (Clarendon, Oxford, 1980).
[40]W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).
[41]N. Mukunda, A. P. Balachandran, Jan S. Nilsson, E. C. G. Sudarshan, and F. Zaccaria, Phys. Rev. D **23**, 2189 (1981).
[42]See for instance Ref. 36.
[43]G. Marmo, E. J. Saletan, A. Simoni, and B. Vitale, *Dynamical Systems* (Wiley, Chichester, 1985).
[44]G. Marmo, G. Morandi, A. Simoni, and E. C. G. Sudarshan, "Quasi-invariance and central extensions,"preprint Univ. Napoli, 1986; G. Marmo, E. J. Saletan, and A. Simoni, Nuovo Cimento B **96**, 159 (1986); see also E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[45]Y. Choquet-Bruhat, C. de Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, Amsterdam, 1977), p. 228.
[46]See for instance P. G. Bergmann and J. Goldberg, Phys. Rev. **98**, 531 (1955); P. G. Bergmann, Rev. Mod. Phys. **33**, 510 (1961).
[47]X. Jaen, J. Llosa, F. Marqués, and A. Montoto, J. Math. Phys. **27**, 519 (1986).
[48]M. Henneaux, Ann. Phys. (NY) **140**, 45 (1982).
[49]E. H. Kerner, J. Math. Phys. **3**, 35 (1962); **6**, 1218 (1965); R. P. Gaida, Yu. B. Klyuchkovskii, and V. I. Tretyak, Theor. Math. Phys. **44**, 194 (1980); **45**, 180 (1980); in *Proceedings of the VII Seminar on Problems of High Energy Physics and Quantum Field Theory* (Protvino, USSR, 1984); Vol. I, p. 99; in *Proceedings of the Workshop on Constraint's Theory and Relativistic Dynamics*, Firenze, 1986, edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987), p. 210.
[50]H. Rund, *The Differential Geometry of Finsler Spaces* (Springer, Berlin, 1959).
[51]R. Giachetti and E. Sorace, Lett. Nuovo Cimento **26**, 1 (1979); A. Barducci, R. Giachetti, and E. Sorace, Lett. Math. Phys. **8**, 105 (1984); G. Marmo, N. Mukunda, and E. C. G. Sudarshan, Phys. Rev. D **30**, 2110 (1984); A. P. Balachandran, G. Marmo, and A. Stern, Nuovo Cimento A **69**, 175 (1982); J. Martin and J. L. Sanz, J. Math. Phys. **19**, 780 (1978).
[52]L. Brink, S. Deser, B. Zumino, P. di Vecchia, and P. Hove, Phys. Lett. B **64**, 435 (1976).
[53]T. F. Jordan, Phys. Rev. **166**, 1308 (1968).
[54]R. N. Hill and E. H. Kerner, Phys. Rev. Lett. **17**, 1156 (1966).
[55]M. Henneaux and L. C. Shepley, J. Math. Phys. **23**, 2101 (1982).
[56]V. V. Dodonov, V. I. Man'ko, and V. D. Skarzhinsky, Hadronic J. **4**, 1734 (1981).
[57]L. Lusanna and B. Milewski, Nucl. Phys. B **247**, 396 (1984).
[58]H. Feshback and F. Villars, Rev. Mod. Phys. **30**, 24 (1984).
[59]L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).

Summable chains of instantons and their symmetries

Yves Brihaye
Université de Mons, Physique Théorique et Mathematique, B-7000 Mons, Belgium
Jutta Kunz
NIKHEF-K, P. O. Box 41882, NL-1009 DB Amsterdam, The Netherlands

(Received 7 December 1988; accepted for publication 5 April 1989)

Among its solutions, the self-duality equations of SU(2) Yang–Mills theory possess configurations called *summable chains of instantons*. The symmetry properties of these solutions are studied and it is shown that they are invariant under an $O(3) \otimes O(2)$ subgroup of the conformal group of Euclidian space-time. This exceptional invariance is exploited in order to establish new summability formulas and to characterize the action densities of these classical solutions of the SU(2) gauge theory.

I. INTRODUCTION

The problem of finding all the finite action, classical solutions of the SU(2) Yang–Mills theory[1] has provided an interesting challenge for mathematicians and physicists. A considerable number of solutions have been obtained in Euclidian space-time by solving the stronger conditions of self-duality,[2] which lead to a system of first-order equations.

The most prominent self-dual solution is the instanton of Belavin, Polyakov, Schwartz, and Tyupkin.[3] Rapidly, a superposition principle for instantons was discovered, thanks to the beautiful ansatz invented by Corrigan, Fairlie, t'Hooft, and Wilczek.[4] For any integer N (labeling the homotopy class), this ansatz allows us to construct a family of solutions depending on $5N+4$ real parameters for $N>2$ (this number is, respectively, 5, 13 in the cases $N=1$ and $N=2$, see Ref. 5), while it is known that the most general solution depends on $8N-3$ parameters.[6] The construction of the extra solutions requires more elaborated techniques.[7]

Among the solutions available in the Corrigan, Fairlie, t'Hooft, Wilczek ansatz, some of them, called summable chains of instantons, enjoy very remarkable properties studied by Boutaleb-Joutei, Chakrabarti, and Comtet in Ref. 8. These authors have shown, namely, that (a) the summable chains of instantons arise naturally in the study of static solutions of the Yang–Mills equations considered in a de Sitter space; (b) the Green's functions for such instantons background can be given in a compact form; (c) in the large N limit they converge to the Prasad–Sommerfield static monopole solution (see also Ref. 9); and (d) they constitute good starting points for generating new self-dual solutions by the technique of Backlund transformations (see Ref. 10 and sources cited therein).

In this paper, we shall derive some new properties of the summable chains of instantons of Ref. 8 and discuss in details their relations with another class of summable solutions introduced in Ref. 11 and reconsidered recently in Ref. 12.

The next two sections are devoted to a summary of the basic equations of the ansatz and a discussion of the solutions that can be qualified as summable chains. In Sec. IV, we demonstrate that these configurations of the gauge fields are invariant under an $O(3) \otimes O(2)$ subgroup of the group of conformal transformations in Euclidian space-time. We derive new summability formulas in Sec. V and exploit them in Sec. VI to characterize the action densities associated with some of these classical solutions.

II. GENERALITIES

The action density of the SU(2) Yang–Mills theory reads

$$L = -\tfrac{1}{2}\,\mathrm{Tr}\, F_{\mu\nu} F_{\mu\nu}, \tag{2.1}$$

$$F_{\mu\nu} = \partial_\nu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu], \tag{2.2}$$

where the gauge potentials A_μ are elements of the algebra of the gauge group, SU(2) in this case. All the solutions that we will study here can be obtained in the Corrigan, Fairlie, t'Hooft, and Wilczek ansatz.[4] The gauge potentials assume the form

$$A^a_\mu(x) = \eta^{(\pm)a}_{\mu\nu}\partial_\nu \ln\rho \quad (a=1,2,3), \tag{2.3}$$

$$\eta^{(\pm)a}_{jk} = \varepsilon_{ajk}, \quad \eta^{(\pm)a}_{k4} = \pm\delta_{ak} \quad (j,k=1,2,3), \tag{2.4}$$

for which the self-duality equation

$$\widetilde{F}_{\mu\nu} \equiv \tfrac{1}{2}\varepsilon_{\mu\nu\rho\sigma}F_{\rho\sigma} = \pm F_{\mu\nu} \tag{2.5}$$

reduces to a scalar equation for the potential ρ:

$$(1/\rho)\Delta\rho = 0, \quad \Delta \equiv \partial^\mu\partial_\mu. \tag{2.6}$$

This equation admits among its solutions the functions below labeled by an integer N and by $5N+4$ real constants[12] [an overall normalization factor is irrelevant from the choice (2.3)]:

$$\rho(x) = \sum_{i=0}^{N} \frac{\lambda_i}{|x-k^{(i)}|^2}, \tag{2.7a}$$

$$\lambda_i > 0, \quad k^{(i)} \neq k^{(j)}, \quad \text{for } i \neq j. \tag{2.7b}$$

In this formalism, the action density is given by

$$L = -\tfrac{1}{2}\Delta\Delta \ln\rho, \quad \text{at } x \neq k^{(i)}. \tag{2.8}$$

It is well known that the singularities at $x=k^{(i)}$ in Eq. (2.7) are gauge artifacts and can be eliminated by a suitable gauge transformation[13]; in fact, only the numerator of ρ, say P, enters in the action density of these configurations;

$$L = -\tfrac{1}{2}\Delta\Delta \ln P, \tag{2.9a}$$

$$L = Q/P^4. \tag{2.9b}$$

From the requirement $\lambda_i > 0$, it is easy to see that P is positive. Naively, Q is a polynomial of degree $8N-4$ but, be-

 0022-2488/89/081913-05$02.50

cause of the properties of the operator $\Delta\Delta$, it appears to be of degree $8N-8$ only, insuring the integrability of the action density (2.9) over space-time:

$$\int L\, d^4x = 8\pi^2 N. \tag{2.10}$$

III. SUMMABLE CHAINS

We will mainly focus our attention on the two classes of potentials given by (from now on $N \equiv n-1$)

$$\rho_{\rm I}(n,x) = \sum_{k=0}^{n-1} \frac{\sec^2\theta_k}{[(x_0-\tan\theta_k)^2 + x_1^2 + x_2^2 + x_3^2]}, \tag{3.1}$$

$$\rho_{\rm II}(n,y) = \sum_{k=0}^{n-1} \frac{1}{[(y_0-\sin{}^2\theta_k)^2 + y_1^2 + y_2^2 + (y_3-\cos 2\theta_k)^2]}, \tag{3.2}$$

with

$$\theta_k = k\pi/n + \varepsilon, \quad 0 \leqslant \varepsilon < \pi/n. \tag{3.3}$$

The parameter ε plays no role in the solutions, indeed it is related to the gauge invariance that occurs in the ansatz (2.7) when all the positions $k^{(i)}$ are chosen on a line or on a circle (Ref. 5); it is, however, useful to keep it in the equations.

The solutions (3.1) and (3.2) have been considered previously in different contexts.

(i) For $\varepsilon = 0$ the potential of Eq. (3.1) is the one used by the authors of Ref. 8, where they introduce the notion of a summable chain of instantons [see their Eq. (2.20)]. Their alternative formula [Eq. (2.19)] can also be reproduced from Eq. (3.1) above with the choice $\varepsilon = 0$ (resp. $\varepsilon = \pi/2n$) for even (resp. odd) values of n.

(ii) The solutions associated with the potential (3.2) are a subset of a large class of self-dual gauge fields invariant under an $O(2) \otimes O(2)$ subgroup of $O(4)$, the group of space-time transformations. This kind of invariance was investigated in Ref. 11, where the solutions are constructed in the framework of the Yang self-duality equations.[2] The potential (3.2) was considered more recently by Chakrabarti for the construction of hyperbolic monopoles (Ref. 12).

By using the transformation laws of a scalar density under the conformal group, one can show that the functions (3.1) and (3.2) are related to each other through the following mapping:

$$x_\mu = 2P_{\mu\nu}[(y_\nu + a_\nu)/|y+a|^2] + a_\mu, \tag{3.4}$$

with

$$a = (0,0,0,1), \quad P_{\mu\nu} = \delta_{\mu\nu} - 2a_\nu a_\mu. \tag{3.5}$$

IV. SYMMETRIES

The function (3.1) is invariant under the subgroup SO(3) of rotations in the subspace x_1,x_2,x_3, supplemented by the reflections of the four coordinate axes; there are, however, many other functions that have the same invariance in the family defined in Eq. (2.7). So, the question arises: what distinguishes the summable chains of instantons in the subclass of $O(3) \otimes Z_2$ invariant solutions?

On the other hand, the solutions of Eq. (3.2) are invariant under an $O(2) \otimes O(2)$ subgroup of the group of space-time rotations. One of these, $O(2)$, acts on the x_1,x_2 subspace while the other one, acts on the x_0,x_3 subspace; this last rotation has to be compensated by an (inessential) translation of the parameter ε.

All invariances discussed so far are linearly realized; however, the correspondence of the functions (3.1) and (3.2) through the conformal mapping (3.4) allows us to carry the invariances of the first function into the other one and vice versa. As a consequence, the two classes of solutions under consideration are in fact invariant under (at least) an $O(3) \otimes O(2)$ subgroup of the conformal group of Euclidian space-time. The two subgroups $O(3) \otimes Z_2$ and $O(2) \otimes O(2)$ are the linearly realized parts of this bigger transformation group, respectively, in the cases I and II.

More explicitly, in the case I the solutions are invariant under the set of transformation defined by

$$x'_\mu = R_{\mu\nu} x_\nu, \tag{4.1}$$

$$X'_\mu = V_{\mu\nu}\left[\frac{X_\nu - (a-b)_\nu X^2}{1 - 2((a-b)\cdot X) + |a-b|^2 X^2}\right], \tag{4.2}$$

with the definitions

$$X'_\mu \equiv (x'_\mu + a_\mu)/2, \quad X_\mu \equiv (x_\mu + a_\mu)/2, \tag{4.3a}$$

$$a = Vb, \tag{4.3b}$$

and where R denotes an orthogonal matrix acting on the subspace x_1,x_2,x_3 and leaving the subspace x_0 invariant, while V denotes an orthogonal matrix acting on the x_0,x_3 subspace and leaving the directions x_1 and x_2 invariant.

With the same conventions, the invariances in case II read

$$y'_\mu = V_{\mu\nu} y_\nu, \tag{4.4}$$

$$Y'_\mu = R_{\mu\nu}\left[\frac{Y_\nu + (a-b)_\nu Y^2}{1 + 2((a-b)\cdot Y) + |a-b|^2 Y^2}\right], \tag{4.5}$$

with

$$Y'_\mu \equiv (y'_\mu - a_\mu)/2, \quad Y_\mu \equiv (y_\mu - a_\mu)/2, \quad a \equiv Rb. \tag{4.6}$$

Let us finally observe that if the matrix V (resp. R) in Eq. (4.2) [resp. (4.5)] is such that $b = a$, then the transformation is linear; if, on the other hand, it is chosen as diag$(1,1,1,-1)$ (i.e., so that $b = -a$), the transformation (4.2) [resp. (4.5)] takes the form

$$x'_\mu = x_\mu/|x|^2, \quad \text{resp. } y'_\mu = y_\mu/|y|^2; \tag{4.7}$$

therefore the asymptotic behavior of the summable chains solutions is related in a simple way to their Taylor expansion around the origin.

We guess that, among all configurations available in the ansatz (2.3)–(2.7), the summable chains of instantons are the ones that have the largest invariance under conformal transformations.

V. RECURRENCE RELATIONS AND SUMMABILITY

We now use the conformal mapping (3.4) to carry some properties of the type II family to the case I. Since we are mainly interested in the action density, we have to study the numerators of the functions $\rho_{\rm I}$ and $\rho_{\rm II}$. Therefore let us define the sequence of polynomials K_n as the numerators of the rational functions $\rho_{\rm II}$,

$$K_n = \frac{1}{n+1}\rho_{\rm II}(n+1,y)\prod_{k=0}^{n}[(y_0-\sin 2\theta_k)^2 + y_1^2 + y_2^2 + (y_3-\cos 2\theta_k)^2]. \quad (5.1)$$

Notice that the definition is such that K_n is of degree $2n$ in the space-time variables.

Using the formalism of Ref. 11, one can prove that these polynomials depend only on the variables u and v, defined as

$$u = y_0^2 + y_3^2, \quad v = y_1^2 + y_2^2 \quad (5.2)$$

and that they obey the following recurrence relation:

$$K_{n+2} = (1+u+v)K_{n+1} - uK_n, \quad (5.3)$$

$$K_0 = 1, \quad K_1 = 1+u+v, \quad (5.4)$$

whose solution [with the initial conditions (5.4)] reads

$$K_n = \frac{1}{\delta}\left[\left(\frac{1+u+v+\delta}{2}\right)^{n+1} - \left(\frac{1+u+v-\delta}{2}\right)^{n+1}\right], \quad (5.5)$$

$$\delta^2 \equiv (1+u+v)^2 - 4v. \quad (5.6)$$

Since the functions $\rho_{\rm I}$ and $\rho_{\rm II}$ are related through a conformal transformation, one expects similar recurrence relations in the case I as well. This is indeed the case; defining the numerator of $\rho_{\rm I}$ as follows:

$$P_n(r^2,t^2) = \frac{4^n}{n+1}\rho_{\rm I}(n+1,x) \times \prod_{k=0}^{n}[(x_0-\tan\theta_k)^2 + r^2]\cos^2\theta_k, \quad (5.7a)$$

$$r^2 \equiv x_1^2 + x_2^2 + x_3^2, \quad t^2 \equiv x_0^2, \quad (5.7b)$$

one obtains, after some algebra, the analog of Eq. (5.3) for the P_n,

$$P_{n+2} = 2(1+r^2+t^2)P_{n+1} - (1+2t^2-2r^2+(r^2+t^2)^2)P_n, \quad (5.8)$$

$$P_0 = 1, \quad P_1 = 2(1+r^2+t^2), \quad (5.9)$$

which allows to sum up these polynomials,

$$P_n = (1/4r)[(t^2+(1+r)^2)^{n+1} - (t^2+(1-r)^2)^{n+1}]. \quad (5.10)$$

The formulas (5.5) and (5.10) above are the counterparts of the summability formula (2.22) presented in Ref. 8 for the full potentials $\rho_{\rm I}$. In our case, only the numerators of the potentials are summed, i.e., the polynomials relevant for the gauge invariant quantities. We have obtained the generalization of the formula (2.22) of Ref. 8 for arbitrary values of ε:

$$\rho_{\rm I}(n,x) = n\left[\frac{\cot n\omega - \cot n\omega^*}{\cot\omega - \cot\omega^*}\right], \quad (5.11)$$

$$\cot\omega = t(\theta) + ir(\theta), \quad (5.12)$$

where $\theta = \varepsilon$ (resp. $\varepsilon + \pi/2n$) for even (resp. odd) values of n and where the quantities $t(\theta)$ and $r(\theta)$ are defined by

$$t(\theta) = \frac{(t^2+r^2-1)\sin 2\theta + 2t\cos 2\theta}{1+r^2+t^2-(t^2+r^2-1)\cos 2\theta + 2t\sin 2\theta}, \quad (5.13a)$$

$$1 + r^2(\theta) + t^2(\theta) = \frac{2(1+r^2+t^2)}{1+r^2+t^2-(t^2+r^2-1)\cos 2\theta + 2t\sin 2\theta}. \quad (5.13b)$$

Similarly, the potential $\rho_{\rm II}$ can be summed over and the formula reads

$$\rho_{\rm II}(n,y) = (n/2i)\left\{\frac{\cot n\omega - \cot n\omega^*}{[(u+v)^2+2u-2v+1]^{1/2}}\right\} \quad (5.14)$$

with

$$\cot\omega = \frac{2(y_0\cos\varepsilon - y_3\sin\varepsilon) + i[(u+v)^2+2u-2v+1]^{1/2}}{u+v+2(y_3\cos\varepsilon + y_0\sin\varepsilon)+1}. \quad (5.15)$$

We observe that the summability formulas relative to the potentials involve multiple trigonometric functions, while the ones for the numerators are written in terms of rational functions.

VI. ACTION DENSITY

The knowledge of the action density is related to the computation of Q defined in Eq. (2.9); we have computed these polynomials in some cases with the help of the algebraic program REDUCE and obtained the following expressions:

Case I:

$$L(n=2,r,t) = 2^4[3/(1+R^2)^4], \quad (6.1)$$

$$L(n=3,r,t) = 2^{10}[Q_2/(3R^4+10r^2+6t^2+3)^4], \quad (6.2)$$

$$L(n=4,r,t) = L(n=2,r,t) + 2^7[\tilde{Q}_3/(R^4+6r^2+2t^2+1)^4] \quad (6.3)$$

$$Q_2 = 27R^8 + 12R^4(r^2+9t^2) + 162t^4 + 42r^2t^2 - 14r^4 + 12(r^2+9t^2) + 27 \quad (6.4)$$

$$\tilde{Q}_3 = 9R^8 + R^4(36t^2-20r^2) + 54t^4 - 4r^2t^2 - 42r^4 + 36t^2 - 20r^2 + 9, \quad (6.5)$$

with the definition

$$R^2 \equiv r^2 + t^2. \tag{6.6}$$

The case $n = 2$, given for completeness, is the famous instanton, in fact, invariant under $O(4)$ transformations. Cases $n = 3$ and 4 are more specific to our analysis, the functions of Eq. (6.2) and (6.3) develop a unique maximum at the origin and decrease monotonically when one moves away from this point. This statement is confirmed by the Taylor expansion of the action density around the origin:

$$L(n+1,r,t) = \tfrac{16}{3} n^2(n+2)^2[1 - 4t^2 - 4\beta_n r^2 + O(R^4)], \tag{6.7}$$

with

$$\beta_n \equiv (4n^4 + 24n^3 + 45n^2 + 20n - 12)/9(n+2)^2. \tag{6.8}$$

It can be related to the asymptotic behavior of L, thanks to the symmetry property discussed in Eq. (4.7),

$$L(n+1,r,t) = \frac{16}{3}\frac{n^2(n+2)^2}{R^8}\left[1 - \frac{4}{R^4}(t^2 + \beta_n r^2) + O\left(\frac{1}{R^4}\right)\right]. \tag{6.9}$$

The occurrence of two terms in Eq. (6.3) is related to the factorization of the polynomial P_3.

As stressed in the Appendix, the polynomials P_n simplify considerably on the time axis ($r = 0$), i.e., on the line where the singularities have been chosen in Eq. (3.1); this property is at the origin of a very simple behavior of L on the time axis

$$L(n+1, r=0, t) = \tfrac{16}{3}[n^2(n+2)^2/(1+t^2)^4]; \tag{6.10}$$

it is, up to a multiplicative factor, the same profile as the one of the single instanton.

We now turn to a comparison of the action (6.2) with the one of the most general solution invariant under $O(3) \otimes Z_2$, in the case $n = 3$; in this purpose, we investigate in details the potential given by

$$\rho = 1 + \lambda/[3r^2 + (\sqrt{3}t - 1)^2] + \lambda/[3r^2 + (\sqrt{3}t + 1)^2], \tag{6.11}$$

where, for simplicity, we choose the scale so that the summable chain corresponds to $\lambda = 4$. The action density associated with this potential reads

$$L(r,t) = (64\lambda^2/P^4)[81R^8 + (1+2\lambda)^2 + 12(r^2 + 9t^2) \times (9R^4 + 1 + 2\lambda) + 36\lambda R^2(3t^2 - r^2) + 18(r^4 + 5r^2t^2 + 57t^4)], \tag{6.12}$$

$$P = 9R^4 + 6\lambda R^2 + 6(r^2 - t^2) + 2\lambda + 1. \tag{6.13}$$

For small values of λ there are two maxima appearing at

$$r = 0, \quad t = \pm t_m, \quad 0 \leqslant t_m \leqslant 1/\sqrt{3}, \tag{6.14}$$

while for large values (in fact, $\lambda > 2.5$) the maximum is unique and located at the origin; it is the case for the summable chain, $\lambda = 4$. It is only in the case of a summable chain that the symmetry (4.7) is present in the function (6.12) and that the polynomial P in Eq. (6.13) factorizes on the time axis, leading to the very simple formula (6.10).

Case II: For $n = 2$ the solution is again the single instanton; in the case $n = 3$ the energy density takes the form

$$L(n=3,u,v) = 16[(K_2^2 + 24uK_2 + 27u^2)/K_2^4], \tag{6.15a}$$

$$K_2 = 1 + u + 2v + (u+v)^2; \tag{6.15b}$$

this function attains its maximal value for $v = 0$, $u \simeq 0.3$ so that the action is concentrated around a circle in space-time. In our conventions, the circle is centered around the origin and located in the plane y_0, y_3, where the poles have been chosen in Eq. (3.2).

For higher values of n, the shape of L remains the same: the maximal value is attained in the plane $v = 0$ and for a value of u approaching 1 when n increases (the first few values are collected in Table I); when the variable u is fixed, the action density decreases monotonically with v.

The behavior of the action density around the origin reads

$$\begin{aligned} L(2,u,v) &= 48(1 - 4(u+v)) + O(R^4),\\ L(3,u,v) &= 16(1 + 22u - 4v) + O(R^4),\\ L(n,u,v) &= 16(1 + 4(u-v)) + O(R^4), \quad n > 3. \end{aligned} \tag{6.16}$$

Unlike Eq. (6.7), here we remark an absence of continuity in the formulas (constant terms and coefficients of u); this is related to the peculiar values of the coefficients of u^k in the polynomials K_n's [see Eq. (A5)].

In this case also, the simple form assumed by the polynomials P_n in the plane $u = 0$ can be exploited to compute the exact behavior of the action density in this plane;

$$L(n,0,v) = 16/(1+v)^4, \quad \text{for } n > 2; \tag{6.17}$$

this means, in particular, that, when one moves perpendicularly from the circle of maximal action density, one recovers, independently of n, the profile of the single instanton case.

Finally, we investigate the large n limit of $\rho_{\rm II}$ and of the action density and obtain the following results:

$$\rho_\infty(y) = \lim_{n\to\infty} (1/n)\rho_{\rm II}(n,y) = [(1+u+v)^2 - 4u]^{1/2}, \tag{6.18}$$

$$L(\infty,u,v) = 16/[(1+u+v)^2 - 4u]^2. \tag{6.19}$$

As a consequence, the maximum of L for large but finite values of n is attained around $v = 0$, $u = 1$, where the function assumes the value

$$L(n+1, u=1, v=0) = \tfrac{1}{3}n^2(n+2)^2. \tag{6.20}$$

TABLE I. The numerical values for $u_{\max}$ (the coordinate of the maximal value of L in case II, $v_{\max} = 0$), and the maximal value of this function for the first few values of n.

n	$u_{\max}$	$L_{\rm II}$
2	0.	48
3	0.30	61.
4	0.55	133.
5	0.68	251.
6	0.77	525.
7	0.83	924.
8	0.87	1524.

VII. CONCLUSIONS

Given a field theory with many symmetries like a gauge theory, it is likely that solutions having an exceptional invariance possess other interesting properties. In the family of self-dual gauge fields available in the Corrigan, Fairlie, t'Hooft, Wilczek ansatz, the summable chains of instantons are probably the configurations that have the largest invariance. In this respect, they merit a special attention.

The type II solutions [Eq. (3.2)] appear to be the only ones invariant under an $O(2) \otimes O(2)$ subgroup of $O(4)$, the group of linear transformations of Euclidian space-time. When investigated in the framework of the conformal group, the subgroup of invariance of these solutions is enlarged to $O(3) \otimes O(2)$ and these solutions can be mapped by a conformal transformation on the summable chains of instantons of Ref. 8, studied here as type I.

Each summable chain solution is mainly characterized by a polynomial that is gauge invariant and whose degree is equal to the topological (or winding) number of the solution; polynomials of solutions with successive winding numbers obey linear recurrence relations [Eqs. (5.3) and (5.8)]. As a mathematical aspect of our analysis, we noticed that summable chains provide explicit formulas for an infinite set of finite trigonometric series involving the angles θ_k. These results, summarized by Eqs. (A2) and (A6), are obtained in an unified way, the key ingredient being the recurrence relations.

We have exploited our new relations to evaluate a gauge invariant quantity (for instance the action density) associated with summable chains. These functions present nice factorization properties in some radial directions of space-time; there, for any values of the topological number n, they assume the same decay profile as the one of the single instanton.

APPENDIX: MORE ON THE POLYNOMIALS P AND Q

In this appendix we write explicitly the first few polynomials P_n and K_n and we present formulas to compute them, in general:

$$\begin{aligned} &P_0 = 1, \quad P_1 = 2(1 + r^2 + t^2), \\ &P_2 = 3 + 10r^2 + 6t^2 + 3(r^2 + t^2)^2, \\ &P_3 = 4 + 28r^2 + 12t^2 \\ &\qquad + 4(r^2 + t^2)(7r^2 + 3t^2) + 4(r^2 + t^2)^3, \\ &P_4 = 5 + 20(3r^2 + t^2) + 126r^4 + 140r^2t^2 + 30t^4 \\ &\qquad + 20(r^2 + t^2)^2(3r^2 + t^2) + 5(r^2 + t^2)^4, \end{aligned} \tag{A1}$$

$$P_n(r^2,t^2) = (n+1) \sum_{d=0}^{n} \sum_{k=0}^{d} A(n,d,k) t^{2k} r^{2d-2k} \tag{A2}$$

$$A(n,d,k) = C_n^d C_d^k \frac{(2n-2k+1)!!}{(2n-2d+1)!!(2d-2k+1)!!}. \tag{A3}$$

We remark that P_n nicely factorizes on the time axis:

$$P_n(0,t^2) = (n+1)(1+t^2)^2 \tag{A4}$$

and that P_1 can be factorized out of P_{2n+1} as a consequence of Eqs. (5.8) and (5.9).

The first few K_n read

$$\begin{aligned} &K_0 = 1, \quad K_1 = 1 + u + v, \\ &K_2 = 1 + u + 2v + (u+v)^2, \\ &K_3 = 1 + u + 3v + (u+3v)(u+v) + (u+v)^3, \\ &K_4 = 1 + u + 4v + u^2 + 6uv + 6v^2 \\ &\qquad + (u+4v)(u+v)^2 + (u+v)^4, \end{aligned} \tag{A5}$$

$$K_n(u,v) = \sum_{d=0}^{n} \sum_{k=0}^{d} C_d^k C_{n-k}^{n-d} u^k v^{d-k}. \tag{A6}$$

In this case also, the recurrence relation (5.3) implies that K_1 can be factorized out of K_{2n+1}; on the other hand, the general expression (A6) simplifies drastically when u or v are equal to zero:

$$\begin{aligned} &K_n(u,0) = 1 + u + u^2 + \cdots + u^n, \\ &K_n(0,v) = (1+v)^n. \end{aligned} \tag{A7}$$

[1] C. N. Yang and R. Mills, Phys. Rev. **96**, 191 (1954).
[2] C. N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[3] A. Belavin, A. Polyakov, A. Schwartz, and Yu. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[4] E. D. Corrigan and D. B. Fairlie, Phys. Lett. B **67**, 69 (1977); F. Wilczek, *Quark Confinement and Field Theory*, edited by D. Stump, and D. Weingarden (Wiley, New York, 1977); G. t'Hooft (private communication).
[5] R. Jackiw, C. Nohl, and C. Rebbi, Phys. Rev. D **15**, 1642 (1977).
[6] M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[7] M. F. Atiyah, V. A. Drinfeld, N. J. Hitchin, and Yu. Manin, Phys. Lett. A **65**, 185 (1978).
[8] H. Boutaleb-Joutei, A. Chakrabarti, and A. Comtet, Phys. Rev. D **23**, 1781 (1981).
[9] P. Rossi, Nucl. Phys. B **149**, 170 (1979).
[10] A. Chakrabarti, Phys. Rev. D **35**, 696 (1987).
[11] Y. Brihaye, Commun. Math. Phys. **86**, 509 (1982).
[12] A. Chakrabarti, J. Math. Phys. **27**, 340 (1986).
[13] J. J. Giambiagi and K. D. Rothe, Nucl. Phys. B **129**, 111 (1977).

Theta function representation of modular group for SU(n)

F. Ardalan,[a] H. Arfaei,[a] and M. Zarkesh
Center for Theoretical Physics and Mathematics, Atomic Energy Organization of Iran, P. O. Box 14155-1339, Tehran, Iran

(Received 25 October 1988; accepted for publication 15 February 1989)

Matrices of the representation of the modular group acting on the space of SU(n) level k Kac–Moody algebra character are calculated explicitly. Also, a simple form is obtained for the case $k = 1$, which is compared with the corresponding matrix elements for the character representation in the case of the toroidally compactified string theory.

I. INTRODUCTION

One of the fundamental requirements of string theory and two-dimensional conformal field theory is modular invariance of the partition function of the theory, which has attracted much attention recently.[1–5] In particular, two-dimensional conformal field theories with the Virasora central charge $c < 1$ have been classified,[2,6,7] and for the string theories compactified on the lattice of the simple Lie algebra, all the partition functions consistent with modular invariance have been found.[8] These latter theories are equivalent to the theories of strings moving on the group manifold of the corresponding lattice in the particular case of the Wies–Zumino–Witten index $k = 1$ (k is the central charge of the corresponding Kac–Moody algebra). For arbitrary k, classification of the modular invariant partition functions of the string moving on a group manifold is not known, however. The only known results are the complete classification in the case $G = \mathrm{SU}(2)$ with arbitrary k,[9] and the left-right symmetric solutions of Gepner and Witten[10] and solutions obtained from them with the aid of the automorphisms of the Kac–Moody algebra.[11,12] A main difficulty in the construction of modular invariant partition functions is the nontrivial behavior of the Kac–Moody character under modular transformations.

In this paper we will provide a simple expression of the generators of the modular group acting on the character of the Kac–Moody algebra with group $G = \mathrm{SU}(n)$ and arbitrary central charge k, and we will obtain an explicit form $k = 1$, comparing it with the similar result for the toroidally compactified string theory. In Sec. II we briefly consider modular transformations and their representation on the Kac–Moody characters for a general simple Lie group. In Sec. III an expression for $G = \mathrm{SU}(n)$ and arbitrary k is found. Section IV contains an even simpler result for $k = 1$ whch is then compared with the corresponding matrices for toroidally compactified string theory on the SU(n) lattice; the known $G = \mathrm{SU}(2)$ and arbitrary modular transformations are also rederived at the end.

II. MODULAR TRANSFORMATION OF KAC–MOODY CHARACTERS

We denote the Kac–Moody algebra corresponding to the simple finite-dimensional algebra G by $\overline{G}$, and its Cartan subalgebra by $\overline{H}$. Then the modular group, $\mathrm{SL}(2,z)/(1,-1)$, acts on $\overline{H}$ according to the following[13]:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}: \quad (\tau,\nu,t) \to \left(\frac{a\tau + b}{c\tau + d}, \frac{\nu}{c\tau,d}, t + \frac{c}{2}\frac{\nu^2}{c\tau + d}\right), \tag{2.1}$$

where $(\tau,\nu,t) \in \overline{H} = L_0 \oplus H \oplus C$. There L_0 is the Virasora generator, H a Cartan subalgebra of G, and C the central element of $\overline{G}$. In the following we will be concentrating on the generators S and T of the modular group,

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{2.2}$$

and their representation on the space of character χ_w of $\overline{G}$,

$$T: \chi_w \to \chi_w e^{\pi i[(w+\rho)^2/(k+g) - \rho^2/g]}, \tag{2.3}$$

$$S: \chi_w \to \sum_{w' \in \Lambda^*/(k+g)\Lambda} S_{ww'} \chi_{w'}, \tag{2.4}$$

where $S_{ww'}$ are the elements of the unitary matrix S expressed in terms of the various algebraic quantities defined subsequently,

$$S_{ww'} = C_1 \sum_{r \in W} \epsilon(r) e^{(-2\pi i/k+g)(w+\rho)\cdot r(w'+\rho)}. \tag{2.5}$$

In the above formula, W is the Weyl group of G, and g is the dual-Coexter number of G defined by $g\delta^{ab} = f^{abc} f^{bde}$, where f^{abc}'s are the structure constants. ρ is the Weyl vector of G defined by $2\rho \cdot \Gamma\alpha_i/\alpha_i^2 = 1$ for all simple roots α_i of G. $\epsilon(r)$ is the determinant of r. Λ is the root lattice of G and Λ^* its weight lattice. C_1 is a constant which we will not specify and is determined by the condition of unitarity of the S matrix.

Note that the characters χ_w are determined by the representation of $\overline{G}$ defined by the central charge k and the highest weight w in the G representation. However, as implied in the sum in Eq. (2.4), the highest weights w are restricted to $\Lambda^*/(k+g)\Lambda$. This, in turn, is equivalent to the condition

$$2(w \cdot \alpha/\alpha^2) \leqslant k, \quad \forall \alpha \text{ roots of } G. \tag{2.6}$$

In contrast to T, the matrix S as expressed in Eq. (2.5) is not accessible for most explicit calculations. In the case of $G = \mathrm{SU}(n+1)$, we will derive a manageable expression for S and simplify it further for $k = 1$ in Sec. IV.

III. THE MATRIX S FOR SU(n+1) AND ARBITRARY k

It is convenient to work in a basis for the weight lattice of SU($n+1$) consisting of the $n+1$ weights λ_i of the funda-

[a] Permanent address: Physics Department, Sharif University of Technology, P. O. Box 11365-8639, Tehran, Iran.

 0022-2488/89/081918-03$02.50

mental representation of SU$(n+1)$, in terms of which the n simple roots α_i are[14]

$$\alpha_i = \lambda_i - \lambda_{i+1}\,, \quad i = 1,\ldots,n; \tag{3.1}$$

and have simple inner products,

$$\lambda_i \cdot \lambda_j = -1/(n+1)\,, \quad i \neq j\,, \tag{3.2}$$

$$\lambda_i^2 = n/(n+1)\,, \quad i = 1,\ldots,n+1\,. \tag{3.3}$$

However, λ_i are dependent,

$$\sum_{i=1}^{n+1} \lambda_i = 0\,; \tag{3.4}$$

consequently in the expansion of weights in terms of λ_i an ambiguity may occur

$$w = \sum_{i=1}^{n+1} l_i \lambda_i\,, \tag{3.5}$$

which we will remove by demanding

$$l_{n+1} = 0\,. \tag{3.6}$$

We now proceed to calculate the matrix S of Eq. (2.5) in this basis. First, we notice that the set of weights w appearing in Eq. (2.5) is constrained by Eq. (2.6), which is equivalent to

$$w \cdot \alpha \leqslant k\,, \quad \alpha = \text{highest root of } G\,. \tag{3.7}$$

But, we have,[15]

$$\alpha = \lambda_1 - \lambda_{n+1}\,, \tag{3.8}$$

which leads to

$$l_1 \leqslant k\,. \tag{3.9}$$

Now in terms of the Dynkin indices of w, $a_i \equiv 2w \cdot \alpha_i / \alpha_i^2$,

$$l_i = \sum_{j=i}^{n} a_j\,. \tag{3.10}$$

Therefore the final form of the constraint on the representations appearing in Eq. (2.5), in terms of the Dynkin indices, is

$$\sum_{i=1}^{n} a_i \leqslant k\,, \tag{3.11}$$

and there are $\binom{n+k}{k}$ such representations. So S is a matrix of dimension $\binom{n+k}{k}$.

To calculate the exponents in Eq. (2.5), we use the identity for

$$\begin{aligned} a &= \sum_{i=1}^{n} a_i \lambda_i\,, \quad a \in \Lambda^*\,, \\ a \cdot b &= \sum_{i=1}^{n} a_i b_i - \frac{1}{n+1}\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{n} b_j\right), \end{aligned} \tag{3.12}$$

and note that the Weyl group W of SU$(n+1)$ is the permutation group of the $n+1$ vectors λ_i.[15] Thus S becomes

$$S_{ww'} = C\eta^{-[1/(n+1)][\Sigma_{i=1}^{n}(w+\rho)_i][\Sigma_{j=1}^{n}(w'+\rho)_j]} \times \sum_{r\in W} \epsilon(r) \prod_{i=1}^{n+1} \eta^{(w+\rho)_i [r(w'+\rho)]_i}\,, \tag{3.13}$$

where

$$\eta := e^{-2\pi i/(k+n+1)}\,. \tag{3.14}$$

However,

$$\sum_{r\in W} \epsilon(r) \prod_{i=1}^{n+1} \eta^{(w+\rho)_i [r(w'+\rho)]_i} = \det T^{(n+k)} \begin{pmatrix} d_1,\ldots,d_n,0 \\ d_1',\ldots,d_n',0 \end{pmatrix}, \tag{3.15}$$

the determinant of the k minor of the matrix $T^{(n+k)}$ at the intersection of the rows labeled by $(d_i,\ldots,d_n,0)$ and the columns labeled by $(d_1',\ldots,d_n',0)$. Here the matrix $T^{(N)}$ is defined by

$$(T^{(N)})_{ij} = \eta^{ij}\,, \quad 0 \leqslant i,j \leqslant N\,, \tag{3.16}$$

and $d_i = (w+\rho)_i$, $d_i' = (w'+\rho)_i$ are related to the Dynkin indices by

$$d_i = \sum_{j=i}^{n} a_i + n + 1 - i \tag{3.17}$$

similarly for d_i'. The final form for S is then

$$S_{ww'} = C\eta^{-[1/(n+1)](\Sigma_i d_i)(\Sigma_j d_j')} \times \det T^{(n+k)} \begin{pmatrix} d_1,\ldots,d_n,0 \\ d_1',\ldots,d_n',0 \end{pmatrix}, \tag{3.18}$$

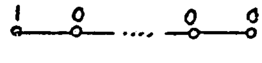
which is an expression amenable to direct calculations much more readily than the original expression of Eq. (2.5). In the next section we will further simplify Eq. (3.18) for the case $k = 1$.

IV. MATRIX S FOR SU(n+1) AND k=1

In this section we will obtain a particularly simple expression for the case $k = 1$ and compare it with the matrix S for the equivalent theory of a toroidally compactified string.

When $k = 1$, the representations of G appearing in the Kac–Moody representation, i.e., satisfying the condition $\Sigma_{i=1}^{n} a_i \leqslant k$, are just the basic representations of SU$(n+1)$, i.e., representations which have one Dynkin index equal to unity and the rest equal to zero,

It is convenient to label the representations by the location of the 1 in their Dynkin diagram. Thus we will assign the number $\mu = 0,\ldots,n$ to the μth basic representation, denoting the singlet representation by zero. Incidentally, as it is well known,[14] the μth basic representation is in fact the totally antisymmetric tensor product of rank μ built out of the fundamental representation



In this notation, it is easily seen that for $\mu \neq 0$ the Dynkin indices of the highest weight $w^{(\mu)}$ of the μth basic representation are $a_i = \delta_{i\mu}$, $d_i^\mu = n + 2 - i$ for $i \leqslant \mu$, and

$$d_i^\mu = n + 1 - i \quad \text{for } i > \mu\,. \tag{4.1}$$

Therefore, for each $\mu \neq 0$, d_i^μ's are n different numbers from the set $1,\ldots,n+1$ excepting the number $n + 1 - \mu$. Thus the subdeterminant

$$T^{(n+1)}\begin{pmatrix} d_1,\ldots,d_n,0 \\ d'_1,\ldots,d'_n,0\end{pmatrix}$$

appearing in the expression Eq. (3.18) for $S_{\mu\nu}$ reduces to $T^{(n+1)}_{\langle n+1-\mu,n+1-\nu\rangle}$, the determinant obtained from $T^{(n+1)}$, by crossing out the $(n+1-\mu)$th row and $(n+1-\nu)$th column. Therefore

$$S_{\mu\nu} = C_1\eta^{(-1/n+1)[(n(n+1)/2)+\mu][(n(n+1)/2)+\nu]} \times \det T^{(n+1)}_{\langle n+1-\mu,n+1-\nu\rangle} \tag{4.2}$$

where we have used the relation

$$\sum_{i=1}^{n} d^{\mu}_i = \frac{n(n+1)}{2} + \mu . \tag{4.3}$$

There remains to calculate the above subdeterminant which we will show to be

$$\det T^{(n)}_{\langle i,j\rangle} = [1/(n+1)] \det T^{(n)}(-1)^{i+j}\eta^{ij}. \tag{4.4}$$

To prove this, recall that

$$[T^{(n)}]_{ij} = \eta^{ij}, \quad \eta = \exp(-2\pi i/(n+2)), \quad i,j = 0,1,\ldots,n . \tag{4.5}$$

[In the rest of the proof we drop the superscript (n) from $T^{(n)}$.] Then, we have the following equations readily verified:

$$T^{-1} = [1/(n+1)]T^*, \tag{4.6}$$

$$T^4 = (n+1)^2 1, \tag{4.7}$$

$$\det T = i^{[n(3n+1)/2]}(n+1)^{(n+1)/2} . \tag{4.8}$$

Finally, to get the det $T_{\langle i,j\rangle}$, notice that

$$(T^{-1})_{ij} = (1/\det T)(-1)^{ij}\det T_{\langle i,j\rangle} . \tag{4.9}$$

Therefore

$$[1/(n+1)]\eta^{-ij} = (1/\det T)(-1)^{i+j}\det T_{\langle i,j\rangle} , \tag{4.10}$$

which leads to Eq. (4.4). Substituting these results in the expression (4.2) for $S_{\mu\nu}$, we get

$$S_{\mu\nu} = [1/(\sqrt{n+1})]e^{[2\pi i/(n+1)]\mu\nu}, \quad \mu,\nu = 0,\ldots,n , \tag{4.11}$$

where we have used the unitarity of $S_{\mu\nu}$ to fix the proportionality constant.

This final $k=1$ result for S can now be compared with the corresponding matrix in the string theory compactified on the torus of the SU$(n+1)$ lattice. There,[8] S obtained from a simple application of the Poisson resummation formula reads

$$S_{ww'} \sim e^{2\pi i w\cdot w'} , \tag{4.12}$$

where w and w' are the highest weights in the respective conjugacy classes of the weight lattices of the group SU$(n+1)$. To get Eq. (4.11) from (4.12), note that from Eqs. (3.2), (3.3), and (3.10), we have

$$w\cdot w' = \mu n - \mu\nu/(n+1) , \tag{4.13}$$

resulting in

$$e^{2\pi i w\cdot w'} = e^{-[2\pi i/(n+1)]\mu\nu}$$

in agreement with Eq. (4.11), obtained from Kac–Moody character transformation. This equivalence is of course to be expected from the celebrated equivalence in the Frenkel–Kac construction of vertex operators.[16]

It should be mentioned in passing that the modular generator T of Eq. (2.3) can also be obtained from the equivalent toroidally compactified string, where it has the form[8]

$$T = e^{i\pi w^2} . \tag{4.14}$$

Again it can easily be seen that

$$(w^{(\mu)})^2 = \mu - \frac{\mu^2}{n+1}, \tag{4.15}$$

$$\frac{(w^{(\mu)}+\rho)^2}{n+2} = \mu - \frac{\mu^2}{n+1} + \frac{n(n+1)}{12}; \tag{4.16}$$

consequently the two prescriptions agree within a constant phase.

In conclusion, we calculate the matrix S from Eq. (3.18) for SU(2) and arbitrary k. There each of the subdeterminants $T^{(k+1)}\binom{d_1,0}{d'_1,0}$, where $d_1 = a_1+1$, $d'_1 = a'_1+1$ is of the form

$$\det T^{(k+1)}\begin{pmatrix} d_1,0 \\ d'_1,0\end{pmatrix} = \det\begin{pmatrix} 1 & 1 \\ 1 & \eta^{d_1d'_1}\end{pmatrix}, \quad \eta = e^{-2\pi i/(k+2)} \tag{4.17}$$

and we get for S:

$$S_{a_1a'_1} = C_1\eta^{(-1/2)d_1d'_1}\det\begin{pmatrix} 1 & 1 \\ 1 & \eta^{d_1d'_1}\end{pmatrix},$$

so that

$$S_{a_1a'_1} \sim \sin[\pi/(k+2)](a_1+1)(a'_1+1) , \tag{4.18}$$

in agreement with the known results.[10]

ACKNOWLEDGMENT

The authors would like to thank Professor G. Khosrovshahi for useful conversations.

[1] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, Nucl. Phys. B **241**, 333 (1984).
[2] D. Friedan, A. Qiu, and S. Shenker, Phys. Rev. D **52**, 1575 (1984).
[3] J. L. Cardy, Nucl. Phys. B **270**, [FS16], 186 (1986).
[4] E. Verlinde, Utrecht preprint THU-88/17, 1988.
[5] C. Vafa, Harvard preprint HUTP-88/AU11, 1988.
[6] P. Goddard, A. Kent, and D. Olive, Commun. Math. Phys. **103**, 105 (1982).
[7] D. Gepner, Nucl. Phys. B **287**, 111 (1987).
[8] F. Ardalan and H. Arfaei, ICTP preprint IC/88/102, 1988.
[9] A. Cappelli, C. Itzykson, and J. B. Zuber, Nucl. Phys. B **280**, [FS18], 445 (1987).
[10] D. Gepner and E. Witten, Nucl. Phys. B **287**, 493 (1986).
[11] D. Bernard, Nucl. Phys. B **288**, 628 (1987).
[12] D. Bernard and J. Thierry-Meig, DAMTP/87-6, 1987.
[13] V. G. Kac and D. Peterson, Adv. Math. **53**, 125 (1984).
[14] B. G. Wybourn, *Classical Group for Physicists* (Wiley, New York, 1974).
[15] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978).
[16] J. Frenkel and V. G. Kac, Invent. Math. **62**, 23 (1980).

Cartan-preserving automorphisms of untwisted and twisted Kac–Moody algebras

N. Gorman and L. O'Raifeartaigh
Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland
W. McGlinn
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Received 8 February 1989; accepted for publication 8 March 1989)

The group $\widetilde{N}$ defined as the normalizer of the Cartan subalgebra in the group of all (inner and outer) automorphisms of affine Kac–Moody (KM) algebras is a natural extension of the Weyl group and is shown to play a fundamental role in the structure of these algebras (both untwisted and twisted). It is a (discrete) Galilean group which incorporates the affine and Weyl group structure of the KM algebra and the space-time structure of the string. It links the Virasoro and KM algebras in a nontrivial way and plays a key role in the "vertex" construction. Moreover, the elements of its covering group can be used to uniquely parametrize (and untwist) the pseudotwisted KM algebras.

I. INTRODUCTION

Experience has shown that the automorphism groups of physical systems and equations often transcend in importance the particular systems and equations to which they apply. For example, the three major automorphisms of classical electromagnetism—the Poincaré, conformal, and gauge automorphisms—have all become cornerstones of modern physics and apply to a much wider class of systems. For this reason the study of automorphism groups of new systems is important and in the present paper we wish to study an automorphism group of Kac–Moody (KM) algebras.[1,2] Of course, some of the automorphisms of KM algebras, namely the conformal (Virasoro) automorphisms[3] and, in the case of the string,[4] the Poincaré automorphisms, have already been well studied, but what we wish to consider here is a different group of automorphisms, namely the group of all automorphisms (both inner and outer) of the algebra which preserve the Cartan subalgebra. The inner automorphism part of this group is known in connection with the affine and Weyl structures of KM algebras, but here we wish to emphasize the importance of the group for the general structure of KM algebras, both untwisted and twisted (and for string theory), and to explore its properties in detail, in particular to distinguish between its inner and outer parts.

The group of all Cartan-preserving automorphisms, i.e., the normalizer of the Cartan subalgebra in the group of automorphisms, will be denoted by $\widetilde{N}$ and the subgroup of inner automorphisms will be denoted by N. The actions of $\widetilde{N}$ and N on the Cartan are determined by the quotient groups

$$\widetilde{W}=\widetilde{N}/C, \quad W=N/C, \tag{1.1}$$

respectively, where C is the group of automorphisms that leave each element of the Cartan separately invariant, i.e., the centralizer of the Cartan (which is automatically inner).

The corresponding groups for ordinary compact simple Lie algebras will be denoted by the same letters with the superscript zero and we note that in this case $\widetilde{W}^0$ and W^0 have a well-known geometrical interpretation, namely, as the symmetry group of the root diagram and its Weyl subgroup, i.e., the subgroup generated by reflections in planes orthogonal to the roots, respectively. In addition, one has

$$\mathscr{T}\equiv \mathrm{Aut}(G_0)/\mathrm{Int}(G_0)=\widetilde{N}^0/N^0=\widetilde{W}^0/W^0, \tag{1.2}$$

where $\mathscr{T}$, the group of "strictly" outer automorphisms of the Lie algebra, is also the symmetry group of the Dynkin diagram. Because coset representatives that form a group exist in all three cases in (1.2), the groups $\widetilde{N}$ and $\widetilde{W}$ actually have the semidirect product structures

$$\widetilde{N}^0=\mathscr{T}\wedge N^0, \quad \widetilde{W}^0=\mathscr{T}\wedge W^0. \tag{1.3}$$

The main results of our investigation of the groups $\widetilde{N}$, N, $\widetilde{W}$, and W are the following.

(i) The group $\widetilde{N}$ has the structure of a Galilean group, which is discrete in the sense that the usual parameters are quantized. The homogeneous (rotation) subgroup of $\widetilde{N}$ is just the corresponding group $\widetilde{N}^0$ for the ordinary Lie subalgebra G_0 of the KM algebra and the new characteristic KM feature is the existence of an inhomogeneous subgroup $\widetilde{A}$, which is a discrete version of the Galilean acceleration subgroup.

(ii) The result (1.2) for the homogeneous part $\widetilde{N}^0$ of $\widetilde{N}$ has the analog

$$\widetilde{A}/A=\mathscr{Z} \tag{1.4}$$

for the inhomogeneous part, where $\mathscr{Z}$ is the central subgroup of the simply connected covering group $\mathscr{G}_0$ of the algebra G_0. Combining (1.2) and (1.4), one obtains the KM analog of (1.2), namely

$$\mathscr{S}\equiv\mathscr{T}\wedge\mathscr{Z}=\widetilde{N}/N=\widetilde{W}/W, \tag{1.5}$$

where $\mathscr{S}$ is the symmetry group[5] of the extended Dynkin diagram. [There is no KM analog of (1.3).]

(iii) The covering group $\widetilde{A}_c$ of the acceleration subgroup $\widetilde{A}$ obtained by extending the parameters of $\widetilde{A}$ from their permitted (discrete) values to the continuum generates the full family of pseudotwisted KM algebras. In fact, $\widetilde{A}_c$ can be used to parametrize the pseudotwisted KM algebras uniquely and untwist them.

 0022-2488/89/081921-12$02.50

(iv) The quotient groups $\widetilde{W} = \widetilde{N}/C$ and $W = N/C$ again have a geometrical significance, namely as the symmetry group of the (Minkowskian) root diagram of the KM algebra and its Weyl subgroup, i.e., the subgroup generated by (Minkowskian) reflections in planes orthogonal to the KM roots.

(v) The action of $\widetilde{N}$ can be extended (uniquely) to include the Virasoro algebra V of the KM algebra; it then links the V and KM algebras in a manner that is reducible, but not fully reducible.

(vi) The elements of N play a key role in the so-called "vertex" construction[3] of the non-Abelian elements of the KM algebra.

(vii) The group $\widetilde{N}$ not only plays the role of an "internal" Weyl group, but in string theory, it also plays the role of an "external" space-time group. In particular, the generators of the acceleration subgroup $\widetilde{A}$ become the center-of-mass coordinates of the string, while the group $\widetilde{N}$ becomes that subgroup of the Poincaré group that is the little group of a lightlike vector in Minkowski space. Since such a subgroup is Galilean, this interpretation of $\widetilde{N}$ gives a physical explanation of its Galilean structure.

The results (i)–(vii) are established and discussed in detail for untwisted KM algebras in Secs. II–VIII. Their extension to twisted KM algebras is not trival and the main modifications that arise are discussed in Secs. IX–XI. One of the principal results of the paper is the derivation of the quantization conditions for the parameters of the Galilean acceleration subgroup. The permitted values of these parameters for the various cases—outer and inner automorphisms and untwisted and twisted algebras—are given in Table III. (The result (1.4) for all untwisted and twisted cases follows immediately from Table III.)

II. RECALL OF PROPERTIES OF COMPACT SIMPLE LIE ALGEBRAS

The simple, compact, simply connected Lie groups[6] will be denoted by $\mathscr{G}_0$ and their Lie algebras

$$[T^a, T^b] = f^{ab}{}_c\, T^c, \quad a,b,c = 1,2,...,n, \tag{2.1}$$

will be denoted by G_0. Here the $f^{ab}{}_c$ are real totally antisymmetric structure constants, uniformly normalized so that (by Schur's lemma) they satisfy

$$f^{ac}{}_d\, f^{db}{}_c = Q\delta^{ab}, \tag{2.2}$$

where Q is a constant which depends only on the group. The Cartan form of the algebra will be written in the usual way as

$$[H^i, E^\alpha] = \alpha^i E^\alpha, \quad [E^\alpha, E^{-\alpha}] = (2/\alpha^2)(\alpha\cdot H), \quad i = 1,...,l, \tag{2.3}$$

where l is the rank, the H's commute, and there are further relations between the E^α's which will not be needed here. From (2.2) the roots α evidently satisfy the completeness relation

$$\sum_\alpha \alpha^i \alpha^j = Q\delta^{ij}, \tag{2.4}$$

where α^i are the components of α.

The roots are characterized by the property that for any two roots α, β,

$$2(\alpha\cdot\beta)/\beta^2 \in \mathbb{Z}, \tag{2.5}$$

where $\mathbb{Z}$ denotes the set of all integers; this leads one to introduce *coroots* $\tilde{\alpha}$ defined by

$$\tilde{\alpha} = 2\alpha/\alpha^2, \text{ so that } (\tilde{\alpha}\cdot\beta)\in\mathbb{Z}. \tag{2.6}$$

The coroots are obviously parallel to the roots and since they in turn satisfy (2.5) they are the roots of an algebra $\widetilde{G}_0$ called the dual algebra. However, the dual algebra is actually the algebra itself (G_0 is self-dual) in all cases except $SO(2n+1)$ and $Sp(2n)$, which are dual to each other.

The roots and coroots generate (by addition and subtraction) infinite lattices called the root and coroot lattices Γ_α respectively. Furthermore, to each set of roots and coroots there corresponds a dual lattice called the weight and coweight lattices Γ_w and $\Gamma_{\tilde{w}}$, respectively, and defined as the set of all weights w and coweights $\tilde{w}$ satisfying

$$(\tilde{\alpha}\cdot w)\in\mathbb{Z}, \quad (\alpha\cdot\tilde{w})\in\mathbb{Z}, \tag{2.7}$$

respectively. From (2.5) it is clear that roots and coroots are special cases of weights and coweights, respectively, but except for $G_0 = E_8$ the converse is not true. Thus in general,

$$\Gamma_w \supset \Gamma_\alpha, \quad \Gamma_{\tilde{w}} \supset \Gamma_{\tilde{\alpha}}. \tag{2.8}$$

Note that α and w scale in the same way and opposite to $\tilde{\alpha}$ and $\tilde{w}$. The algebras for which the roots have the same length are called simply laced algebras and for the other algebras (for which the roots can have only two distinct lengths) it will be convenient to denote the long and short roots by λ and σ, respectively. For the root and coroot lattices of simply laced algebras we have the equality

$$\Gamma_\alpha = (\alpha^2/2)\Gamma_{\tilde{\alpha}} \tag{2.9}$$

and for the corresponding lattices of general algebras we have the inclusions

$$(\lambda^2/2)\Gamma_{\tilde{\alpha}} \subseteq \Gamma_\alpha \subseteq (\sigma^2/2)\Gamma_{\tilde{\alpha}}, \tag{2.10}$$

which will be useful later.

As is well known, the positive weights w are the highest weights of the irreducible representations of the algebras and fall naturally into congruency classes [the generalizations of the SU(3) triality classes], where two weights w and w' are said to be congruent ($w\sim w'$) if and only if they differ by an element of the root lattice:

$$w\sim w' \Leftrightarrow w - w' \in \Gamma_\alpha. \tag{2.11}$$

The lowest positive weight $w^{(g)}$ in each congruency class will be called a fundamental weight and the representation with $w^{(g)}$ as the highest weight will be called a fundamental representation.

It is easy to see that the group elements

$$\exp 2\pi i(\tilde{w}\cdot H) \tag{2.12}$$

of $\mathscr{G}_0$ are central and it is not difficult to show that all elements of the center are of this kind. Also, since

$$\exp 2\pi i(\tilde{\alpha}\cdot H) = 1, \tag{2.13}$$

the central elements corresponding to the two coweights $\tilde{w}_1$ and $\tilde{w}_2$ are the same if and only if the coweights are congruent (modulo $\tilde{\alpha}$). This shows that the order p of the center is just the number of congruency classes of the dual group and hence of the group itself. Equation (2.12) also shows that the irreducible representations fall naturally into congruency classes, each characterized by a fundamental representation. Since $p = l + 1$ only for SU $(l+1)$, but $p \leqslant 4$ otherwise, the number of fundamental representations is $\leqslant 4$, except for SU$(l+1)$, in which case the fundamental representations are just the trivial representation and the l-primitive (totally antisymmetric) tensor representations.

A crucial group for our considerations will be the group $\widetilde{N}^0$ of all automorphisms of the Lie algebra G_0 that preserve the Cartan subalgebra. It can be shown that the group $\widetilde{N}^0$ is actually a semidirect product of the form

$$\widetilde{N}^0 = \mathscr{T} \wedge N^0, \tag{2.14}$$

where N^0 is the inner part of $\widetilde{N}^0$ and $\mathscr{T}$ is the group $\mathrm{Aut}(G_0)/\mathrm{Int}(G_0)$ of "strictly outer" automorphisms of G_0. The action of N^0 on the Cartan subalgebra is by definition implemented by the quotient group

$$\widetilde{W}^0 = \widetilde{N}^0/C^0, \tag{2.15}$$

where C^0 is the centralizer of the Cartan subalgebra in $\widetilde{N}^0$ and, since the roots are just the spectra of the base elements of the Cartan, it follows that $\widetilde{W}^0$ is the symmetry group of the root diagram. From (2.14) and (2.15), one sees that $\widetilde{W}^0$ has the semidirect product structure

$$\widetilde{W}^0 = \mathscr{T} \wedge W^0, \tag{2.16}$$

where $W^0 = N^0/C^0$ is the corresponding inner symmetry group of the root diagram. It is well known that W^0 defined in this way is isomorphic to the Weyl group, i.e., the group generated by reflections in planes orthogonal to the roots. In general, it is not possible to sharpen the quotient in (2.16) to a semidirect product $N^0 = W^0 \wedge C^0$; for example, among the classical groups it is possible only for SU$(2n+1)$.

For future reference we recall that the Weyl reflections in root space are induced by the reflections in the Cartan subalgebra obtained by conjugation with the elements

$$g^\alpha = \exp[(i\pi/2)(E^\alpha + E^{-\alpha})] \tag{2.17}$$

of the inner normalizer N^0. [However, it should be noted that the finite group, called DT(G_0),[7] generated by the g^α in the defining representation where $(g^\alpha)^4 = 1$, is not isomorphic to the Weyl group [which is generated by adj(g^α)], but has it as a quotient, i.e., $W^0 = \mathrm{DT}(G_0)/(Z_2)^l$, where $(Z_2)^l$ is the subgroup of the Cartan group generated by the $(g^\alpha)^2$.] Since $(\mathrm{adj}(g^\alpha))^2 = 1$, the action of the g^α on the full Lie algebra G_0 must be involutive and is of the form

$$\beta\cdot H \rightarrow \gamma\cdot H, \quad E^\beta \rightarrow \epsilon(\beta,\gamma)E^\gamma, \tag{2.18}$$

where $\epsilon(\beta,\gamma) = \pm 1$.

It is possible to choose the subgroup $\mathscr{T}$ in (2.16) so that it transforms the l primitive roots α_i of G_0 into themselves (and is thus the symmetry group of the Dynkin diagram) and so that its action on the Lie algebra is generated by the transformations

$$\tau(\alpha_i\cdot H) = \alpha_j\cdot H, \quad \tau(E^{\alpha_i}) = E^{\alpha_j}, \quad \tau\in\mathscr{T} \tag{2.19}$$

on the Cartan subalgebra and the l primitive E^{α_i}.

It is well known that in practice the group $\mathscr{T}$ can only be one of the permutation groups $\mathscr{T} = 1, S_2, S_3$, where S_3 occurs only for the algebra of Spin (8) and S_2 occurs only for the other algebras whose simply connected groups $\mathscr{G}_0$ are listed in the top row of Table I; the maximal $\mathscr{T}$-invariant subgroups $\widehat{\mathscr{G}}_0$ of $\mathscr{G}_0$ are given in the second row. The relative positions of the $\widehat{\mathscr{G}}_0$ and $\mathscr{G}_0$ root lattices are determined by the positions of the invariant α's in G_0 relative to λ and σ in $\widehat{G}_0$, as given in the third row of Table I. Note that the $\mathscr{G}_0$ and $\widehat{\mathscr{G}}_0$ are all simply laced and nonsimply laced, respectively (and that all groups except E_7 and E_8 appear). The group SU$(2n+1)$ is exceptional in that it is the only group for which (i) $\tau(\alpha) = \alpha \neq \lambda$, (ii) dim $\widehat{G}_0 <$ dim $\widehat{G}_0^\perp$, and (iii) no circle in the Dynkin diagram is $\mathscr{T}$ invariant. All the groups $\mathscr{G}_0$ except SU$(2n+1)$ will be called generic. In all cases the action of $\mathscr{T}$ is to permute the inequivalent representations of the same dimension [e.g., the vector and the two primitive spinor representations of Spin(8), the two primitive spinor representations of Spin$(2n)$, and the complex conjugate representations of SU(n) and E_6].

The relative positions of the roots may be understood by observing that the τ-invariant projections of the α's are the weights of $\widehat{\mathscr{G}}_0$ in $\mathscr{G}_0$ and that for the generic case, because dim $\widehat{G}_0 >$ dim $\widehat{G}_0^\perp$, the longest projections $\tau(\alpha) = \alpha$ (longest weights) belong to $\widehat{\mathscr{G}}_0$. Hence, the $\tau(\alpha) = \alpha$ must coincide with the long roots λ of $\widehat{\mathscr{G}}_0$. For SU$(2n+1)$, because dim $\widehat{G}_0 <$ dim $\widehat{G}_0^\perp$, the $\tau(\alpha) = \alpha$ no longer belong to $\widehat{\mathscr{G}}_0$ and that they coincide with the 2σ may be seen directly from the fact that $\widehat{\mathscr{G}}_0$ and $\widehat{\mathscr{G}}_0^\perp$ are just the antisymmetric and traceless-symmetric second-rank tensor representations of $\widehat{\mathscr{G}}_0 = \mathrm{SO}(2n+1)$.

III. RECALL OF PROPERTIES OF KM ALGEBRAS

The generalization[2,3] of simple compact Lie algebras G_0 to (untwisted affine) KM algebras is

$$[T^a_m, T^b_n] = if^{ab}{}_c\, T^c_N + m\delta^{ab}\delta_N K,$$
$$n,m\in\mathbb{Z}, \quad N = n+m, \tag{3.1}$$

where K is a central charge (commutes with all T^a_m) and $\delta_N = 0, 1$ for $n\neq 0$, $n = 0$. It is not actually necessary for n, m to be integers (elements of any additive group such as the half-integers or the real line would do for consistency). The discrete, equal spacing of m, n reflects the assumption that the Fourier-transformed space is the circle. The Cartan form of (3.1) is

TABLE I. Groups with nontrivial outer automorphisms, their maximal $\mathscr{T}$-invariant subgroups, and the relative positions of the respective roots.

$\mathscr{G}_0$	Spin(8)	Spin$(2n)$, $n\geqslant 5$	SU$(2n)$, $n\geqslant 2$	SU$(2n+1)$	E_6
$\widehat{\mathscr{G}}_0$	G_2	Spin$(2n-1)$	Sp$(2n)$	SO$(2n+1)$	F_4
$\alpha = \tau(\alpha)$	$\alpha=\lambda$	$\alpha=\lambda$	$\alpha=\lambda$	$\alpha=2\sigma$	$\alpha=\lambda$

$$[H^i_m, H^j_n] = mK\delta^{ij}\delta_N, \quad [H^i_m, E^\alpha_n] = \alpha^i E^\alpha_N,$$
$$[E^\alpha_m, E^{-\alpha}_n] = (2/\alpha^2)(\alpha\cdot H_N + mK\delta_N), \tag{3.2}$$

with further obvious relations for the E^α_m among themselves; it is worth noting that even in the Abelian case ($E^\alpha_m = 0$) the algebra (3.2) is nontrivial and is in fact the algebra of oscillators corresponding to a free (Euclidean) quantized field, e.g., the transverse part of the bosonic string algebra. It is usual to adjoin to any KM algebra a scale operator D defined up to a central addition by

$$[D, T^a_m] = -mT^a_m. \tag{3.3}$$

Then because the H^i_m do not commute and the H^i_0 and D identify the E^α_m uniquely (up to phases), it is convenient to use

$$\{D, H^i_0, K\} \tag{3.4}$$

as the KM–Cartan subalgebra. Thus, apart from the trivial central term K, the KM–Cartan subalgebra contains only one more operator than the ordinary Cartan subalgebra.

There are some important Lie subalgebras of the KM algebra, namely the Lie algebra

$$G_0 = \{H^i_0, E^\alpha_0\} \equiv \{H^i, E^\alpha\} \tag{3.5}$$

on which the KM algebra is based and the $SU(2)^\alpha_n$ subalgebras with the generators

$$\{E^\alpha_n, E^{-\alpha}_{-n}, (2/\alpha^2)(\alpha\cdot H_0 + nK)\}. \tag{3.6}$$

[The construction for $SU(2)^\alpha_n$ can actually be extended to algebras G_n isomorphic to G_0 by defining them as the closure under commutation of the generators E^β_n, $E^{-\beta}_{-n}$, $(2/\beta^2)(\beta\cdot H_n + nK)$, where the β are the l-primitive (simple) roots, but this extension will not be needed here.]

The representations of the KM algebras (3.1) in which we shall be interested are the so-called *highest weight* (unitary) representations defined by the annihilation and self-adjoint conditions

$$T^a_m|0\rangle = 0, \quad m>0; \quad (T^a_m)^\dagger = T^a_{-m}, \quad K^\dagger = K > 0. \tag{3.7}$$

The states $|0\rangle$, which are the highest weight states of $-D$, are called the vacuum states and carry a unitary representation of the Lie group $\mathscr{G}_0$ generated by the T^a_0. (Thus the vacuum has a degeneracy equal to the dimension of this representation.) The unitarity of the representations of the $SU(2)^\alpha_n$ groups, generated by the algebras in (3.8), implies that the quantity $2(\alpha\cdot H_0 + nK)/\alpha^2$ has an integer spectrum and hence, that K (which is positive in any case) must be quantized according to

$$2K/\lambda^2 \in \mathbb{Z}_+, \quad \lambda = \text{long root}. \tag{3.8}$$

Since $\langle 0|2(\alpha\cdot H_0 + nK)/\alpha^2|0\rangle$ can be written as $|E^{-\alpha}_{-n}|0\rangle|^2 \geqslant 0$ for $n>0$ and any vacuum state, one also sees that the values of $|\alpha\cdot H_0|$ on the vacuum states are bounded by K. In particular, for the lowest value of K, namely $K = \lambda^2/2$, it follows that the permitted vacuum representations must be subsets of the $(l+1)$-primitive representations. (The permitted subsets are exhibited in Fig. 3 of Ref. 3.)

Finally, it should be noted that the group elements

$$\exp 2\pi i D, \quad \exp 2\pi i(\tilde{w}\cdot H_0) \tag{3.9}$$

of KM algebras are central and thus fixed for each irreducible representation. Hence, the highest weight irreducible KM representations fall into the same congruency classes as the Lie group representations.

IV. THE GROUP $\tilde{N}$ OF CARTAN-PRESERVING AUTOMORPHISMS OF KM ALGEBRAS

After these preliminaries we now proceed to the main purpose of the paper, which is to consider the group $\tilde{N}$ of Cartan-preserving automorphisms and the Weyl group W of KM algebras. Let us first consider the group $\tilde{N}$. Since the KM algebra is an invariant subalgebra of the combined D–KM system it is preserved by any automorphism; thus the action of $\tilde{N}$ must be of the form

$$H^i_0 \to R^i_j(H^j_0 + u^j K), \quad D \to \lambda D + v\cdot H_0 + \eta K, \quad K \to K, \tag{4.1}$$

where R^i_j, u^i, v^i, λ, and η are parameters. Since the space $\{\mathbf{H}_n\}$ is the unique eigenspace of H^i_0 and D, with the eigenvalues zero and $\{n\}$, respectively (with each n occurring once and only once), one sees from (4.1) that the set $\{\mathbf{H}_n\}$ must transform into itself under $\tilde{N}$ and hence, by the preservation of the spectrum of D, the coefficient λ must be unity. Furthermore, for $\lambda = 1$, one sees from the KM commutators that the only permitted $\tilde{N}$ transformations of the H^i_n and E^α_n are (up to phases)

$$H^i_n \to R^i_j H^j_n, \quad n \neq 0; \quad E^\alpha_m \to E^{\bar{\alpha}}_{m+(v\cdot\alpha)}, \tag{4.2}$$

where $\bar{\alpha}^i = (R^{-1})^i_j \alpha^j$ and from the KM algebra relation

$$\begin{aligned}(2/\alpha^2)(\alpha\cdot H_0 + mK) &= [E^\alpha_m, E^{-\alpha}_{-m}] \to [E^\alpha_{m+(v\cdot\alpha)}, E^{-\alpha}_{-(m+(v\cdot\alpha))}]\\ &= (2/\alpha^2)(\alpha\cdot H_0 + (m+v\cdot\alpha)K),\end{aligned} \tag{4.3}$$

one then sees that the vectors $\mathbf{u}$ and $\mathbf{v}$ in (4.1) must be identical. Thus the most general group of Cartan-preserving automorphisms is of the form

$$H^i_0 \to R^i_j(H^j_0 + v^j K), \quad D \to D + v\cdot H_0 + \eta K. \tag{4.4}$$

However, the group (4.4) is actually the direct product of a one-dimensional subgroup $\tilde{T}$ defined as

$$H^i_0 \to H^i_0, \quad D \to D + \eta K \tag{4.5}$$

and a residual subgroup $\tilde{N}$ defined as

$$H^i_0 \to R^i_j(H^j_0 + v^j K), \quad D \to D + v\cdot H_0 + (v^2/2)K; \tag{4.6}$$

since the group $\tilde{T}$ is trivial, corresponding to the arbitrariness in the origin of D in definition (3.4), we shall henceforth refer to $\tilde{N}$ as the group of Cartan-preserving automorphisms.

In the transformations (4.6) the matrices R^i_j are easily identified as the matrices of the Cartan-preserving automorphisms $\tilde{N}^0$ of the ordinary Lie algebra and the structure of the group $\tilde{N}$ is easily seen to be a semidirect product of the form

$$\tilde{N} = \tilde{N}^0 \wedge \tilde{A}, \tag{4.7}$$

where $\tilde{A}$ is the l-parameter subgroup

$$H^i_0 \to H^i_0 + v^i K, \quad D \to D + v\cdot H_0 + (v^2/2)K. \tag{4.8}$$

Thus the essential KM feature is the existence of the group $\tilde{A}$; we now discuss this group in more detail.

Because of the discreteness of the root spectrum, the parameters $\mathbf{v}$ must satisfy quantization conditions, but since these depend on the Lie group in question and are absent when the spectrum of D is continuous, let us first identify the group $\tilde{A}$ for general $\mathbf{v}$. The identification is made by observing that the transformations (4.8) preserve the form

$$2KD - H_0^2 \tag{4.9}$$

and that this form is Galilean (from its analogy with either the form $2mE - p^2$ of nonrelativistic mechanics, or the form $p_+p_- - p^2$ of relativistic mechanics with one of the lightlike vectors $p_\pm$ kept fixed). This observation identifies $\tilde{A}$ as the group of Galilean accelerations in the l-dimensional Euclidean root space.

Note that the transformations (4.8) can be implemented by the unitary operators

$$\exp iv\cdot X, \tag{4.10}$$

where $[X^i,H_0^j] = iK\delta^{ij}$, $[X^i,D] = iH_0^i$ and the algebra in (4.10) will be recognized as both the algebra of the position, momentum, and energy in nonrelativistic quantum mechanics and the algebra of the center-of-mass, total momentum, and energy of the left- or right-moving relativistic closed string in the light-cone gauge[4] (see Sec. VIII).

However, for (untwisted) KM algebras (and for the closed string) the spectrum of D is integral and hence, $\mathbf{v}$ is not free, but quantized. In fact, from (4.2) one sees that for the E_m^α to transform into themselves it is necessary and sufficient that

$$(v\cdot\alpha)\in\mathbb{Z}, \text{ or equivalently, } \mathbf{v}\in\Gamma_{\tilde{w}}. \tag{4.11}$$

Thus the quantization condition for $\mathbf{v}$ is that it be an element of the coweight lattice. Thus finally, the group $\tilde{N}$ of Cartan-preserving automorphisms is given by (4.7), where $\tilde{A}$ is given by (4.8) with $\mathbf{v}\in\Gamma_{\tilde{w}}$.

To see the effect of these automorphisms on the congruency classes of representations one considers the lowest-level representations $K = \lambda^2/2$ for which

$$\mathbf{H}_0 \to \mathbf{H}_0 + \mathbf{v}(\lambda^2/2), \quad \mathbf{v}\in\Gamma_{\tilde{w}}. \tag{4.12}$$

Since the $\mathbf{H}_0$ is weight valued, (4.12) shows that the automorphisms change the weights by the translations that lie in the lattice $\frac{1}{2}\lambda^2\Gamma_{\tilde{w}}$, and thus the effect can be seen by identifying this lattice. It can easily be deduced from any list of roots and weights that the identification is as follows:

$$(\lambda^2/2)\Gamma_{\tilde{w}} = \begin{cases} \Gamma_w, & \text{for all simply laced groups,} \\ \Gamma_\alpha, & \text{for Spin}(2n+1) \text{ and Sp}(2n), \\ \Gamma_\lambda, & \text{for } G_2 \text{ and } F_4. \end{cases} \tag{4.13}$$

From (4.13) one then sees at once that the outer automorphisms connect representations belonging to all congruency classes for the simply laced groups, connect only congruent representations for Spin$(2n+1)$ and Sp$(2n)$, and connect only representations that are congruent with respect to the long roots for G_2 and F_4. [Note that the $\tilde{A}$ automorphisms also permute the groups $SU(2)_n^\alpha$ and change G_0 into the group $\{E^\alpha_{(\alpha\cdot v)}, E^{-\alpha}_{-(\alpha\cdot v)}, \alpha\cdot(H_0 + vK)\}$.]

Let us finally consider the action of the automorphisms $\tilde{N}$ on the vectors in weight space, which is defined as the $(l+2)$-dimensional space spanned by the eigenvalues $P = (d,\mathbf{p},k)$ of the Cartan operators $\{D,\mathbf{H}_0,K\}$ and which, on account of (4.9), has a Minkowski metric with inner products of the form

$$P\cdot P' = (dk' + d'k) - p\cdot p'. \tag{4.14}$$

Since the Cartan group C acts trivially, the action of $\tilde{N}$ is implemented by the quotient group $\tilde{W} = \tilde{N}/C$ and, from the fact that $\tilde{N}$ is the maximal group of Cartan-preserving automorphisms, it follows that $\tilde{W}$ is the symmetry group of the KM root diagram.

The action of the subgroup $\tilde{W}^0 = \tilde{N}^0/C^0$ of $\tilde{N}/C$ corresponds, of course, to rotations of $\mathbf{p}$ with d, k fixed; thus what we are really interested in are the transformations of P which are induced by $\tilde{A}$ in (4.8). It is easy to see that these transformations are of the form

$$(d,\mathbf{p},k) \to (d + v\cdot p + (v^2/2)k, \mathbf{p} + \mathbf{v}k, k), \tag{4.15}$$

which can be written in the more compact notation

$$P \to P - [2(P\cdot V)/V^2]V, \tag{4.16}$$

where $V = (1,\mathbf{v},0)$. From (4.16) one sees that the transformations are the exact analogs in Minkowski space of reflections in Euclidean space; in particular, they preserve the norm $(P'^2 = P^2)$ and thus are Lorentz transformations. However, the transformations also preserve the lightlike vectors $(0,0,k)$, so that they belong to "the little group of a lightlike vector," which is known to be a Galilean subgroup of the Lorentz group.

V. INNER GROUP N AND WEYL GROUP W OF CARTAN-PRESERVING AUTOMORPHISMS

Thus far we have considered the group $\tilde{N}$ of general Cartan-preserving automorphisms of KM algebras which, modulo the Cartan-preserving automorphisms $\tilde{N}^0$ of ordinary Lie algebras, consisted of the Galilean transformations $\tilde{A}$ in (4.8) with $\mathbf{v}\in\Gamma_{\tilde{w}}$. Let us now consider the conditions for the automorphisms $\tilde{A}$ to be inner.

Since the integer spacing of D and the root spacing of $\mathbf{H}_0$ is preserved by KM algebras, it is clear that necessary conditions for the automorphisms (4.8) to be inner are that

$$\Delta\mathbf{H}_0 \equiv \mathbf{v}K\in\Gamma_\alpha, \quad \Delta D \equiv v\cdot H_0 + (v^2/2)K\in\mathbb{Z}. \tag{5.1}$$

However, since $K = n\lambda^2/2$; $n\in\mathbb{Z}_+$; and the trivial representation $\mathbf{H}_0 = 0$ of G_0 is permitted for all values of n, including $n = 1$, the second condition in (5.1) splits into the two separate conditions

$$v\cdot H_0\in\mathbb{Z}, \quad (v^2/2)K\in\mathbb{Z}; \tag{5.2}$$

since in general $\mathbf{H}_0$ can be any weight, conditions (5.2) and the first condition in (5.1) may be written as

$$\mathbf{v}K\in\Gamma_\alpha, \quad \mathbf{v}\in\Gamma_{\tilde{\alpha}}, \quad (v^2/2)K\in\mathbb{Z}. \tag{5.3}$$

We now show that the second condition in (5.3) implies the other two and thus is the only condition that is really necessary. First, we show that irrespective of the value of K, the first two conditions in (5.3) imply the third: For this, one first expands $\mathbf{v}$ and $\mathbf{v}K$ in terms of the l-primitive roots α_i. From (5.3) one obtains

$$\mathbf{v} = \sum_{i=1}^{l} m_i \tilde{\alpha}_i; \quad \mathbf{v}K = \sum_{i=1}^{l} n_i \alpha_i, \quad m_i, n_i \in \mathbb{Z}, \tag{5.4}$$

where, since the α_i's are linearly independent, one has

$$\alpha_i^2 n_i = 2Km_i. \tag{5.5}$$

However, then

$$v^2 K = (\mathbf{v}K)\cdot\mathbf{v} = \sum_i n_i m_i (\alpha_i \cdot \tilde{\alpha}_i) + \sum_{i \neq j} n_i m_j (\alpha_i \cdot \tilde{\alpha}_j)$$
$$= 2\left\{ \sum_i n_i m_i + \sum_{i>j} n_i m_j (\alpha_i \cdot \tilde{\alpha}_j) \right\}, \tag{5.6}$$

where the equality of the terms for $i<j$ and $i>j$ follows from (5.5). Since the terms within the curly brackets of (5.6) are integers, this establishes the result. Next, we show that since $K_{\min} = \lambda^2/2$, the second condition in (5.3) implies the first. Indeed, from the inclusions (2.10) we have $\Gamma_{\tilde{\alpha}} \subseteq (2/\lambda^2)\Gamma_\alpha$ and hence,

$$\mathbf{v} \in \Gamma_{\tilde{\alpha}} \Rightarrow \mathbf{v} \in (2/\lambda^2)\Gamma_\alpha = K_{\min}^{-1} \Gamma_\alpha$$
$$\Rightarrow \mathbf{v}K \in \Gamma_\alpha. \tag{5.7}$$

It is worth emphasizing, for later reference in the twisted case, that the condition $(v^2/2)K \in \mathbb{Z}$ in (5.3) is a consequence of the other two for any value of K, whereas the first condition in (5.3) is a consequence of the second only because $K_{\min} = \lambda^2/2$. In any case, for untwisted algebras the final result is that the necessary condition (5.1) for the automorphisms to be inner reduces to

$$\mathbf{v} \in \Gamma_{\tilde{\alpha}}, \tag{5.8}$$

i.e., that $\mathbf{v}$ be an element of the coroot lattice.

We shall now show that (5.8) is also a sufficient condition by constructing the operators in the KM group which implement the transformations with the parameters $\mathbf{v} = \tilde{\alpha}$: For this, we consider the elements

$$g_n^\alpha(\theta) = \exp[(i\theta/2)(E_n^\alpha + E_{-n}^{-\alpha})], \tag{5.9}$$

of the Lie subgroups $SU(2)_n^\alpha$ of (3.7). From the KM algebra one easily verifies that

$$(g_n^\alpha(\theta))^\dagger H_0^i g_n^\alpha(\theta) = H_0^i + \alpha^i[(i/2)(E_n^\alpha - E_{-n}^{-\alpha})]\sin\theta$$
$$- (\alpha^i/\alpha^2)(\alpha\cdot H_0 + nK)(1-\cos\theta), \tag{5.10}$$

$$(g_n^\alpha(\theta))^\dagger D g_n^\alpha(\theta) = D - n[(i/2)(E_n^\alpha - E_{-n}^{-\alpha})]\sin\theta$$
$$+ (n/\alpha^2)(\alpha\cdot H_0 + nK)(1-\cos\theta), \tag{5.11}$$

and hence, for $\theta = \pi$ [when the $g_n^\alpha(\theta)$ become the elements of the Weyl group of $SU(2)_n^\alpha$], one has

$$(g_n^\alpha(\pi))^\dagger H_0^i g_n^\alpha(\pi) = H_0^i - \tilde{\alpha}^i(\alpha\cdot H_0 + nK),$$
$$(g_n^\alpha(\pi))^\dagger D g_n^\alpha(\pi) = D + n\tilde{\alpha}\cdot H_0 + [(n\tilde{\alpha})^2/2]K. \tag{5.12}$$

Thus the transformations $g_n^\alpha(\pi)$ preserve the Cartan subalgebra $\{D, \mathbf{H}_0, K\}$ and in fact are just automorphisms of the form (4.7) which are mixtures of ordinary Weyl rotations induced by α and Galilean accelerations $\tilde{A}$ induced by $n\tilde{\alpha}$. Hence, if one cancels the ordinary Weyl rotations by defining

$$U_n^\alpha = g_n^\alpha(\pi)(g_0^\alpha(\pi))^{-1}, \tag{5.13}$$

one finds that

$$(U_n^\alpha)^{-1} H_0^i U_n^\alpha = H_0^i - n\tilde{\alpha}^i K,$$
$$(U_n^\alpha)^{-1} D U_n^\alpha = D - n\tilde{\alpha}\cdot H_0 + [(n\tilde{\alpha})^2/2]K. \tag{5.14}$$

For $n = 1$ these are just the generators of the Galilean transformations (4.8) with $\mathbf{v} \in \Gamma_{\tilde{\alpha}}$, as required. Thus, finally, $\mathbf{v} \in \Gamma_{\tilde{\alpha}}$ is the necessary and sufficient condition for the general Cartan-preserving automorphisms $\mathbf{v} \in \Gamma_{\tilde{\omega}}$ to be inner.

Since the quantization conditions for $\tilde{A}$ and A are $\mathbf{v} \in \Gamma_{\tilde{\omega}}$ and $\mathbf{v} \in \Gamma_{\tilde{\alpha}}$, respectively, it follows that the quotient group $\tilde{A}/A$ is isomorphic to the lattice group $\Gamma_{\tilde{\omega}}/\Gamma_{\tilde{\alpha}}$ which, from the discussion of Sec. II, is just the center $\mathscr{Z}$ of the corresponding simply connected Lie group.

Note that for $\mathbf{v} \in \Gamma_{\tilde{\alpha}}$ the "reflections" (4.15) of Sec. IV reduce exactly to the "reflections" by which previous authors[3] defined the Weyl group W of KM algebras. Since these reflections are produced by the nontrivial action of the normalizer group N on the Cartan subalgebra $\{D, \mathbf{H}_0, K\}$ for $\mathbf{v} \in \Gamma_{\tilde{\alpha}}$, it follows that just as in the case of ordinary Lie algebras, the Weyl group defined by reflections is isomorphic to the quotient N/C, where C is the centralizer of the Cartan subalgebra.

Note, also, that the inner automorphism relation

$$\exp(i\tilde{\alpha}\cdot X) = U_1^\alpha = \exp[(i\pi/2)(E_1^\alpha + E_{-1}^{-\alpha})]$$
$$\times \exp[(-i\pi/2)(E_0^\alpha + E_0^{-\alpha})] \tag{5.15}$$

could be regarded as a formal definition of the operator $\mathbf{X}$. From this point of view $\mathbf{X}$ is then defined by the KM algebra, but since it is independent of α, it acquires a universality which transcends the particular KM algebra used to define it. Indeed, as we have seen, $\mathbf{X}$ can be used to formally generate the outer automorphisms also by extending the range of the parameter $\mathbf{v}$ in $\exp(i\mathbf{v}\cdot X)$ from $\Gamma_{\tilde{\alpha}}$ to $\Gamma_{\tilde{\omega}}$. Furthermore, if the spectrum of D is taken to be continuous the range of $\mathbf{v}$ can be extended to the continuum. In particular, $\exp(i\mathbf{v}\cdot X)$ for continuous $\mathbf{v}$ generates automorphisms of the Abelian KM (string) algebra. In this sense $\mathbf{X}$ may be regarded as a remnant of the non-Abelian part of the KM algebra in the limit when the strictly non-Abelian part E_n^α is contracted to zero.

VI. EXTENSION OF $\tilde{N}$ TO THE VIRASORO ALGEBRA

It is well known that every KM algebra admits a Virasoro automorphism and is implemented by generators L_n which satisfy the Virasoro algebra

$$[L_n, L_m] = (n-m)L_{n+m} + (c/12)n(n^2-1)\delta_{n+m}, \tag{6.1}$$

where c is an arbitrary constant, and have the action

$$[L_n, T_m^a] = -mT_{m+n}^a \tag{6.2}$$

on the KM algebra. An interesting feature of the Cartan-preserving automorphisms $\tilde{N}$ (both inner and outer) is that they can be extended to include the Virasoro algebra. In fact, it is easy to verify that if v denotes the usual parameter of Galilean accelerations, the required extension of $\widetilde{N}$ to the combined KM and V system is

$$L_n \to L_n + v\cdot H_n + (v^2/2)K\delta_n \qquad (6.3)$$

and is unique. One sees from (6.3) that L_0 has the same transformation law as D in (4.8), but this is not surprising since L_0 and D coincide in the sense that $L_0 - D$ is central to the whole KM–V system.

It is worth noting that the automorphism (6.3) is not an automorphism of the Virasoro algebra L_n alone, but transforms the Virasoro algebra into a mixture of itself and the elements $\mathbf{H}_n$, K of the KM algebra. In fact, the spaces spanned by $\{L_n,\mathbf{H}_n\}$ for fixed $n\neq 0$ and by $\{L_0,\mathbf{H}_0,K\}$ form reducible, but not fully reducible, representations of the Galilean accelerations (4.8) and indeed of the full group $\widetilde{N} = \widetilde{N}^0 \wedge \widetilde{A}$. Thus with respect to $\widetilde{N}$ any combined KM–V algebra decomposes into the representations $\{L_0,\mathbf{H}_0,K\}$, $\{L_n,\mathbf{H}_n\}$ for fixed $n\neq 0$, and $\{E^{\alpha}_n\}$ for all n, α. Further, apart from the usual decomposition of the Cartan algebras according to short and long roots, which yields two $\widetilde{N}^0$ orbits, the representations each consist of a single orbit. Thus, finally, the action of $\widetilde{N}$ on the KM–V algebra is analogous to the action of $\widetilde{N}^0$ on an ordinary Lie algebra; indeed, $\widetilde{N}$ just extends the action of $\widetilde{N}^0$ on G_0 by extending the $\widetilde{N}^0$ orbits to include (all) the different values of n.

For highest weight representations of KM algebras it is well known that the V generators L_n can be realized as bilinears in the KM generators, so that the Virasoro automorphism can be thought of as an internal automorphism. The realization is through the so-called Sommerfield–Sugawara (SS) construction[3,8,9]

$$\left(K+\frac{1}{2}Q\right)L^{\rm SS}_n = \sum_{pq} T^a_p\, T^a_q\, \theta(q-p)\delta(p+q-n)$$
$$= \sum_{p,q}\left(H_p\cdot H_q + \sum_{\alpha}\frac{\alpha^2}{2}E_p^{-\alpha}E^{\alpha}_q\right)$$
$$\times\theta(q-p)\delta(p+q-n), \qquad (6.4)$$

where Q is the group-invariant constant defined in Eq. (2.2) and $\theta(s)$, the normal-ordering stepfunction, is defined as $\theta(s) = 0, \frac{1}{2}$, and 1 for $s>0$, $=0, <0$, respectively; it is instructive to verify that for the $L^{\rm SS}_n$ the transformations (6.3) are just those induced by the $\widetilde{N}$ automorphisms of the KM algebra. Actually, since the ordinary Weyl rotations are trivial for (6.3) and since for the Galilean accelerations v the change in the H term in (6.4) is

$$\Delta\left(\sum_{p,q} H_p\cdot H_q\theta\delta\right) = \Delta\left(\frac{1}{2}H_0^2\delta_n + H_0\cdot H_n(1-\delta_n)\right)$$
$$= K\left(v\cdot H_n + \frac{v^2}{2}K\delta_n\right), \qquad (6.5)$$

the only nontrivial part of the verification is to show that the change in the E term in (6.4) due to v is

$$\Delta\left(\sum\frac{\alpha^2}{2}E_p^{-\alpha}E^{\alpha}_q\theta\delta\right) = \frac{1}{2}Q\left(v\cdot H_n + \frac{v^2}{2}K\delta_n\right). \qquad (6.6)$$

To show this, one notes that the change $E^{\alpha}_q \to E^{\alpha}_{q+m}$, where $m = (\alpha\cdot v)$, induced in the KM generators E by the v transformations may be absorbed in the θ function, so that

$$\Delta\left(\sum_{\alpha,p,q}\alpha^2 E_p^{-\alpha}E^{\alpha}_q\theta(q-p)\delta(q+p-n)\right)$$
$$= -\sum_{\alpha,p,q}\alpha^2 E_p^{-\alpha}E^{\alpha}_q\chi(q-p,m)\delta(q+p-n), \qquad (6.7)$$

where $\chi(q-p,m)$ is the difference

$$\chi(q-p,m) = \theta(q-p) - \theta(q-p-2m). \qquad (6.8)$$

However, since the factor $\chi\delta$ makes the sum over p, q finite and $\chi\delta$ is odd in $q-p$, one may make, without change in (6.5), the replacement

$$E_p^{-\alpha}E^{\alpha}_q \to \tfrac{1}{2}[E_p^{-\alpha},E^{\alpha}_q] = (1/\alpha^2)(-\alpha\cdot H_n - qK\delta_n) \qquad (6.9)$$

and then (6.7) reduces to

$$\Delta\left(\sum\alpha^2E^{-\alpha}E^{\alpha}\theta\delta\right) = \sum_{\alpha}[(\alpha\cdot H_n)x + K\delta_n y], \qquad (6.10)$$

where x and y are the numerical factors

$$x = \sum_q\chi(2q-n,m) = m, \quad y = \sum_q q\chi(2q,m) = \frac{m^2}{2}. \qquad (6.11)$$

Then recalling that m is actually $(\alpha\cdot v)$ and using the completeness relation (2.4) for the α's one has

$$\Delta\left(\sum\alpha^2E^{-\alpha}E^{\alpha}\theta\delta\right)$$
$$= \sum_{\alpha}\left[(v\cdot\alpha)(\alpha\cdot H_n) + \frac{(v\cdot\alpha)(\alpha\cdot v)}{2}K\delta_n\right]$$
$$= Q\left(v\cdot H_n + \frac{v^2}{2}K\delta_n\right), \qquad (6.12)$$

as required.

A corollary of the verification for $L^{\rm SS}_n$ is that for any L_n satisfying (6.1) and (6.2), the quantities

$$KL_n - H_0\cdot H_n, \quad n\neq 0; \quad 2KL_0 - H_0^2,$$
$$QL_n - \sum\left(\frac{\alpha^2}{2}E_p^{-\alpha}E^{\alpha}_q\theta\delta\right); \quad H_p\cdot H_q, \quad p,q\neq 0 \qquad (6.13)$$

are *separately* invariant with respect to the Cartan-preserving group $\widetilde{N}$, whereas only their sum is invariant with respect to the entire KM algebra. This observation explains why the combination $2KD - H_0^2$ introduced in (4.9) is the invariant of the Cartan algebra with respect to the group $\widetilde{N}$, whereas

$$(2K+Q)D - H_0^2 - 2\sum_{n=1}^{\infty}H_{-n}\cdot H_n$$
$$-\sum_n\sum_{\alpha}\alpha^2E^{\alpha}_{-n}E^{\alpha}_n\theta(n), \qquad (6.14)$$

with a different ratio of coefficients for D and H_0^2, is the invariant with respect to the full KM algebra. The interaction with the non-Abelian part renormalizes D and $\mathbf{H}_0$ in a different manner!

VII. USE OF THE NORMALIZER N IN THE VERTEX CONSTRUCTION

For the level 1 representations of simply laced KM algebras it is possible[3] to construct the elements E_n^α from three more primitive ingredients, namely the Lie algebra G_0, the Abelian (oscillator) KM algebra generated by $\{\mathbf{H}_n, K\}$, and the generator $\mathbf{X}$ of Galilean accelerations introduced in Sec. IV. This can be seen explicitly be recalling that in this construction the E_n^α are defined as

$$E_n^\alpha = \frac{\gamma^\alpha}{2\pi i} \oint \frac{dz}{z^{n+1}} V^\alpha(z), \tag{7.1}$$

where the integration is around the unit circle; the γ^α are elements of a unitary Clifford algebra defined by

$$\gamma^\alpha \gamma^{-\alpha} = 1, \quad \gamma^\alpha \gamma^\beta = \epsilon(\alpha,\beta)\gamma_{\alpha+\beta}, \quad \alpha+\beta \neq 0, \tag{7.2}$$

where the $\epsilon(\alpha,\beta)$ are structure constants of G_0 defined by

$$[E_0^\alpha, E_0^\beta] = \epsilon(\alpha,\beta) E_0^{\alpha+\beta}; \tag{7.3}$$

and the "vertex" $V^\alpha(z)$ is a product of the form

$$V^\alpha(z) = V_0^\alpha(z) \prod_{n>1} V_{-n}^\alpha(z) V_n^\alpha(z), \tag{7.4}$$

where

$$V_0^\alpha = \exp(i\tilde{\alpha}\cdot X)\exp((i \ln z)\tilde{\alpha}\cdot H_0), \quad V_n^\alpha = \exp[i(z^n/n)\tilde{\alpha}\cdot H_n], \quad n \neq 0. \tag{7.5}$$

The commutation relations of the V^α's [and the Clifford algebra (7.1)] then guarantee that (for level 1 representations of simply laced KM algebras) the E_n^α in (7.1) satisfy the usual KM commutation relations.

This so-called vertex construction of the E_n^α is well known, but the points that we wish to make here are the following. First, it is not $\mathbf{X}$ but the exponentiated quantities $\exp(\tilde{\alpha}\cdot X)$ that are used in the construction, and these are elements of the acceleration subgroup A of the inner Cartan-preserving automorphism group N. Second, the only part of the vertex $V^\alpha(z)$ which is not constructed from the oscillator KM algebra, namely, $V_0^\alpha(z)$ in (7.5), is also an element of the group N. Thus the vertex construction for simply laced algebras is actually a construction of the E_n^α from G_0, the oscillator KM algebra, and N, the normalizer of the Cartan in $\mathrm{Int}(G_0)$. (For nonsimply laced algebras the vertices corresponding to the short roots have an extra fermionic factor,[10] but this does not change the role of N in any essential way.)

VIII. SPACE-TIME SIGNIFICANCE OF $\tilde{N}$ FOR STRING THEORY

We now show that the Cartan-preserving groups $\tilde{N}$ of KM algebras have a remarkable space-time significance within the context of string theory. For this, one recalls[4] that the transverse part of the bosonic string algebra in the light-cone gauge is just an Abelian (oscillator) KM algebra

$$[H_m^i, H_n^j] = m\delta^{ij}\delta_N K, \quad [D, H_m^i] = -mH_m^i, \quad i,j = 1,\dots,l, \tag{8.1}$$

to which one can adjoin the $\tilde{N}$ group generator X^i of (4.10) defined as

$$[X^i, H_0^j] = iK\delta^{ij}, \quad [X^i, H_n^j] = 0, \quad n \neq 0. \tag{8.2}$$

One first notices that the generators M^{ij} of orthogonal transformations of the vectors X^i, H_n^i can be constructed from these operators by the standard formula

$$M^{ij} = X^{[i}H_0^{j]} + \sum_{n\neq 0} \frac{1}{n} H_{-n}^{[i} H_n^{j]}, \quad i,j = 1,\dots,l, \tag{8.3}$$

where $[ij]$ denotes antisymmetrization. Now, in the light-cone gauge formulation of string theory, the l-dimensional weight space $\mathbb{R}_l$ is regarded as a transverse Euclidean subspace of an $(l+2)$-dimensional Minkowski space whose remaining two dimensions are spanned by two lightlike vectors (k^μ and κ^μ, say) orthogonal to $\mathbb{R}_l$ and the oscillator KM algebra (8.1) is extended to the two extra dimensions by defining

$$k\cdot H_n = \delta_n K, \quad \kappa\cdot H_n = L_n^{\mathrm{SS}}(\mathbf{H}); \quad k\cdot X = 0, \quad [\kappa\cdot X, K] = 2iK, \tag{8.4}$$

where $L_n^{\mathrm{SS}}(\mathbf{H})$ denotes the SS–Virasoro operators of the transverse string algebra (8.1). When the Minkowski space has the critical dimension $l+2 = 26$ the quantities

$$M^{\mu\nu} = X^{[\mu}P^{\nu]} + \sum_{n\neq 0} \frac{1}{n} H_{-n}^{[\mu} H_n^{\nu]}, \quad P^\mu = H_0^\mu, \quad \mu = 0,\dots,l+1 \tag{8.5}$$

close to form a Poincaré algebra and are identified as the generators of the physical Poincaré group.

Equations (8.1)–(8.5) merely summarize the light-cone gauge treatment of the bosonic string, but the point we wish to make here is the following. If one considers the little group of this physical Poincaré group that leaves the scalar $k\cdot P$ invariant one finds that it is generated by P^μ, $M^{\mu\nu}k_\nu$ and the original $SO(l)$ generators (8.3); if one writes these generators in terms of the base elements of the string algebra one finds

$$\kappa\cdot P = L_0 = D, \quad P^i = H_0^i, \quad k\cdot P = K, \quad M^{ij}, \quad M^{i\nu}k_\nu = X^i. \tag{8.6}$$

Thus the generators of the little group within the Poincaré group of $k\cdot P$ are exactly the generators of the Cartan-preserving group $\tilde{N}$. From this, one sees that, for the string, the Cartan-preserving $\tilde{N}$ is the little group of a lightlike vector in Minkowski space; this provides a physical explanation of its Galilean structure. Note that the algebra of the little group $\tilde{N}$ closes whether or not $l+2 = 26$. Note, also, that since $\tilde{N}$ can then be interpreted as a KM automorphism group or as a space-time group, it plays the role of both an internal and external symmetry group!

IX. RECALL OF PROPERTIES OF TWISTED KM ALGEBRAS

Twisted KM algebras are generated from linear (orthogonal) automorphisms of the Lie algebras G_0 as follows. Let Λ by any such automorphism of G_0 and T_ϵ^b a (possibly complex) basis of G_0 compatible with the diagonalization of Λ, i.e., let

$$T_\epsilon^b \xrightarrow{\Lambda} e^{2\pi i\epsilon} T_\epsilon^b, \tag{9.1}$$

where b is a common eigenvalue of some set of labeling operators which are invariant with respect to Λ (but are otherwise complete) and $\exp(2\pi i\epsilon)$ are the eigenvalues of the linear transformation Λ, as indicated. Then the KM algebra is twisted by changing it to

$$[T^a_{n+\epsilon(a)},T^b_{m+\epsilon(b)}] = if^{ab}{}_c\, T^c_{N+\epsilon(c)} + (n+\epsilon(a))K\delta^{ab}\delta_{N+\epsilon(c)},$$
$$N=n+m,\quad \epsilon(c)=\epsilon(a)+\epsilon(b) \tag{9.2}$$

(where for a complex basis δ^{ab} is off diagonal with appropriate normalization). It is usual to require that Λ be cyclic of finite order, $\Lambda^r=1$, $r\in\mathbb{Z}$, in which case the ϵ's are fractions of the form $n+s/r$, $s=0,1,\ldots,r-1$, but this requirement is actually not necessary for the consistency of (9.2).

There is a sharp difference between the twists generated by inner and outer automorphisms. If Λ is inner, then it can be written as $\Lambda=\exp(2\pi i\,\mathrm{adj}(\mathscr{H}))$, where $\mathscr{H}$ is an element of the Lie algebra. Then, without loss of generality, $\mathscr{H}$ can be conjugated into the Cartan subalgebra, i.e., $\mathscr{H}$ may be written as $\mathscr{H}=v\cdot H_0$ for some l-vector $\mathbf{v}$ and in the Cartan basis the action of Λ on G_0 is then evidently

$$H_0 \xrightarrow{\Lambda(v)} H_0,\quad E^\alpha_0 \xrightarrow{\Lambda(v)} e^{2\pi i(\alpha\cdot v)}E^\alpha_0. \tag{9.3}$$

According to the rule (9.2), the twisted KM algebra then takes the form

$$[H^i_n,E^\alpha_{n+(v\cdot\alpha)}]=\alpha^i E^\alpha_{n+(v\cdot\alpha)},$$
$$[E^\alpha_{n+(v\cdot\alpha)},E^{-\alpha}_{-(n+v\cdot\alpha)}]=(2/\alpha^2)(\alpha\cdot H_0+(n+\alpha\cdot v)K), \tag{9.4}$$

etc. However, one sees by inspection that such algebras can be converted into ordinary untwisted KM algebras by the redefinitions $E^\alpha_{(n+v\cdot\alpha)}=\widetilde{E}^\alpha_n$, $\mathbf{H}_0+\mathbf{v}K=\mathbf{H}_0$; hence, they are called trivially twisted or pseudotwisted.

As a matter of fact, if one returns to the Cartan-preserving transformations $\widetilde{A}$ of Sec. V, one sees that the above redefinitions are just the $\widetilde{A}$ transformations, with the range of the parameter vector $\mathbf{v}$ extended from $\mathbf{v}\in\Gamma_{tw}$ to arbitrary real $\mathbf{v}$. Thus there is a one-to-one correspondence between the KM algebras twisted by inner automorphisms and the group of Galilean accelerations $\widetilde{A}$, or, more precisely, between these twisted KM algebras and the quotient group $\widetilde{A}_c/\widetilde{A}$, where $\widetilde{A}_c$ is the covering group of $\widetilde{A}$. Thus the quotient group $\widetilde{A}_c/\widetilde{A}$ can be used to parametrize the inner-twisted KM algebras uniquely and untwist them.

In contrast, the twisted algebras (9.2) generated by (nontrivial) outer automorphisms Λ cannot be untwisted. Such twisted algebras can be understood most easily by replacing the usual complete set of labeling operators $\{\mathbf{H}_0\}$, where $\mathbf{H}_0$ is the Cartan of G_0, by the complete set $\{\widehat{\mathbf{H}}_0,\Lambda\}$, where $\widehat{\mathbf{H}}_0$ is the Cartan of the maximal Λ-invariant subalgebra $\widehat{G}_0$ of G_0. Hence, we digress to discuss the form of the Lie algebra in the $\{\widehat{\mathbf{H}}_0,\Lambda\}$ basis.

First, since Λ must be an element of one of the groups $\mathscr{T}=1$, S_2, or S_3 of outer automorphisms of $\mathscr{G}_0$, the Λ-invariant subgroups $\widehat{\mathscr{G}}_0$ coincide with the $\mathscr{T}$-invariant subgroups $\widehat{\mathscr{G}}_0$ listed in Table I. Second, it is clear that in an $\{\widehat{\mathbf{H}}_0,\Lambda\}$ basis the Lie algebra of $\mathscr{G}_0$ must be of the form

$$\widehat{G}_0\sim\{\widehat{\mathbf{H}}_0,F^\sigma_0,F^\lambda_0\},\quad \widehat{G}^\perp_\epsilon\sim\{\widehat{\mathbf{H}}^\perp_\epsilon,F^\omega_\epsilon\}_{\epsilon\neq0}, \tag{9.5}$$

where the subscripts correspond to the eigenvalues of Λ, $\widehat{G}_0$ is in the usual Cartan form (where σ,λ are the short and long roots), and $\{\widehat{G}^\perp_\epsilon,\widehat{\mathbf{H}}^\perp_\epsilon\}$ correspond to the orthogonal complements $\{\widehat{G}^\perp_0,\mathbf{H}^\perp_0\}$ of $\{\widehat{G}_0,\widehat{\mathbf{H}}_0\}$ in $\{G_0,\mathbf{H}_0\}$, respectively. Thus the ω in (9.5) are the weights of the representation $\widehat{G}^\perp_0$ of $\widehat{G}_0$ and the problem is to determine which ω's are permitted. The ω's are evidently projections of the roots α of G_0 onto the root diagram of $\widehat{G}_0$ and hence, if $\tau(\alpha)$ denotes the Λ transform of α [according to (2.19)] and $m(\alpha)$ its multiplicity (i.e., the number of roots α that project onto the same ω), then

$$\tau(\alpha)=\alpha\Rightarrow\omega=\alpha,\quad m(\alpha)=1;$$
$$\tau(\alpha)\neq\alpha\Rightarrow|\omega|<|\alpha|,\quad m(\alpha)=2,3. \tag{9.6}$$

From (9.6) it follows that the F^ω for $\omega=\alpha$ come singly and the F^ω for $|\omega|<|\alpha|$ come in pairs or triplets; from this result and the relations $\tau(\alpha)=\alpha=\lambda$ and $\tau(\alpha)=\alpha=2\sigma$ of Table I, one sees that the only possibilities for the F^ω_ϵ in (9.5) are

$$F^\omega_{\epsilon\neq0}=F^\sigma_\epsilon\quad\text{generically},$$
$$F^\omega_{\epsilon\neq0}=\{F^\sigma_\epsilon,F^\lambda_\epsilon,F^{2\sigma}_\epsilon\}\quad\text{for SU}(2n+1). \tag{9.7}$$

The normalization of the F^ω_ϵ is not fixed by (9.7), but is determined in the usual way by the requirement

$$[F^\omega_\epsilon,F^{-\omega}_{-\epsilon}]=2(\omega\cdot\widehat{H}_0/\omega^2). \tag{9.8}$$

Explicit expressions for the F^ω_ϵ in terms of the roots E^α of $\mathscr{G}_0$ are given in Table II.

After this digression we return to the twisted algebras. From (9.5) and (9.7), it is not difficult to convince oneself that in the $\{\widehat{H}_0,\Lambda\}$ basis the twisted KM algebra (9.2) must take the following form:
generically,

$$\{\widehat{\mathbf{H}}_n,F^\sigma_n,F^\alpha_n\},\quad \{\mathbf{H}^\perp_{n+\epsilon},F^\sigma_{n+\epsilon}\}_{\epsilon\neq0},\quad \alpha=\lambda,\quad n\in\mathbb{Z},$$

SU$(2n+1)$,

$$\{\widehat{\mathbf{H}}_n,F^\sigma_n,F^\lambda_n\},\quad \{\mathbf{H}^\perp_{n+1/2},F^\sigma_{n+1/2},F^\lambda_{n+1/2},F^\alpha_{n+1/2}\},$$
$$\alpha=2\sigma,\quad n\in\mathbb{Z}, \tag{9.9}$$

with the commutation relations

$$[\widehat{\mathbf{H}}_p,F^\omega_q]=\omega F^\omega_{p+q},$$
$$[F^\omega_p,F^{-\omega}_q]=(2/\omega^2)(\omega\cdot\widehat{H}_{p+q}+pK\delta_{p+q}),\ \text{etc.},$$
$$p,q\in\mathbb{Z}+\epsilon. \tag{9.10}$$

The quantization of K for the twisted algebras is determined in the usual way by requiring that the SU$(2)^\omega_p$ algebras generated by $\{F^\omega_p,F^{-\omega}_{-p},(\omega\cdot\widehat{H}_0+pK)/\omega^2\}$ are self-adjoint, i.e., are such that the spectrum of the operator

$$2(\omega\cdot\widehat{H}_0+pK)/\omega^2,\quad p\in\mathbb{Z} \tag{9.11}$$

is integer and positive (indefinite) on vacuum states for $p>0$. For the untwisted $\widehat{G}_0$ algebras $[\omega=(\sigma,\lambda),p\in\mathbb{Z}]$ these conditions yield

$$2K/\lambda^2\in\mathbb{Z}_+,\quad (\nu\cdot\hat{h})\leqslant K, \tag{9.12}$$

where $\hat{h}$ is a highest weight and the ν are the roots of $\widehat{G}_0$, as discussed in Sec. III; we wish to consider the modifications required for the twisted cases (which are characterized by fractional values of p).

TABLE II. Relationship between the generators F^{ω}_{ϵ}, which are eigenoperators of the outer automorphism Λ, and the Cartan–Weyl generators E^{α} for $\omega \neq \alpha$. Here $\eta(\epsilon) = \exp(2\pi i \epsilon)$, $\kappa^2 = \alpha^2/2\omega^2 (=1 \text{ or } 2)$, and the ϵ_α are determined by the condition $\Lambda(E^{\alpha}) = \epsilon_\alpha E^{\tau(\alpha)}$ and the values of the ϵ_{α_i} for the l-primitive roots. (The natural choice $\epsilon_{\alpha_i} = 1$ implies $\epsilon_\alpha = \pm 1$.) For $\omega = \alpha = \tau(\alpha)$, $F^{\omega}_{0} = E^{\alpha}$.

Algebra	$\Lambda^r = 1$	ϵ	ω	F^{ω}_{ϵ}
Generic	$r=2$	$\{0,\frac{1}{2}\}$	σ	$E^{\alpha} + \eta(\epsilon)\epsilon_\alpha E^{\tau(\epsilon)}$
Spin(8)	$r=3$	$\{0,\frac{1}{3},\frac{2}{3}\}$	σ	$E^{\alpha} + \eta^2(\epsilon)\epsilon_\alpha E^{\tau(\alpha)} + \eta(\epsilon)\epsilon_\alpha\epsilon_{\tau(\alpha)} E^{\tau^2(\alpha)}$
Su$(2n+1)$	$r=2$	$\{0,\frac{1}{2}\}$	σ,λ	$\kappa(E^{\alpha} + \eta(\epsilon)\epsilon_\alpha E^{\tau(\alpha)})$

For the generically twisted algebras p is fractional only for the short roots $\omega = \sigma$ (for which it is a multiple of σ^2/λ^2) and hence, the only modification of (9.12) is

$$K/\sigma^2 \in \mathbb{Z}_+, \quad (\sigma\cdot h) \leqslant (\sigma^2/\lambda^2)K. \tag{9.13}$$

The modification (9.13) does not change the quantization of K, but reduces the permitted vacuum representations. In particular, for the lowest value of K, namely $K=\sigma^2$, it can be seen by inspection of tables of highest weights that (9.13) reduces the permitted vacuum representations to the fundamental representations of the congruency classes, i.e., to the trivial representations of G_2 and F_4, the trivial and fundamental spinor representations of SO$(2n+1)$, and the trivial and defining representations of Sp$(2n)$.

For the SU$(2n+1)$ twisted algebra $\omega = (\sigma,\lambda,2\sigma)$ and p is half-integer and half-odd-integer for $\omega = (\sigma,\lambda)$ and $\omega = 2\sigma$, respectively. For $\omega = (\sigma,\lambda)$ the only modification to (9.12) is

$$K/\lambda^2 \in \mathbb{Z}_+, \quad (v\cdot \hat{h}) \leqslant \tfrac{1}{2}K, \tag{9.14}$$

which doubles the quantization of K, but imposes no further restriction on the vacuum representations. For $\omega = 2\sigma$ the conditions on the operator (9.11) become

$$\sigma\cdot \hat{H}_0/\sigma^2 + K/4\sigma^2 \in \mathbb{Z}_+ \tag{9.15}$$

and since $(\sigma\cdot H_0)/\sigma^2$ is half-odd-integer and integer for the spinor and tensor representations of SO$(2n+1)$, respectively, one finds that for the spinorial representations,

$$K/2\lambda^2 \in \mathbb{Z}_+ - \tfrac{1}{2}, \quad (4\sigma\cdot \hat{h}^{s}) \leqslant K \tag{9.16}$$

and for the tensorial representations,

$$K/2\lambda^2 \in \mathbb{Z}_+, \quad (4\sigma\cdot \hat{h}^{t}) \leqslant K. \tag{9.17}$$

Thus the even and odd values of K/λ^2 found in (9.14) are correlated to the congruency classes of representations of SO$(2n+1)$. For the spinorial class, the lowest level is $K=\lambda^2$ and the only permitted vacuum representation is the fundamental spinor, but for the tensorial class the lowest level is $K=2\lambda^2$ and it is not difficult to check that there are $n+1$ vacuum representations permitted by (9.14) and (9.17), namely, the trivial and first-n totally antisymmetric representations of SO$(2n+1)$.

X. THE CARTAN-PRESERVING GROUP $\widetilde{N}_t$ FOR TWISTED KM ALGEBRAS

With the twisted KM algebras in hand we may now consider the group $\widetilde{N}_t$ of Cartan-preserving automorphisms of twisted KM algebras. Since the "Cartan" is in this case $\{D,\Lambda,\mathbf{H}_0,K\}$, where Λ is outer, the natural definition of $\widetilde{N}_t$ is the group of automorphisms that preserve $\{D,\hat{\mathbf{H}}_0,K\}$. In other words, $\widetilde{N}_t$ is just the group of Cartan-preserving automorphisms of the ordinary KM algebra associated with $\hat{G}_0$,

$$\widetilde{N}_t = \widetilde{N}^0_t \wedge \widetilde{A}_t, \tag{10.1}$$

but with possibly different quantization conditions for the vector parameter in the Galilean subgroup $\widetilde{A}_t$. To emphasize that the vector parameter lies in the root space of the *subgroup* $\hat{G}_0$ we shall denote it by $\mathbf{u}$ rather than $\mathbf{v}$. As in the untwisted case the quantization of $\mathbf{u}$ can be determined by inspecting the change in the subscripts on the non-Abelian bases of the KM algebra induced by $\mathbf{u}$, i.e., by inspection of

$$F^{\omega}_{n+\epsilon} \to F^{\omega}_{n+\epsilon+(u\cdot\omega)}. \tag{10.2}$$

From (10.2) one sees that $\mathbf{u}$ transforms the $F^{\omega}_{n+\epsilon}$ into themselves if and only if

$$\begin{aligned} &(u\cdot\lambda), \quad r(u\cdot\sigma)\in\mathbb{Z} \text{ generically;} \\ &2(u\cdot v)\in\mathbb{Z} \text{ for SU}(2n+1) \end{aligned} \tag{10.3}$$

[where $\Lambda^r = 1$ and $v(=\sigma,\lambda)$ are the roots of $\hat{G}_0$]. However, since for every group $r=\lambda^2/\sigma^2$ the generic conditions (10.3) can be written as

$$(u\cdot\tilde{v})\in(2/\lambda^2)\mathbb{Z}, \tag{10.4}$$

one then sees that the required quantization conditions are

$$\mathbf{u}\in(2/\lambda^2)\Gamma_w \text{ generically}, \quad \mathbf{u}\in\tfrac{1}{2}\Gamma_{\tilde{w}} \text{ for SU}(2n+1), \tag{10.5}$$

where w and $\tilde{w}$ are the weights and coweights, respectively, of the subalgebra $\hat{G}_0$. Since for $\hat{G}_0 = \text{SO}(2n+1)$ the root lattice is the cubic lattice (which is self-dual), one has

$$\Gamma_{\tilde{w}} = (2/\lambda^2)\Gamma_v \text{ for SO}(2n+1), \tag{10.6}$$

which may be used to reexpress (10.5) as

$$\begin{aligned} &\mathbf{u}\in(2/\lambda^2)\Gamma_w \text{ generically,} \\ &\mathbf{u}\in(1/\lambda^2)\Gamma_v \text{ for SU}(2n+1). \end{aligned} \tag{10.7}$$

Thus the groups (10.1) with $\mathbf{u}$ quantized as in (10.7) constitute the groups of Cartan-preserving automorphisms of the twisted algebras.

A. Necessary conditions for inner automorphisms

Let us now consider the question as to whether the automorphisms are inner. Since the root spacing of the $\hat{\mathbf{H}}_0$ and the $1/r$ spacing of D are preserved by the twisted KM algebra it is clear that necessary conditions for the automorphisms to be inner are

$$\Delta\hat{\mathbf{H}}_0 = \mathbf{u}K\in\Gamma_v, \quad \Delta D = u\cdot\hat{H}_0 + (u^2/2)K\in(1/r)\mathbb{Z}, \tag{10.8}$$

where $\nu = (\sigma,\lambda)$ are the roots and Γ_ν the root lattice of the subgroup $\hat{G}_0$.

For the SU$(2n+1)$ case these conditions are automatically satisfied by the u's in (10.7) since $K/\lambda^2 \in \mathbb{Z}$; for the generators $\mathbf{u} = \sigma/2\sigma^2$ of the Γ_ν/λ^2 lattice $\Delta(2D)$ may be written as

$$\Delta(2D) = [F^{2\sigma}_{1/2}, F^{-2\sigma}_{-1/2}] \qquad (10.9)$$

and is thus integer valued. Thus for SU$(2n+1)$ the inner conditions (10.8) do not place any restrictions on the general automorphisms.

For the generic case it is convenient to write (10.8) in the form

$$\begin{aligned} \Delta\hat{\mathbf{H}}_0 &= (r\mathbf{u})(K/r) \in \Gamma_\nu, \\ \Delta(rD) &= (ru)\cdot\hat{H}_0 + [(ru)^2/2](K/r) \in \mathbb{Z} \end{aligned} \qquad (10.10)$$

since then one sees that the conditions are exactly the same as in the untwisted case except that $(D,\mathbf{u},K)$ are replaced by $(rD, r\mathbf{u}, K/r)$. Hence, by the argument of Sec. V, the quadratic term $u^2K/2$ can play no role and (10.8) reduce to

$$\mathbf{u}K \in \Gamma_\nu, \quad ru\cdot\hat{H}_0 \in \mathbb{Z}. \qquad (10.11)$$

Since for the generic case $K_{\min} = \lambda^2/2$ and $\hat{\mathbf{H}}_0$ takes its values in the weight lattice of $\hat{G}_0$, the conditions in (10.11) reduce to

$$\mathbf{u} \in (2/\lambda^2)\Gamma_\nu, \quad \mathbf{u} \in (1/r)\Gamma_{\check{\nu}} = (\sigma^2/\lambda^2)\Gamma_{\check{\nu}} \qquad (10.12)$$

in this case. From the inclusions (2.10) one then sees that, as in the untwisted case, the first condition is the dominant one and thus

$$\mathbf{u} \in (2/\lambda^2)\Gamma_\nu \text{ generically.} \qquad (10.13)$$

Thus in the generic case conditions (10.8) reduce the weight lattice of (10.7) to the root lattice.

B. Sufficient conditions for inner automorphisms

To see that the conditions on u are also sufficient one constructs the group elements which generate the required automorphisms, namely,

$$U^\omega_{1/r} = g^\omega_{1/r}(g^\omega_0)^{-1}, \quad g^\omega_{1/r} = \exp[(i\pi/2)(F^\omega_{1/r} + F^{-\omega}_{-1/r})], \qquad (10.14)$$

where for the generic case $\omega = \sigma$ and for SU$(2n+1)$ $\omega = \sigma,\lambda,2\sigma$ and (since ordinary Weyl rotations are scale invariant) $g^{2\sigma}_0 = g^\sigma_0$. By using the formal extrapolation $n \to 1/r$ and $\alpha \to \omega$ in (5.14), to avoid repeating the algebra, one sees that

$$\begin{aligned} (U^\omega_{1/r})^{-1} D U^\omega_{1/r} &= D - u\cdot\hat{H}_0 + (u^2/2)K, \\ (U^\omega_{1/r})^{-1}\hat{\mathbf{H}}_0 U^\omega_{1/r} &= \hat{\mathbf{H}}_0 - \mathbf{u}K, \end{aligned} \qquad (10.15)$$

where, since $\omega = \sigma$ and $(\sigma,\lambda,2\sigma)$,

$$\mathbf{u} = (1/r)\tilde{\omega} = \begin{cases} (2/\lambda^2)\sigma, & \text{generically,} \\ \sigma/\sigma^2, \lambda/\lambda^2, \frac{1}{2}(\sigma/\sigma^2), & \text{for SU}(2n+1), \end{cases} \qquad (10.16)$$

respectively, and the u's generate the required lattices. The quantization conditions for the inner and outer automorphisms are summarized in Table III.

TABLE III. Quantization conditions for the vector parameter of the acceleration subgroups $\tilde{A}$ and $\tilde{A}_t$ of the KM Cartan-normalizer groups. In the untwisted case $\Gamma_{\check{\alpha}}$ and $\Gamma_{\check{\omega}}$ are the coroot and coweight lattices, respectively, of the Lie algebra G_0. In the twisted cases λ denotes the long roots and Γ_ν, Γ_w the root and weight lattices, respectively, of the Λ-invariant subalgebra of G_0, where Λ is the outer automorphism of G_0 defining the twisted KM algebra. The result (1.4) follows by inspection of the table, but note that for the twisted SU$(2n+1)$ case the identity of the inner and outer automorphisms corresponds to the fact that the Λ-invariant subgroup is SO$(2n+1)$, which [in contrast to Spin$(2n+1)$] has unit center.

KM algebras	General	Inner
Untwisted	$\mathbf{v}\in\Gamma_{\check{\omega}}$	$\mathbf{v}\in\Gamma_{\check{\alpha}}$
Generically twisted	$\mathbf{u}\in(2/\lambda^2)\Gamma_w$	$\mathbf{u}\in(2/\lambda^2)\Gamma_\nu$
Twisted SU$(2n+1)$	$\mathbf{u}\in(1/\lambda^2)\Gamma_\nu$	$\mathbf{u}\in(1/\lambda^2)\Gamma_\nu$

XI. EXTENSION OF $\tilde{N}_t$ TO THE VIRASORO ALGEBRA FOR TWISTED KM ALGEBRAS

Since the existence of the Virasoro automorphisms (6.2) of untwisted KM algebras is actually independent of the range of m and the structure of G_0, the automorphisms extend automatically to the twisted case, where they take the form

$$[L_n, T^a_{m+\epsilon}] = -(m+\epsilon)T^a_{N+\epsilon}, \quad m,n\in\mathbb{Z}, \quad N = n+m \qquad (11.1)$$

provided only that the indices on the L_n's themselves remain integers and the L_n satisfy the usual Virasoro commutation relations (6.1) (with any center c). It can also be shown that the SS construction (6.4) remains valid in the twisted case, provided the p, q range over all the fractional values permitted by the twist, and the Q is that of the full algebra G_0 [and, if the canonical center $cn(n^2-1)/12$ is required, that L_0 should have a constant $\dim(G_0/\hat{G}_0)/16(2K+Q)$ added to it].[9,11] In the $\{\hat{H}_0,\Lambda\}$ basis (9.9) of the twisted KM algebras, the SS construction is

$$\left(K + \frac{1}{2}Q\right)L^{\rm ss}_n = \sum_{pq\in\mathbb{Z}+\epsilon}\left(H_p\cdot H_q + \sum_\omega \frac{\omega^2}{2}F^{-\omega}_p F^\omega_q\right) \times\theta(q-p)\delta(p+q-n), \qquad (11.2)$$

where

$$\mathbf{H}_p = \hat{\mathbf{H}}_p, \quad p\in\mathbb{Z}; \quad \mathbf{H}_p = \mathbf{H}^\perp_p, \quad p\in\mathbb{Z}+\epsilon, \quad \epsilon\neq 0; \qquad (11.3)$$

$\omega = \sigma,\lambda$ in the generic case; $\omega = \sigma,\lambda,2\sigma$ for SU$(2n+1)$; and ϵ takes the values 0, $\frac{1}{2}$ or 0, $\frac{1}{3}$, $\frac{2}{3}$.

It is not difficult to see from (11.1) that the Cartan-preserving automorphisms of the twisted algebra can be extended to include the Virasoro algebra by letting

$$L_n \to L_n + u\cdot\hat{H}_n + (u^2/2)K\delta_n, \quad n\in\mathbb{Z}, \qquad (11.4)$$

just as in the untwisted case, and that the extension is unique.

We should now like to verify that for the SS construction (11.2) for the twisted algebras, the Cartan-preserving automorphisms induce the transformations (11.4). As in the untwisted case, the verification for the H part of $L^{\rm ss}_n$ is trivial, so that the verification actually reduces to showing that the changes in the non-Abelian contributions to $L^{\rm ss}_n$ due to twisting produce no change in the *variation* of $L^{\rm ss}_n$ (with respect to Galilean accelerations $\mathbf{u}$). The change in the non-Abelian contribution to L_n^{ss} may be written as

$$\sum_{pq\in Z}\sum_{\alpha(\omega)}\frac{\alpha^2(\omega)}{2}E_p^{-\alpha(\omega)}E_q^{\alpha(\omega)}\theta\delta\to\sum_{pq\in Z+\epsilon}\frac{\omega^2}{2}F_p^{-\omega}F_q^{\omega}\theta\delta \tag{11.5}$$

for *each* weight ω, where $\alpha(\omega)$ denote the roots α which project onto ω under the outer automorphism of the algebra G_0. From (9.9) one sees that for $\alpha(\omega)=\omega$ there is only one such root and ϵ is fixed [$\epsilon=0$ for the generic case and $\epsilon=\frac{1}{2}$ for SU$(2n+1)$], while for $\alpha(\omega)\neq\omega$ there are $r=2$, 3 such roots and ϵ runs over all r values ($\epsilon=0,\frac{1}{2}$ or $\epsilon=0,\frac{1}{3},\frac{2}{3}$). Therefore, the problem is to establish the equality of the variations of (11.5), which is equivalent to establishing the equality

$$\sum_{pq}\sum_{\alpha(\omega)=\omega}\frac{\alpha^2(\omega)}{2}[E_p^{-\alpha(\omega)},E_q^{\alpha(\omega)}]\chi(q-p,m)\delta$$
$$=\sum_{pq\in Z+\epsilon}\frac{\omega^2}{2}[F_p^{-\omega},F_q^{\omega}]\chi(q-p,m)\delta \tag{11.6}$$

for the $\alpha(\omega)$ and ϵ just discussed, where $m=(u\cdot\omega)$ and χ is defined in (6.8). This can be done by direct computation, or, more easily, by formally extending the computation in (6.7) to include the appropriate values of m and K. Using the KM algebra according to (9.9) and (9.10) and eliminating the δ function, the problem then reduces to establishing that

$$\sum_{q\in Z}\sum_{\alpha(\omega)}(\alpha(\omega)\cdot H_n+qK\delta_n)\chi(2q-n,m)$$
$$=\sum_{q\in Z+\epsilon}(\omega\cdot\hat{H}_n+qK\delta_n)\chi(2q-n,m), \tag{11.7}$$

or, since the sum of $\alpha(\omega)\cdot H_n$ over the permitted values of $\alpha(\omega)$ is just $r\omega\cdot\hat{H}_n$, to establishing that

$$r\sum_{q\in Z}(\omega\cdot\hat{H}_n+qK\delta_n)\chi(2q-n,m)$$
$$=\sum_{q\in Z+\epsilon}(\omega\cdot\hat{H}_n+qK\delta_n)\chi(2q-n,m), \tag{11.8}$$

where $\omega\cdot\hat{H}_n$ and K are independent of q. Let us consider the cases $\alpha(\omega)\neq\omega$ and $\alpha(\omega)=\omega$ in turn. For $\alpha(\omega)\neq\omega$ one has $r=2$, 3 and ϵ takes all $r=2$, 3 values. Hence, what has to be established in this case are the numerical identities

$$r\sum_{q\in Z}\chi(2q-n,m)=\sum_{rq\in Z}\chi(2q-n,m),$$
$$r\sum_{q\in Z}q\chi(2q,m)=\sum_{rq\in Z}q\chi(2q,m). \tag{11.9}$$

For $\alpha(\omega)=\omega$ one has $r=1$ and either $\epsilon=0$ (generic case) or $\epsilon=\frac{1}{2}[\mathrm{SU}(2n+1)]$. In the generic case (11.8) becomes an identity and for SU$(2n+1)$ it reduces to establishing the identities

$$\sum_{q\in Z}\chi(2q-n,m)=\sum_{q\in Z+1/2}\chi(2q-n,m), \tag{11.10a}$$

$$\sum_{q\in Z}q\chi(2q,m)=\sum_{q\in Z+1/2}q\chi(2q,m). \tag{11.10b}$$

However, by adding the lhs of (11.10a) and (11.10b) to Eqs. (11.10a) and (11.10b), respectively, these equations reduce to (11.9) for $r=2$. Hence, all that need be established finally are the numerical identities (11.9). However, by using the identities

$$\sum_{rq\in Z}f(q)\chi(2q-n,m)\equiv\sum_{q\in Z}f\left(\frac{q}{r}\right)\chi\left(\frac{2q}{r}-n,m\right)$$
$$\equiv\sum_{q\in Z}f\left(\frac{q}{r}\right)\chi(2q-rn,rm), \tag{11.11}$$

where f is any regular function, the second of which follows from the scale invariance of χ, (11.9) reduce to

$$r\sum_{q\in Z}\chi(2q-n,m)=\sum_{q\in Z}\chi(2q-rn,rm),$$
$$r^2\sum_{q\in Z}q\chi(2q,m)=\sum_{q\in Z}q\chi(2q,rm), \tag{11.12}$$

respectively, and the validity of (11.12) then follows immediately from (6.11). This completes the verification of (11.4) for the twisted case.

ACKNOWLEDGMENTS

WM would like to express gratitude for a very pleasant stay at DIAS, where some of this work was done, and thank Mike M^c Gettrick for helpful discussions. NG would like to express appreciation for hospitality received at the Department of Physics, University of Notre Dame. The authors thank L. Michel, D. Williams, and J. B. Zuber for useful discussions.

This work was partially supported by National Science Foundation Grant No. INT-8814944.

[1]V. Kac, Funct. Anal. Prilozhen **1**, 82 (1967); Funct. Anal. Prilozhen **8**, 77 (1974); (Funct. Anal. Appl. **1**, 328 (1967); Funct. Anal. Appl. **8**, 68 (1974); R. Moody, Bull. Am. Soc. **73**, 217 (1967).

[2]V. Kac, *Infinite-Dimensional Lie Algebras—An Introduction* (Birkhauser, Boston, 1983); (Cambridge U.P., Cambridge, 1985), 2nd ed.

[3]P. Goddard and D. Olive, Int. J. Mod. Phys. A **1**, 303 (1986).

[4]M. Green, J. Schwarz, and E. Written, *Superstring Theory* (Cambridge U. P., Cambridge, 1987).

[5]D. Olive and N. Turok, Nucl. Phys. B **215** (**FS7**), 470 (1983).

[6]J. F. Cornwell, *Group Theory in Physics* (Academic, New York, 1984), Vol. 2.

[7]L. Michel, J. Patera, and R. T. Sharp, J. Math Phys. **29**, 777 (1988).

[8]C. Sommerfield, Phys. Rev. **176**, 2019 (1968); H. Sugawara, Phys. Rev. **170**, 1659 (1968).

[9]N. Gorman, W. M^c Glinn, L. O'Raifeartaigh, and D. Williams, Int. J. Mod. Phys. A **4**, 1235 (1989).

[10]P. Goddard, W. Nahm, D. Olive, and A. Schwimmer, Commun. Math. Phys. **112**, 385 (1987).

[11]V. Kac and D. Peterson, Adv. Math. **53**, 125 (1984).

Path-dependent techniques for electromagnetism with magnetic charges

G. M. Buendía
Departamento de Física, Universidad Simón Bolívar, Apartado Postal 89000, Caracas 1080A, Venezuela
U. Percoco
Centro de Física, Instituto Venezolano de Investigaciones Científicas, Apartado Postal 21827, Caracas 1020A, Venezuela
A. Trias
Department de Matematiques, Universitat Politecnica de Catalunya, E.T.S. Enginyeria de Telecomunicació, Barcelona 08034, Spain

(Received 30 August 1988; accepted for publication 22 February 1989)

An extension of the Green's function formalism to the path dependent formulation of electromagnetism with magnetic charges is presented. The explicit form of the electromagnetic potential in terms of a generalized current density is calculated. As an example, the radiation field created by a moving particle with both types of charge is found. A technique is developed for the computation of ordinary potentials in terms of path dependent potentials. It is shown how the problem of a moving magnetic charge can be treated without the difficulty of dealing with moving strings of singularities. As an illustration of the method, the electromagnetic potential for a magnetic monopole moving in a straight line is calculated.

I. INTRODUCTION

In recent years magnetic monopoles have received a great deal of attention in the context of grand unified theories in which the existence of magnetic charges seems to be an unavoidable consequence of the fact that the gauge group is non-Abelian.[1] But in spite of their recent popularity, the principal motivation for the existence of magnetic monopoles goes back to Dirac's original work in 1931.[2] Dirac demonstrated that the existence of isolated magnetic charges symmetrizes electromagnetism and, most importantly, provides a quantization condition for the electric charge.

However, as it is well known, Dirac's original theory presents some disturbing features. When magnetic charges are present, the electromagnetic potentials exhibit string singularities. A Lagrangian formulation with these kinds of singularities has proven extremely difficult to work with.

In 1980, Gambini and Trias[3] (GT) (inspired by the earlier works of Cabibbo–Ferrari[4] and Wu–Yang[5]) presented a formulation where path dependence was introduced as the key element in a nonsingular description of electromagnetism with magnetic monopoles. In the GT approach string singularities lose their pathological character; thus they play the same role that point singularities play in ordinary electromagnetism (without magnetic charges).

GT proposed a Lagrangian formulation manifestly self-dual and without singularity terms or kinematical constraints. They quantized this Lagrangian and stated the dynamical equations. This formulation contains a large gauge group that allows, by an adequate choice of gauge, the recovery of previous formulations.

Motivated by the fact that the GT path-dependent formalism has been extremely successful in solving the formal difficulties of Dirac's original theory (and by the increasing experimental search for classical magnetic monopoles), we consider it important to extend standard techniques for solving differential equations in the framework of path-dependent functionals. This would open up the possibility of carrying out practical calculations, and of making predictions concerning the behavior of magnetic monopoles that could be tested in the near future.

In this paper we show how the Green's function technique can be extended to the path-dependent formalism. With this powerful tool in hand, we are able to give an explicit form for the electromagnetic potential in terms of a generalized current density that includes both the electric and magnetic monopole contributions.

The techniques developed enable us to treat the difficult problem of studying a moving magnetic charge, a problem for which a satisfactory treatment does not exist. By selecting the trajectory of the particle as the natural path to define the path-dependent potential, the complications of dealing with moving string singularities that arise in more traditional approaches to the problem[6,7] are eliminated.

We give an expression for the electromagnetic field tensor created by a moving particle with electric and magnetic charges. The result shows the perfect dual behavior between both types of charges. When the magnetic charge is zero, we recover the classical result for the radiation field of a moving electric charge.

As a special case, we study the radiation of a magnetic monopole moving along a straight line.

In Sec. II, we present a brief review of the GT formalism. In this section we define the notation and the path-dependent functionals that are going to appear throughout this work.

II. PATH-DEPENDENT FORMALISM

In this section, we are going to present some of the basic aspects of the Gambini–Trias path-dependent formulation of electromagnetism with magnetic charges.[2]

In order to emphasize the duality between magnetic and electric charges, a complex notation is introduced.

 0022-2488/89/081933-04$02.50

The autodual complex electromagnetic tensor is defined by

$$\mathbb{F}_{\mu\nu}(x) = F_{\mu\nu}(x) + i\overline{F}_{\mu\nu}(x). \tag{1}$$

The $\mathbb{F}_{\mu\nu}(x)$ satisfies the Maxwell equation,

$$\mathbb{F}_{\mu\nu}{}^{,\nu} = -J_\mu(x), \tag{2}$$

where $J_\mu(x)$ is the complex current density, which includes the contribution of both types of charges.

Now, a complex path-dependent potential is introduced as

$$A_\mu(x,P(x)) = \int_{P(x)}^{x} dx'^{\nu}\, \mathbb{F}_{\nu\mu}(x') + \text{gauge term}, \tag{3}$$

where $P(x)$ is a continuous path going from spatial infinity to the point x.

To operate with these path dependent objects, a differential operator called the parallel derivative is defined. The way it acts is as follows:

$$\partial_{:\mu} G(x,P(x)) = \lim_{\Delta x\to 0} \frac{G(x+\Delta x, P{:}\,(x+\Delta x)) - G(x,P(x))}{\Delta x}, \tag{4}$$

where $G(x,P(x))$ is any path-dependent functional, Δx is some displacement in the μ direction, and P: $(x+\Delta x)$ is obtained by parallel transporting $P(x)$ to the point $x+\Delta x$.

In this formalism, the point function $\mathbb{F}_{\mu\nu}(x)$ is written in terms of the path-dependent potentials as

$$\mathbb{F}_{\mu\nu}(x) = A_{\nu:\mu}(x,P(x)) - A_{\mu:\nu}(x,P(x)) - i\int_{P(x)}^{x} dx'^{\lambda} J_{\mu\nu\lambda}(x), \tag{5}$$

where $J_{\mu\nu\lambda}(x)$ is the dual of the complex current density. Equation (5) is obtained by integrating the dual of (2) and using (3).

Since parallel derivatives are commutative, the addition of a parallel four gradient, $\Lambda_{:\mu}(x,P(x))$, to the potential, leaves invariant the relation (5); therefore the gauge invariance of $\mathbb{F}_{\mu\nu}$ is maintained.

Ordinary point-function potentials can be obtained from the path-dependent potential by assigning a fixed path to every point x.

III. GREEN'S FUNCTION IN THE PATH-DEPENDENT FORMALISM

In the following section, we write the equation for the complex path-dependent potential and show how this equation can be solved by extending the Green's function technique to the path-dependent formalism.

From (2) and (5) we find that the path-dependent potential satisfies

$$\Box{:}\, A_\mu(x,P(x)) = \mathcal{J}_\mu(x,P(x)), \tag{6}$$

where we have introduced the parallel D'Alembertian, $\Box{:} = \partial_{:\lambda}\partial^{:\lambda}$, and, $\mathcal{J}_\mu(x,P(x))$, the complex path-dependent current density defined as

$$\mathcal{J}_\mu(x,P(x)) = \int_{P(x)}^{x} dx'^{\lambda}\Big[-J_{\lambda,\mu}(x') + J_{\mu,\lambda}(x') - iJ_{\mu\nu\lambda}{}^{,\nu}(x')\Big] + \Lambda_{:\nu:\mu}{}^{:\nu}(x,P(x)), \tag{7}$$

where we wrote the explicit form of the gauge term.

Until this point, we have formulated an equation for the path-dependent potential that is completely analogous to the one that appears in ordinary electromagnetism. In fact, for point-dependent functionals, substituting $\Box{:}$ by $\Box$, and selecting the gauge term such that

$$\Box{:}\,\Lambda = \int_{P(x)}^{x} dx'^{\lambda} J_\lambda(x'), \tag{8}$$

we recover from the real part of (6) the well-known relation for the electromagnetic potential in the absence of magnetic charges,

$$\Box A_\mu(x) = j^{\text{electric}}_{\mu(x)}. \tag{9}$$

Now, in analogy with the standard techniques for solving differential equations based on the properties of Green's function, we write the solution of (6) as

$$\mathscr{A}_\mu(x,P(x)) = \mathring{\mathscr{A}}_\mu(x,P(x)) + \int\cdots\int DP'\, d^4x'\, G(x,x';P,P') \times \mathcal{J}(x',P'(x')), \tag{10}$$

where the first term on the right is the solution of the homogeneous equation. The second term contains the sum over all paths P' that end in the point x.

The path-dependent Green's function $G(x,x';P,P')$ that appears in the above equation satisfies

$$\Box{:}\, G(x,x';P,P') = \delta^4(x,x',P-P'). \tag{11}$$

In order to define the path-dependent quantities appearing in (11), and provide a practical way to evaluate the integrals in (10), we approximate the continuous path by polygons defined in R^4. We also take the appropriate measures to ensure the convergence.

We define the polygons by $N+1$ ordered points and a unit vector K that gives the asymptotic direction of the polygon at infinity. The first point, x_0, coincides with x, the arrival point of the continuous path. The remaining N points describe changes of direction.

Then, a path-dependent functional can be approximated in the polygonal space as

$$F(x,P(x)) \to F(x_0,x_1,\ldots,x_N;K). \tag{12}$$

In this representation the delta function that appears in (11) can be written as

$$\delta^4(x_0 - X'_0; P-P') \to \prod_{i=0}^{N}[(2\pi)^{4(N+1)}]^{-1} \times d^4p_i\, e^{-i(x_i - x_i')p_i} \times d^4q\, \frac{e^{-i(k-k')q}}{(2\pi)^4}. \tag{13}$$

After applying the parallel D'Alambertian operator to the Green's function defined in the polygonal space, we get an integral expression for (11) that can be solved by standard techniques of complex variable integration, giving

$$G(x,x';P,P') \to D_{\rm ret}(x_N - x'_N)\prod_{i=1}^{N}\delta^4(x_{i-1}-x'_{i-1}-x_i+x'_i), \tag{14}$$

where

$$D_{\rm ret}(x-x') = (1/4\pi|\mathbf{x}-\mathbf{x}'|)\delta(x_0-x'_0-|\mathbf{x}-\mathbf{x}'|) \tag{15}$$

is the well-known retarded Green's function.

It is interesting to notice how the Green's function operates in the polygonal space. The term $D_{\rm ret}(x-x')$ propagates the information in the usual way (between any pair of points, from one polygon to the other). The product of the delta functions automatically propagates the information to the remaining points of the arrival polygon.

Writing the path-dependent current density in the polygonal space, and substituting (14) in (10), we get the expression for the path-dependent potential,

$$\begin{aligned}\mathscr{A}_\mu(x_0,\ldots,x_N;k) &= \mathring{\mathscr{A}}_\mu(x_0,\ldots,x_N;k)\\ &+\int d^4x'_0\, D_{\rm ret}(x_0-x'_0)\sum_{i=1}^{N}(x_i^\lambda - x_{i-1}^\lambda)\\ &\times \hat{J}_{\lambda\mu}(x_i-x_0+x'_0;x_{i-1}-x_0+x'_0)\\ &+k^\lambda\int d^4x'_0\, D_{\rm ret}(x_0-x'_0)\mathring{J}_{\lambda\mu}(x_N-x_0+x'_0;k)\\ &+\int d^4x'_0\, D_{\rm ret}(x_0-x'_0)\Lambda{:}\mu{:}\nu^\nu(x,P(x)),\end{aligned}\tag{16}$$

where we define

$$\hat{J}_{\lambda\mu}(x_i;x_{i-1}) = \int_0^{-1} d\alpha\, j_{\lambda\mu}[x_i+\alpha(x_i-x_{i-1})], \tag{17}$$

$$\mathring{J}_{\lambda\mu}(x_N;k) = \int_\infty^0 d\alpha\, j_{\lambda\mu}(x_N+k\alpha). \tag{18}$$

The third term of (16) gives the contribution of the asymptotic term of the current density; the fourth term contains the explicit form of the gauge term.

Thus we have succeeded in finding an explicit expression for the complex path-dependent potential in terms of the complex current density. To illustrate the technique and to show the limiting process from the polygonal to the continuous path, we are going to calculate the above relation for a moving charged particle.

IV. PATH-DEPENDENT POTENTIAL FOR A MOVING CHARGED PARTICLE

For a particle moving along a trajectory, parametrized by s, the complex current density is given by

$$J_\mu(x) = G\int ds\, \dot{z}_\mu(s)\delta^4[x-z(s)], \tag{19}$$

where G is the complex charge, $G = e + ig$.

Substituting the above expression in (16), (17), and (18) and computing the integrals, we are left with the problem of recovering the continuous path. This is done in the following way: first, we let the distance between consecutive points go to zero, and then we let x_N go to infinity in such a way that the asymptotic direction of the polygon is parallel to the continuous trajectory. For this last step, we assume that there is a region D (as large as desired) such that outside it the fields are zero. We assume that in this region, the particle moves in a straight line. Under this assumption, we find that in the continuous limit, the asymptotic term of the current density does not contribute to the potential, i.e., the third term of (16) is zero.

Thus the complex path-dependent potential for the moving charge takes the form

$$\begin{aligned}\mathscr{A}_\mu(x,P(x)) &= \mathring{\mathscr{A}}_\mu(x,P(x)) - \frac{G}{4\pi}\int dx'^\lambda\\ &\times\Big\{[(\ddot{z}_\lambda R+\dot{z}_\lambda-\dot{z}_\lambda Q)(x-z)_\mu\\ &+(-\ddot{z}_\mu R-\dot{z}_\mu+\dot{z}_\mu Q)(x'-z)_\lambda]\frac{1}{R^3}\\ &+i\epsilon_{\mu\nu\lambda\alpha}(x'-z)^\nu[\ddot{z}^\alpha R+\dot{z}^\alpha-\dot{z}^\alpha Q]\frac{1}{R^3}\Big\}\\ &+\text{gauge terms},\end{aligned}\tag{20}$$

where, following standard notation, we have defined

$$\begin{aligned}R &= (x-z)^\sigma\dot{z}_\sigma,\\ Q &= (x-z)^\sigma\ddot{z}_\sigma.\end{aligned}\tag{21}$$

From (3) we can identify the complex electromagnetic tensor $\mathbb{F}_{\lambda\mu}(x)$; taking its real part we obtain

$$\begin{aligned}F_{\lambda\mu}(x) &= \mathrm{Re}\,\mathbb{F}_{\lambda\mu}(x)\\ &= (e/4\pi)[(\ddot{z}_\mu R+\dot{z}_\mu-\dot{z}_\mu Q)(x'-z)_\lambda\\ &+(-\ddot{z}_\lambda R-\dot{z}_\lambda+\dot{z}_\lambda Q)(x'-z)_\mu]\\ &\times(1/R^3)+(g/4\pi)\epsilon_{\mu\nu\lambda\alpha}\\ &\times(x'-z)^\nu[\ddot{z}^\alpha R+\dot{z}^\alpha-\dot{z}^\alpha Q](1/R^3).\end{aligned}\tag{22}$$

In this expression we can clearly see the contribution to the field tensor due to the electric and the magnetic part of the charge. The dual behavior between both types of charges is evident.

Notice that when $g=0$, i.e., no magnetic charge, we recover the classical relation for the electromagnetic field tensor due to a moving electric charge. [The real part of (20) is the well-known Liénard–Wiechert potential.]

V. RADIATION OF A MAGNETIC CHARGE MOVING IN 1-D

As an application of the results and the techniques developed in the preceding sections, we are going to calculate the potential and the field tensor for a magnetic charge that moves in a straight line.

Dirac's result for the potential of a static magnetic charge has not been extended to the case of a moving magnetic charge, because of the difficulties involved in working with moving strings of singularities. In the path-dependent formulation we solve this problem by selecting a fixed path, the trajectory of the particle. It is easy to see why this is the correct choice; because it is the only path that always has a singularity where the particle is, it is precisely its trajectory.

For a magnetic charge that moves along the x_1 axis, with a trajectory parametrized as

$$z(s) = (z_0(s), z_1(s), 0, 0), \tag{23}$$

the electromagnetic potential can be written as

$$A_\alpha(x) = \frac{1}{2}\int_{-\infty}^{x} dx'^1\, F_{1\alpha}(x') + \frac{1}{2}\int_{\infty}^{x} dx'^1\, F_{1\alpha}(x'), \tag{24}$$

where for convenience, we assign two trajectories to each point. One trajectory goes from $-\infty$ to the point x, and the other from ∞ to the point x.

Notice that (24) is now a standard point function, because we have frozen the path dependence by choosing a fixed path.

From (22), setting $e = 0$, we get for the particle described by (23),

$$F_{1\alpha}(x) = (g/4\pi)\epsilon_{\mu 01\alpha}(x-z)^\mu \times [(x_1 - z_1)(\ddot{z}_1\dot{z}_0 - \ddot{z}_0\dot{z}_1) + \dot{z}_0](1/R^3). \tag{25}$$

Once more, the dual behavior between the electric and magnetic charges is manifest in the result. The electric field in the x_1 direction is zero; a magnetic charge creates electric field only in the direction perpendicular to its trajectory.

As an example, we calculate the electromagnetic potential given in (24), when the magnetic charge is moving along the x_1 axis with $\ddot{z}^\sigma(s)\ddot{z}_\sigma(s) = -a^2 = \text{const}$. In this case, we get

$$\begin{aligned} A_\mu &= (A_0, \mathbf{A}) \\ &= (0, x_3\hat{j} + x_2\hat{k})\frac{g}{4\pi}\frac{1}{x_2^2 + x_3^2} \\ &\quad\times\left[1 + \frac{1 + a^2x^2}{\sqrt{4a^2(x_2^2 + x_3^2) + (1 + a^2x^2)}}\right], \end{aligned} \tag{26}$$

where $x^2 = c^2t^2 - x_1^2 - x_2^2 - x_3^2$, and $\hat{j}$, $\hat{k}$ are the unity vectors in the x_2, x_3 directions, respectively.

Thus we have found an electromagnetic potential for a moving magnetic monopole that is only singular along the trajectory of the particle, i.e., $x_2 = x_3 = 0$, free of moving strings of singularities. It is important to remark that the potential given in (26) is an ordinary point function, so all the standard techniques of electromagnetism can be applied to it.

VI. CONCLUSIONS

In this paper we have presented an extension of the Green's function method to the path-dependent formulation of electromagnetism with magnetic charges. The method we have developed offers a powerful tool for carrying out practical calculations in situations where string singularities are involved.

We have calculated the explicit form of the electromagnetic potential in terms of a generalized current density which contains contribution from both types of charge. As a special case, we calculated the electromagnetic field tensor for a moving particle, showing that the classical result is recovered when the magnetic charge is set equal to zero.

We have also developed techniques for the computation of ordinary potentials in terms of path-dependent potentials. This technique has enabled us to study the radiation field created by a magnetic monopole moving in a straight line.

[1] G. 't Hoof, Nucl. Phys. B **79**, 276 (1974); A. Polyakov, JETP Lett. **20**, 194 (1974).

[2] P. A. M. Dirac, Proc. R. Soc. London, Ser. A **133**, 60 (1931); Phys. Rev. **74**, 817 (1948).

[3] R. Gambini and A. Trias, J. Math. Phys. **21**, 1539 (1980).

[4] N. Cabibbo and E. Ferrari, Nuovo Cimento **23**, 1147 (1962).

[5] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975); **14**, 437 (1976).

[6] P. Vinciarelli, Phys. Rev. D **6**, 3419 (1972); P. Jordan, Ann. Phys. (Leipzig) **32**, 66 (1938); A. Barut, J. Phys. A **11**, 2073 (1978).

[7] R. Gambini and A. Trias, Phys. Rev. D **23**, 553 (1981).

The analytical wavefunctions of the SU(3) limit in the interacting boson model

Gui Lu Long
Department of Physics, Tsinghua University, Beijing, People's Republic of China
Hong Zhou Sun
Department of Physics, Tsinghua University, Beijing, People's Republic of China and The Institute of Theoretical Physics, Academia Sinica, Beijing, People's Republic of China

(Received 25 March 1988; accepted for publication 8 March 1989)

The Su(3) limit intrinsic states in the interacting boson model (IBM) are given by the group theoretical method. A method to calculate the physical states from the intrinsic states is given. Using this method, the physical states of the SU(3) limit ground state band are given completely. Some preliminary results are also obtained for beta and gamma band states. The wavefunctions obtained are expressed in terms of the five building blocks of the IBM wave functions belonging to Sun, Zhang, and Feng. In the SU(3) limit it is found that the building block T_d^+, which represents three d bosons coupled to zero angular momentum, appears even in the ground state band wavefunctions. The wavefunctions have very interesting physical implications.

I. INTRODUCTION

It is well known that the interacting boson model[1–3] has been successful in the description of the energy spectrum and E2 transitions for medium and heavy nuclei. There are three group chains in the model. They are the U(5), O(6), and SU(3) limits, respectively. They correspond to the vibrational, gamma-unstable, and rotational collective motions in the geometrical picture.

Wavefunctions in the interacting boson model can be constructed in a building up process[1–3] by using the coefficient of fractional parentage. This method has been applied by Scholten in his PHINT[4] computer code. However, in some cases we need to know the analytical forms of the wavefunctions, especially in limiting cases. Because the analytical solution to a problem is a major advantage of the model, analytical wavefunctions are very important. Sun *et al.*[5] have established an analytical basis for interacting boson wavefunctions, and using the analytical basis, have given the analytical wavefunctions for the U(5) and O(6) limits.[5,6] In this paper the SU(3) limit wavefunctions are discussed.

II. THE INTRINSIC STATES

In the interacting boson model, the SU(3) group generators are[7]

$$L_q = \sqrt{10}(d^+d)^1_q,$$
$$Q_u = (s^+d + d^+s)^2_u + \sqrt{(7/4)}\,(d^+d)^2_u. \tag{1}$$

We can rewrite the generators in irreducible tensor form:

$$\begin{aligned}
A &= -2\sqrt{2}\,Q_0,\\
\nu_0 &= \tfrac{1}{2}L_0,\quad \nu_{\pm 1} = \mp\sqrt{(2/3)}Q_{\mp 2},\\
V_{\pm 1/2} &= \pm\tfrac{1}{2}\{\sqrt{(2/3)}Q_{\pm 1}\\
&\qquad \pm\tfrac{1}{2}L_{\pm 1}\},\\
T_q &= (-1)^{1/2+q}(V_{-q})^+.
\end{aligned} \tag{2}$$

Using these generators, we directly obtain the following commutation relations:

$$\begin{aligned}
&[A,\nu_{\pm 1}] = 0,\quad [A,\nu_0] = 0,\\
&[\nu_0,\nu_{\pm 1}] = \pm\nu_{\pm 1},\quad [\nu_1,\nu_{-1}] = -\nu_0,\\
&[A,T_q] = 3T_q,\quad [A,V_q] = -3V_q,\\
&[\nu_0,T_q] = qT_q,\\
&[\nu_{\pm 1},T_q] = \mp\sqrt{(1/2\mp q)(1/2\pm q+1)/2}\;T_{q\pm 1},\\
&[\nu_0,V_q] = qV_q,\\
&[\nu_{\pm 1},V_q] = \mp\sqrt{(1/2\mp q)(1/2\pm q+1)/2}\;V_{q\pm 1};
\end{aligned} \tag{3}$$

and

$$\begin{aligned}
&(VT)^1_m - (TV)^1_m = \sqrt{1/2}\,\nu_m,\\
&(VT)^0_0 + (TV)^0_0 = -1/2\sqrt{2}\,A.
\end{aligned} \tag{4}$$

This tells us that T_q and V_q are two sets of irreducible tensor operators. Here, $\nu_0\,\nu_{\pm 1}$ form an angular momentum-like algebra SU(2). The irreducible representations (IR's) have been studied by many authors.[8–11] One of us (Sun) has given the general results of SU(3) IR's.[7] The basis of vectors of IR $(\lambda\mu)$ can be labeled as the common eigenvectors of A, ν^2, and ν_0 operators,

$$|(\lambda\mu)\epsilon\,\Lambda k\rangle. \tag{5}$$

They satisfy the following equations:

$$\begin{aligned}
&A\,|(\lambda\mu)\epsilon\,\Lambda k\rangle = \epsilon|(\lambda\mu)\epsilon\Lambda k\rangle,\\
&\nu^2|(\lambda\mu)\epsilon\,\Lambda k\rangle = \Lambda(\Lambda+1)|(\lambda\mu)\epsilon\Lambda k\rangle,\\
&\nu_0|(\lambda\mu)\epsilon\,\Lambda K\rangle = k\,|(\lambda\mu)\epsilon\Lambda k\rangle.
\end{aligned} \tag{6}$$

For instance, in $(\lambda\mu) = (2,0)$, the eigenvalues of ϵ and Λ are listed in Table I.

We can prove that the following operators,

 0022-2488/89/081937-05$02.50

TABLE I. The ϵ,Λ in SU(3) (2,0) IR.

ϵ	Λ
4	0
1	$\frac{1}{2}$
-2	1

$$b^{+}_{00} = \sqrt{2/3}\, d^{+}_{0} - \sqrt{1/3}\, s^{+},$$

$$b^{+}_{1/2,\pm 1/2} = d^{+}_{\pm 1}, \tag{7}$$

$$b^{+}_{10} = \sqrt{1/3}\, d^{+}_{0} + \sqrt{2/3}\, s^{+}, \quad b^{+}_{1\pm 1} = d^{+}_{\pm 2},$$

form a set of irreducible tensor operators $U^{(2,0)}(\epsilon\Lambda k)$ of the SU(3) group:

$$b^{+}_{00}|0\rangle = |(2,0)4,0,0\rangle,$$

$$b^{+}_{1/2k}|0\rangle = |(2,0)1,\tfrac{1}{2},k\rangle, \tag{8}$$

$$b^{+}_{1k}|0\rangle = |(2,0)-2,1,k\rangle,$$

that is,

$$U^{(2,0)}(\epsilon\Lambda k) = b^{+}_{\Lambda k},$$

$$\epsilon = 4 - 6\Lambda. \tag{9}$$

We can prove that the following operators,

$$a^{+}_{0} = b^{+}_{00},$$

$$g^{+}_{m} = (b^{+}_{00} b^{+}_{1})^{1}_{m} - \tfrac{1}{2}(b^{+}_{1/2} b^{+}_{1/2})^{1}_{m}, \tag{10}$$

satisfy the following commutation relations:

$$[A,a^{+}_{0}] = 4a^{+}_{0}, \quad [A,g^{+}_{m}] = 2g^{+}_{m},$$

$$[\nu_{\pm 1},a^{+}_{0}] = 0,$$

$$[\nu_{\pm 1},g^{+}_{m}] = \mp\sqrt{(1\mp m)(1\pm m+1)/2}\, g^{+}_{m\pm 1},$$

$$[T_{\pm 1/2},a^{+}_{0}] = 0, \quad [T_{\pm 1/2},g^{+}_{m}] = 0. \tag{11}$$

We can also construct a SU(3) invariant operator,

$$e^{+} = \sqrt{1/3}\, \{b^{+}_{00}(b^{+}_{1} b^{+}_{1})^{0}_{0} - ((b^{+}_{1/2} b^{+}_{1/2})^{1} b^{+}_{1})^{0}_{0}\}. \tag{12}$$

This operator commutes with all generators of the SU(3) group. Using the operators a^{+}_{0}, g^{+}_{1}, and e^{+}, we can construct all the highest weight states of the SU(3) group that contain n bosons. This means we can construct the wavefunctions associated with the following group chain:

$$\mathrm{U}(6) \supset \mathrm{SU}(3) \supset \mathrm{U}(1) \otimes \mathrm{SU}(2). \tag{13}$$

The corresponding highest weight states of SU(3) are

$$|[n](\lambda\mu)\epsilon_{\max}\Lambda_0\Lambda_0\rangle, \tag{14}$$

where

$$\epsilon_{\max} = 2\lambda + \mu, \quad \Lambda_0 = \mu/2. \tag{14'}$$

They satisfy the following equation:

$$T_q|[n](\lambda\mu)\epsilon_{\max}\Lambda_0\Lambda_0\rangle = 0. \tag{15}$$

It is easy to show that the following states satisfy the above equation:

$$a^{+\,n}_{0}|0\rangle,$$

$$\cdots$$

$$a^{+\,(n-2p)}_{0} g^{+\,p}_{1}|0\rangle, \tag{16}$$

$$\cdots$$

$$a^{+\,(n-2p-3q)}_{0} g^{+\,p}_{1} e^{+\,q}|0\rangle,$$

$$\cdots .$$

We notice that

a^{+}_{0} creates one boson with $\epsilon = 4$, $\Lambda = 0$,

g^{+}_{1} creates two bosons with $\epsilon = 2$, $\Lambda = 1$,

e^{+} creates three bosons with $\epsilon = 0$, $\Lambda = 0$.

Using the relation $\epsilon_{\max} = 2\lambda + \mu$, $\Lambda_0 = \mu/2$, we obtain the SU(3) highest weight state vectors as follows:

$$|[n](\lambda\mu)\epsilon_{\max}\Lambda_0\Lambda_0\rangle = C a^{+\,(n-2p-3q)}_{0} g^{+\,p}_{1} e^{+\,q}|0\rangle, \tag{17}$$

where C is a normalization constant and

$$\epsilon_{\max} = 4(n-2p-3q) + 2p,$$

$$\Lambda_0 = p,$$

$$\lambda = 2n - 4p - 6q, \tag{17'}$$

$$\mu = 2p.$$

Thus it comes out that a^{+}_{0}, g^{+}_{1}, and e^{+} are the building blocks of the SU(3) highest weight vectors. From (17) and (17′), we also prove the decomposition rule for U(6) $\supset$ SU(3):

$$[n] = \sum_{pq} (2n-4p-6q,2p), \tag{18}$$

i.e.,

$$[n] = (2n0) + (2n-4,2) + (2n-8,4) + \cdots$$
$$+ (2n-6) + (2n-10) + (2n-14,4)$$
$$+ \cdots + \cdots . \tag{18'}$$

The intrinsic states of the SU(3) are[12]

$$X((\lambda\mu)K) = \begin{cases} |[n](\lambda\mu)\epsilon_{\max}\Lambda_0 K/2\rangle, \\ |[n](\lambda\mu)\epsilon_{\min}\Lambda'_0 K/2\rangle. \end{cases} \tag{19}$$

In IBM, the ground state band is generated from the $(2n,0)$ IR, and beta and gamma bands are generated from the $(2n-4,2)$ IR. The intrinsic states for these bands are

$$X((2n,0)K=0)$$
$$= \sqrt{(1/n!)}\, a^{+\,n}_{0}|0\rangle, \tag{20}$$

$$X((2n-4,2)K=0)$$
$$= \sqrt{[2/(2n-1)(n-2)!]}\, a^{+\,(n-2)}_{0} g^{+}_{0}|0\rangle, \tag{21}$$

$$X((2n-4,2)K=2)$$
$$= \sqrt{[2/(2n-1)(n-2)!]}\, a^{+\,(n-2)}_{0} g^{+}_{1}|0\rangle. \tag{22}$$

III. THE ANALYTICAL WAVEFUNCTIONS OF THE GROUND STATE BAND

The intrinsic states are related to the intrinsic states by the following formulas:

$$X((\lambda\mu)KLM) = \sum_L C((\lambda\mu)K,L)\psi((\lambda\mu)K,L,M=K), \tag{23}$$

where $\psi((\lambda\mu)KLM)$ are the physics states and $C((\lambda\mu)KL)$ is the Elliott coefficient.

In order to calculate the physics states, one usually has to go to a fixed frame and carry out the overlap integral of $X((\lambda\mu)KLM)$ with the rotational functions.[12] Because of practical difficulties, the explicit expressions of the physics states or the analytical wavefunctions of the interacting boson model have not been calculated. In order to calculate the analytical wavefunctions, we used a projection method. We start to introduce the method just by using it.

For the ground state band, the intrinsic state is

$$X((2n,0)K=0) = \sqrt{(1/n!)}\, a_0^{+\,n}|0\rangle.$$

Expanding $a_0^{+\,n}$, we obtain

$$a^{+\,n} = \sum_i C_n^i \left(-\sqrt{\frac{1}{3}}s^+\right)^i \left(\sqrt{\frac{2}{3}}d_0^+\right)^{(n-i)}. \tag{24}$$

Because the s bosons do not contribute to angular momentum, we need only to calculate the projections of the d boson part

$$\left(\sqrt{\frac{2}{3}}d_0^+\right)^n = \sum_L W(n,L,M=0), \tag{25}$$

where $W(n,L,M=0)$ is the projection onto the angular momentum L. The summation runs from $L=0$ to its maximum, $L=2n$. It is easy to show that there is no projection onto odd number angular momentum.

The angular momentum raising and lowering operators in the interacting boson model are defined as follows:

$$L_+ = 2d_2^+ d_1 + 2d_{-1}^+ d_{-2} + \sqrt{6}\, d_1^+ d_0 + \sqrt{6}\, d_0^+ d_{-1}, \tag{26}$$

$$L_- = 2d_1^+ d_2 + 2d_{-2}^+ d_{-1} + \sqrt{6}\, d_0^+ d_1 + \sqrt{6}\, d_{-1}^+ d_0.$$

We notice here that

$$L_\pm = \mp\sqrt{2}\, L_{\pm 1}. \tag{26'}$$

We see that

$$\left[L_+ \left(\sqrt{\frac{2}{3}}d_0^+\right)^n\right] = \sum_L \sqrt{L(L+1)}\; W(n,L,M=1). \tag{27}$$

We define the successive commutators of the L_+, with an operator A, as

$$[L_+^{(\Delta)}A] \equiv \overbrace{[L_+[L_+\cdots[L_+}^{\Delta\text{ times}}A]\cdots]]. \tag{28}$$

The successive commutators of L_+ with both sides of Eq. (25) are

$$\left[L_+^{(\Delta)}\left(\sqrt{\frac{2}{3}}d_0^+\right)^n\right] = \sum_L \sqrt{\frac{(L+\Delta)!}{(L-\Delta)!}}\; W(n,L,M=\Delta). \tag{29}$$

When $\Delta = 2n$, only one term $W(n,2n,2n)$ is nonzero in the summation on the right-hand side (rhs) of (29); other terms such as $W(n,2n-2,2n)$, etc., vanish because the third component of L exceeds the L value ($M > L$),

$$[L_+^{(2n)}(\sqrt{2/3}\, d_0^+)^n] = \sqrt{(4n)!}\; W(n,2n,2n). \tag{30}$$

Therefore, after calculating the successive commutators on the left-hand side (lhs) of (30), one obtains maximum angular momentum projection:

$$W(n,2n,2n) = \sqrt{1/(4n)!}\;(2n)!2^n d_2^{+\,n}. \tag{31}$$

When $\Delta = 2n-2$, Eq. (29) becomes

$$\begin{aligned}[L_+^{(2n-2)}(\sqrt{2/3}\, d_0^+)^n] &= \sqrt{(4n-2)!/2}\; W(n,2n,2n-2) \\ &\quad + \sqrt{(4n-4)!}\; W(n,2n-2,2n-2). \end{aligned} \tag{32}$$

For the same reason, only two terms are nonzero; therefore after calculating the successive commutators of the lhs of (32) and $W(n,2n,2n-2)$ on the lhs of (32) by using the lowering operators L_- twice,

$$W(n,2n,2n-2) = (1/2\sqrt{2n(2n-1)}) \times [L_-[L_-W(n,2n,2n)]], \tag{33}$$

one obtains the next maximum projection,

$$W(n,2n-2,2n-2) = -\frac{(2n)!(n-1)2^n}{(2n-1)(4n-1)\sqrt{6(4n-4)!}}\, D_2^+\, d_2^{+\,(n-2)}. \tag{34}$$

Generally, the projections of the d boson part onto the angular momentum $L = 2n-2k$ are obtained to be

$$\begin{aligned} &W(n,2n-2k,2n-2k) \\ &= \sum_{i,j} \frac{(-1)^k 2^{n+2k-5i-5j}(2n-k)!(n-k)!n!(2n+2i+2j-2k)!(4n-4k+1)!\sqrt{2}^{\,i}}{\sqrt{6}^{\,(k-2i-2j)} i! j! (k-2i-3j)!(2n-2k)!(n-k+i+j)!(n-2k+2i+3j)!(4n-2k+1)!\sqrt{(4n-4k)!}} \\ &\quad \times P_d^{+\,i} T_d^{+\,j} D_2^{+\,(k-2i-3j)} d_2^{+\,(n-2k+2i+3j)}, \end{aligned} \tag{35}$$

where the P_d^+, T_d^+, D_2^+, and T_3^+ are the building blocks of the analytical wavefunction bases,[6] their explicit expressions are

$$P_d^+ = \sqrt{1/2}(2d_2^+ d_{-2}^+ - 2d_1^+ d_{-1}^+ + d_0^{+\,2})$$

$$\propto (d^+ d^+)_0^0,$$

$$\begin{aligned} T_d^+ &= 1/3(6d_2^+ d_0^+ d_{-2}^+ - (3/2)\sqrt{6}\, d_{-1}^{+\,2} d_2^+ \\ &\quad + 3d_1^+ d_0^+ d_{-1}^+ - d_0^{+\,3} - (3/2)\sqrt{6}\, d_1^{+\,2} d_{-2}^+) \\ &\propto (d^+ d^+ d^+)_0^0, \end{aligned} \tag{36}$$

$$
\begin{aligned}
D_2^+ &= d_2^+ d_0^+ - \sqrt{(3/8)}\, d_1^{+2} \\
&\propto (d^+ d^+)_2^2, \\
T_3^+ &= 2 d_2^{+2} d_{-1}^+ \sqrt{6}\, d_2^+ d_1^+ d_0^+ + d_1^{+3} \\
&\propto (d^+ d^+ d^+)_3^3.
\end{aligned}
$$

The physical wavefunctions for the ground state band are

$$
\begin{aligned}
\psi((2n,0)k &= 0, 2n-2k, 2n-2k) \\
&= c((2n0)\circ 2n-2k)\frac{1}{\sqrt{n!}}\sum_{n_s}\frac{n!}{(n-n_s)!\,n_s!} \\
&\quad \times(-\sqrt{1/3}\,s^+)^{n_s} W(n-n_s, \\
&\quad 2n-2k, 2n-2k)|0\rangle. \qquad (37)
\end{aligned}
$$

For practical use, we have listed the analytical wavefunctions for $n = 1$ through $n = 6$, in Table II.

IV. THE ANALYTICAL WAVEFUNCTIONS OF BETA AND GAMMA BANDS

Using the same method described in the previous section, we calculated the analytical wavefunctions of the gamma band for $L = 2n-2$, $2n-3$. They are

$$
\psi((2n-4,2)k=2, L=2n-2, M=2n-2) = \sqrt{\frac{2}{(n-2)!(2n-1)}}\, d_2^{+(n-2)}\left(\sqrt{(2/3)}\, D_2^+ - \sqrt{(1/3)}\, s^+ d_2^+\right)|0\rangle,
$$

$$
\psi((2n-4,2)k=2, L=2n-3, M=2n-3) = \sqrt{\frac{1}{2(n-1)(2n-1)(n-3)!}}\, d_2^{+(n-3)} T_3^+ |0\rangle.
$$

TABLE II. The analytical wavefunctions of the ground state band for $n = 1$ to $n = 6$.

	$n=1$
$L=0$	$-s^+$
$L=2$	d_2^+
	$n=2$
$L=0$	$(1/6)\sqrt{10}s^{+2} + (2/3)\sqrt{1/5}$
$L=2$	$-((1/3)\sqrt{7}s^+ d_2^+ + (2/3)\sqrt{(2/3)}D_2^+)$
$L=4$	$\sqrt{(1/2)}d_2^{+2}$
	$n=3$
$L=0$	$-((1/18)\sqrt{14}s^{+3} + (2/15)\sqrt{7}s^+ P_d^+ + (4/15)\sqrt{(1/7)}T_d^+)$
$L=2$	$\sqrt{(7/30)}s^{+2}d_2^+ + (4/\sqrt{105})s^+ D_2^+ + (2/\sqrt{105})P_d^+ d_2^+$
$L=4$	$-(\sqrt{(11/30)}s^+ d_2^{+2} + (4/\sqrt{165})d_2^+ D_2^+)$
$L=6$	$\sqrt{(1/6)}d_2^{+3}$
	$n=4$
$L=0$	$(1/12)\sqrt{(2/3)}s^{+4} + (2/5)\sqrt{(1/3)}s^{+2}P_d^+ + (8/35)\sqrt{(1/3)}s^+ T_d^+ + (4/35)\sqrt{(2/3)}P_d^{+2}$
$L=2$	$-((1/90)\sqrt{330}s^{+3}d_2^+ - (2/105)\sqrt{105}s^{+2}D_2^+ + (1/105)\sqrt{660}s^+ P_d^+ d_2^+ + (4/1155)\sqrt{330}P_d^+ D_2^+ + (4/1155)\sqrt{165}\, T_d^+ d_2^+)$
$L=4$	$\sqrt{(143/1260)}s^{+2}d_2^{+2} + \sqrt{(416/3465)}s^+ D_2^+ d_2^+ + \sqrt{(10/1001)}P_d^+ d_2^{+2} + \sqrt{(64/45\,045)}D_2^{+2}$
$L=6$	$-(\sqrt{(5/42)}s^+ d_2^{+3} + \sqrt{(4/105)}D_2^+ d_2^{+2})$
$L=8$	$\sqrt{(1/24)}d_2^{+4}$
	$n=5$
$L=0$	$-((1/540)\sqrt{110}s^{+5} + (2/135)\sqrt{55}s^{+3}P_d^+ + (4/315)\sqrt{55}s^{+2}T_d^+ + (4/315)\sqrt{110}s^+ P_d^{+2} + (8/3465)\sqrt{110}P_d^+ T_d^+)$
$L=2$	$(1/540)\sqrt{1430}s^{+4}d_2^+ (4/954)\sqrt{715}s^{+3}D_2^+ + (2/315)\sqrt{715}s^{+2}P_d^+ d_2^+ + (8/3465)\sqrt{1430}s^+ P_d^+ D_2^+$ $+ (8/3465)\sqrt{1430}s^+ T_d^+ d_2^+ + (4/3003)\sqrt{1430}P_d^{+2}d_2^+ + (16/45\,045)\sqrt{1430}T_d^+ D_2^+$
$L=4$	$-((1/378)\sqrt{3003}s^{+3}d_2^{+2} + (2/693)\sqrt{6006}s^{+2}D_2^+ d_2^+ + \sqrt{(50/3003)}s^+ P_d^+ d_2^{+2} + (8/3)\sqrt{(1/3003)}s^+ D_2^{+2}$ $+ (8/3)\sqrt{(1/3003)}P_d^+ D_2^+ d_2^+ + (2/3)\sqrt{(2/3003)}\, T_d^+ d_2^{+2})$
$L=6$	$(1/126)(\sqrt{595}s^{+2}d_2^{+3} + (4/315)\sqrt{(595/2)}s^+ D_2^+ d_2^{+2} + (8/5355)\sqrt{595}\, D_2^{+2}d_2^+ + (1/765)\sqrt{190}\, P_d^+ d_2^{+3}$
$L=8$	$-((1/36)\sqrt{38}^+ d_2^{+4} + (4/171)\sqrt{19}\, D_2^+ d_2^{+3})$
$L=10$	$\sqrt{(1/120)}d_2^{+5}$
	$n=6$
$L=0$	$(1/162\,160)(1001\sqrt{65}s^{+6} + 6006\sqrt{130}s^{+4}P_d^+ + 6864\sqrt{130}s^{+3}T_d^+ + 20\,592\sqrt{65}s^{+2}P_d^{+2} + 7488\sqrt{65}s^+ P_d^+ T_d^+$ $+ 4320\sqrt{130}P_d^{+3} + 1152\sqrt{65}T_d^{+2})$
$L=2$	$(-1/54\,054)(1001\sqrt{130}s^{+5}d_2^+ + 2860\sqrt{65}s^{+4}D_2^+ + 5720\sqrt{65}s^{+3}P_d^+ d_2^+ + 3120\sqrt{130}s^{+2}P_d^+ D_2^+$ $+ 3120\sqrt{65}s^{+2}T_d^+ d_2^+ + 3600\sqrt{130}s^+ P_d^{+2}d_2^+ + 960\sqrt{130}s^+ T_d^+ D_2^+ + 960\sqrt{65}P_d^{+2}D_2^+ + 480\sqrt{130}P_d^+ T_d^+ d_2^+)$
$L=4$	$(1/262\,548\sqrt{14})(2431\sqrt{442}s^{+4}d_2^{+2} + 7072\sqrt{221}s^{+3}D_2^+ d_2^+ + 6120\sqrt{221}s^{+2}P_d^+ d_2^{+2} + 1632\sqrt{442}s^{+2}D_2^{+2}$ $+ 6528\sqrt{442}s^+ P_d^+ D_2^+ d_2^+ + 1632\sqrt{221}s^+ T_d^+ d_2^{+2} + 1008\sqrt{442}P_d^{+2}d_2^{+2} + 384\sqrt{221}P_d^+ D_2^{+2} + 384\sqrt{442}T_d^+ D_2^+ d_2^{+2})$
$L=6$	$(-1/87\,210\sqrt{77})(1615\sqrt{1615}s^{+3}d_2^{+3} + 1938\sqrt{3230}s^{+2}D_2^+ d_2^{+2} + 798\sqrt{3230}s^+ P_d^+ d_2^{+3} + 912\sqrt{1615}s^+ D_2^{+2}d_2^+$ $504\sqrt{1615}P_d^+ D_2^+ d_2^{+2} + 32\sqrt{3230}D_2^{+3} + 84\sqrt{3230}T_d^+ d_2^{+3})$
$L=8$	$(\sqrt{1436/396})s^{+2}d_2^{+4} + (8/1881)\sqrt{(1463/2)}s^+ D_2^+ d_2^{+3} + \sqrt{(1/2926)}P_d^+ d_2^{+4} + (8/9)\sqrt{1/1463)}D_2^{+2}d_2^{+2}$
$L=10$	$-((1/330)\sqrt{(1265/2)}s^+ d_2^{+5} + ((1/759)\sqrt{1265}\, D_2^+ d_2^{+4})$
$L=12$	$\sqrt{(1/720)}d_2^{+6}$

TABLE III. The analytical wavefunction of the beta band for $n = 3$.

$L=0$	$\sqrt{(1/2)}((1/3)\sqrt{(2/3)}s^{+3} - (1/5)\sqrt{(1/3)}s^+P_d^+ + (1/5)\sqrt{(1/3)}T_d^+)$
$L=2$	$\sqrt{(7/65)}(-\sqrt{2}s^{+2}d_2^+ + (2/7)s^+D_2^+ + (6/7)P_d^+d_2^+)$
$L=4$	$\sqrt{(2/15)}(s^+d_2^{+2} - \sqrt{2}D_2^+d_2^+)$

TABLE IV. The analytical wavefunction of the gamma band for $n = 3$.

$L=2$	$\sqrt{(42/55)}(-(1/3)s^{+2}d_2^+ - (11/21)\sqrt{2}s^+D_2^+ + (2/21)\sqrt{2}P_d^+d_2^+)$
$L=3$	$(1/2)\sqrt{(1/5)}T_3^+$

We can calculate the analytical wavefunctions for the beta band also. For example, we have listed the analytical wavefunctions of the beta and gamma bands for $n = 3$ in Tables III and IV.

V. CONCLUSIONS AND DISCUSSIONS

From Tables III and IV, we can see that this projection method is a kind of "stripping" process. For example, operating the angular momentum raising operator $L_{\max}$ times on both sides of (23), one "strips" all other angular momentum components except the $L_{\max}$ component, and one obtains the $L_{\max}$ projection. Operating the angular raising operator $L_{\max} - 1$ times on both sides of (23), one "strips" all other angular momentum components except $L_{\max}$ and $L_{\max} - 1$, and obtains a combination of $L_{\max}$ and $L_{\max} - 1$ projections; after subtracting the $L_{\max}$ projection, one obtains the $L_{\max} - 1$ projection. One necessary condition that must be present for this method to work is that $L_{\max}$ should be finite in the summation on the rhs of (23). Therefore, this method can be applied to both IBM4[13] and the unitary fermion dynamical symmetry model,[14] in which there exists an angular momentum truncation.

From the wavefunction obtained here, one can see that in the SU(3) limit, the wavefunctions are relatively complicated in comparison with the U(5) and O(6) limit wavefunctions. The building block T_d^+ arises even in the ground state band. As was already discussed in Ref. 5, T_d^+ appears generally in a higher energy band in the U(5) and O(6) limit. This constructs a particular feature of the SU(3) limit wavefunction. In the U(5) limit, the number of d bosons n_d, and the d boson senority number τ, are all good quantum numbers. In the O(6) limit, n_d is no longer a good quantum number; states with different d pair number are mixed up, but the d boson senority number τ is still a good quantum number. In the SU(3) limit, both n_d and τ are not good quantum numbers. States with different d boson senority numbers are mixed up in the SU(3) limit. In this respect, the SU(3) limit wavefunctions are more collective. Of course, the analytical wavefunctions can be used to calculate the matrix element of transition operators in IBM. This idea is currently being pursued. It is interesting to reveal what this SU(3) limit wavefunction implies in the microscopy of the interacting boson model.

In summary, we have given both the intrinsic states and the analytical wavefunctions in the SU(3) limit of the interacting boson model. The explicit expressions for the analytical wavefunctions of the ground state band are given completely. Some analytical wavefunctions for beta and gamma bands are also given. These wavefunctions are convenient in practical calculations. They are quite important in the understanding of the SU(3) limit properties of the interacting boson model.

ACKNOWLEDGMENTS

The authors are grateful to Professor Q. Z. Han and Professor M. Zhang for their kind interest and many stimulating discussions in the course of this work. Acknowledgment is also made to Professor D. H. Feng for useful discussions. They also thank Professor E. G. Zhao and Professor W. W. Wang for discussions.

This work was supported by the Chinese Science Fund.

[1] A. Arima and F. Iachello, Ann. Phys. (NY) **99**, 253 (1976).
[2] A. Arima and F. Iachello, Ann. Phys. (NY) **111**, 239 (1978).
[3] A. Arima and F. Iachello, Ann. Phys. (NY) **123**, 468 (1979).
[4] O. Scholten (computer code).
[5] H. Z. Sun, M. Zhang, and D. H. Feng, Phys. Lett. B **163**, 7 (1985).
[6] H. Z. Sun, W. W. Wang, and D. H. Feng (unpublished).
[7] H. Z. Sun, Chin. J. Nucl. Phys. **3**, 119 (1981).
[8] H. Z. Sun, Phys. Ener. Fort. Phys. Nucl. **4**, 73 (1980).
[9] H. Z. Sun, Acta Phys. **21**, 56 (1965).
[10] A. McDonald and S. P. Rosen, J. Math. Phys. **14**, 1006 (1973); S. Datta, X. Majumdar, and B. K. Basu *ibid.* **14**, 1248 (1973).
[11] J. P. Draayer and Y. Akiyama, J. Math. Phys. **14**, 1904 (1973).
[12] J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128, 562 (1958).
[13] Q. Z. Han, H. Z. Sun, and G. H. Li, Phys. Rev. C **35**, 786 (1987).
[14] Q. Z. Han, H. Z. Sun, J. J. Wang, and G. L. Long, Chin. Phys. Lett. **4**, 505 (1987).

Nonsymmetric vortices for the Ginzberg–Landau equations on the bounded domain

Y. Y. Chen
Department of Mathematics, California State University, San Bernardino, California 92407-2397

(Received 1 March 1988; accepted for publication 15 March 1989)

It is shown by means of the Cronström gauge condition [Phys. Lett. B **90**, 267 (1980)] and the constrained variational principle that there exist nonsymmetric solutions to the steady-state Ginzburg–Landau equations with arbitrary parameter $\lambda > 0$ in bounded domain Ω contained in two-dimensional Euclidean space.

I. INTRODUCTION

The Ginzburg–Landau equations were formulated in 1950 by Ginzburg and Landau[1] to describe the phenomenon of superconductivity. In the two-dimensional case, the Ginzburg–Landau equations to the equilibrium states involve a complex valued function φ and a two-component real valued vector $A = (A_1,A_2)$ and can be written[2]

$$\sum_{k=1}^{2} (\partial_k - iA_k)^2\varphi = \frac{\lambda}{2} (|\varphi|^2 - 1)\varphi, \tag{1.1}$$

$$\partial_k (\partial_k A_j - \partial_j A_k) = \operatorname{Im}[\varphi(\partial_j + iA_j)\overline{\varphi}], \quad k,j = 1,2, \quad k \neq j, \tag{1.2}$$

where ∂_k denotes $\partial/\partial x_k$, $i = \sqrt{-1}$, $\varphi = \varphi_1(x_1,x_2) + i\varphi_2(x_1,x_2)$ and $A = (A_1(x_1,x_2), A_2(x_1,x_2))$ are defined on the domain Ω contained in $\mathbb{R}^2$ where Ω is a convex bounded open set with the origin and where the boundary $\partial\Omega$ of Ω is of class C^2, and λ is a positive numerical parameter characterizing the superconducting physical material involved.

The associated Ginzburg–Landau functional is[2]

$$\begin{aligned} I_\lambda(\varphi,A) = \frac{1}{2}\int_\Omega \Big\{&(\partial_1 A_2 - \partial_2 A_1)^2 + (\partial_1\varphi_1 + A_1\varphi_2)^2 \\ &+ (\partial_1\varphi_2 - A_1\varphi_1)^2 + (\partial_2\varphi_1 + A_2\varphi_2)^2 \\ &+ (\partial_2\varphi_2 - A_2\varphi_1)^2 \\ &+ \frac{\lambda}{4}(1 - |\varphi|^2)\Big\} dx_1\, dx_2. \end{aligned}$$

Using the terminology from differential forms, the functional can be rewritten as

$$I_\lambda(\varphi,A) = \frac{1}{2}\int_\Omega \Big\{|dA|^2 + |D_A\varphi|^2 + \frac{\lambda}{4}(1 - |\varphi|^2)^2\Big\} dx_1\, dx_2,$$

where

$$A = A_1(x_1,x_2)dx_1 + A_2(x_1,x_2)dx_2$$

is a real valued one-form on Ω, d is the exterior differentiation map,

$$dA = (\partial_1 A_2 - \partial_2 A_1) dx_1\, dx_2,$$

and

$$D_A\varphi = (d - iA)\varphi = (\partial_1 - iA_1)\varphi\, dx_1 + (\partial_2 - iA_2)\varphi\, dx_2$$

denotes the covariant derivative of φ with respect to A.

It can be easily demonstrated that the functional I_λ will be invariant under the gauge transformation

$$(\varphi,A) \to (\varphi e^{i\psi}, A + \nabla\psi), \tag{1.3}$$

and the two quantities $|\varphi|^2$ and $*dA = \partial_1 A_2 - \partial_2 A_1$ are also gauge invariants. These gauge invariants are critical for studying the existence of solutions of the Ginzburg–Landau equations by variational principles.

In this paper we will study the existence of infinitely many distinct solutions (φ,A) of (1.1) and (1.2) which are characterized by having a nonzero total flux[2]

$$2\pi N = \int_\Omega dA = \int_\Omega (\partial_1 A_2 - \partial_2 A_1) dx_1\, dx_2, \tag{1.4}$$

where N is an arbitrary nonzero constant, and φ satisfies the boundary condition

$$|\varphi|^2 = 1 \quad \text{a.e. on } \partial\Omega. \tag{1.5}$$

(φ,A) satisfies the following natural boundary conditions on $\partial\Omega$:

$$\partial_1 A_2 - \partial_2 A_1 = \text{const} \tag{1.6}$$

and

$$(A_1 - \varphi_1\partial_1\varphi_2 + \varphi_2\partial_1\varphi_1,\ A_2 - \varphi_1\partial_2\varphi_2 + \varphi_2\partial_2\varphi_1)\cdot\mathbf{n} = 0 \tag{1.7}$$

where $\mathbf{n}$ is the normal vector to $\partial\Omega$. Then we will show that these solutions are not radially symmetric provided the domain Ω is not a disk in the plane.

Cronström[3] in 1979 proved that if the vector $A = (A_1,A_2)$ is continuously differentiable up to the second order, then there is a gauge transformation with form (1.3) such that the new vector $\widetilde{A} = (\widetilde{A}_1,\widetilde{A}_2)$ satisfies the Cronström gauge condition $x_1\widetilde{A}_1 + x_1\widetilde{A}_2 = 0$. We will use Cronström's gauge condition and the constrained variational principle to prove the existence and smoothness of the vortex solutions.

In a previous research paper[4] written with Berger, we studied the radially symmetric vortices of the form $\varphi = R(r)e^{in\theta}$, $A = S(r)d\theta$, for the Ginzburg–Landau equations on $\mathbb{R}^2$ with arbitrary positive λ. We found that for fixed $\lambda > 0$ these equations possess a countably infinite number of distinct solutions characterized by their vortex number N, the integer defined by

$$N = \frac{1}{2\pi}\int_\Omega dA,$$

 0022-2488/89/081942-09$02.50

and we also studied the behavior of the symmetric vortices as $\lambda \to \infty$.

In related papers for the special self-dual case (with fixed $\lambda = 1$), Taubes[5] found all smooth finite energy vortices of the Ginzburg–Landau equations on $\mathbb{R}^2$ without regard to symmetric considerations. However when $\lambda \neq 1$, the self-duality used by Taubes breaks down[6] and new approaches are needed. The self-dual solutions prompt us to consider what results in case $\lambda = 1$ can be carried over to the case $\lambda \neq 1$.

In a recent paper, Bodylev[7] proved that there exist at least two gauge inequivalent solutions of the Ginzburg–Landau equations in a bounded domain in $\mathbb{R}^3$.

The present paper is organized as follows. In Sec. II we define the function spaces for φ and A and investigate the properties of the spaces. In Sec. III we prove that the infimum of the functional I_λ over the function space is attained. In Sec. III we prove that the minimizing solution (φ,A) is a generalized solution. In Sec. IV we prove that the (φ,A) is a smooth solution of the Ginzburg–Landau equations satisfying (1.1), (1.2), (1.4)–(1.7).

II. FUNCTION SPACES

We construct the function spaces of φ and A where $I_\lambda(\varphi,A)$ attains its infimum. Following Cronström we fix the gauge by assuming that A satisfies the Cronström gauge condition

$$x_1A_1(x_1,x_2) + x_2A_2(x_1,x_2) = 0. \tag{2.1}$$

If we denote A as a one-form in Cartesian coordinates as well as in polar coordinates, respectively, i.e.,

$$\begin{aligned} A &= A_1(x_1,x_2)dx_1 + A_2(x_1,x_2)dx_2 \\ &= S(r,\theta)d\theta + T(r,\theta)dr, \end{aligned}$$

then A_1, A_2, S, and T have the following relations:

$$S = -x_2A_1 + x_1A_2, \quad T = (1/r)(x_1A_1 + x_2A_2). \tag{2.2}$$

Therefore, condition (2.1) simply means $T \equiv 0$ and $A = S(r,\theta)d\theta$. It follows that

$$A_1 = -x_2S/r^2, \quad A_2 = x_1S/r^2, \tag{2.3}$$

$$\begin{aligned} &A_1^2 + A_2^2 = S^2/r^2, \\ &{*dA} = \partial_1A_2 - \partial_2A_1 = (1/r)\partial_rS. \end{aligned} \tag{2.4}$$

We note that if A satisfies both the Cronström gauge condition (2.1) and the Coulomb gauge condition $\operatorname{div} A = 0$, then $A = S(r)d\theta$, which has been studied in Ref. 4.

We now discuss necessary conditions on (φ,A) such that $I_\lambda(\varphi,A)$ is finite. From the definition of I_λ we have automatically that ${*dA} \in L_2(\Omega)$ and $1 - |\varphi|^2 \in L_2(\Omega)$ which implies $\varphi_1,\varphi_2 \in L_2(\Omega)$. In the proof of Lemma 2.1 we will see that

$${*dA} = (1/r)\partial_rS \in L_2(\Omega)$$

implies

$$(1/r)[S(r,\theta) - S(0,\theta)] \in L_2(\Omega).$$

According to (2.3), the continuity of A_1 and A_2 at the origin forces $S(0,\theta) = 0$. Therefore we may assume $(1/r)S(r,\theta) \in L_2(\Omega)$ which is equivalent to A_1 and $A_2 \in L_2(\Omega)$ [cf. (2.4)]. Furthermore, in Lemma 3.2 we will claim that we may assume $|\varphi| \leqslant 1$ a.e. on Ω. This property with $A_1,A_2 \in L_2(\Omega)$ will give $\partial_i\varphi_j \in L_2(\Omega)$ for $i,j = 1,2$ and $\partial_i|\varphi|^2 \in L_2(\Omega)$ for $i = 1,2$.

In the following we define the function spaces for φ and A.

The function space Σ_A of A is defined by

$$\begin{aligned} \Sigma_A = \{A = (A_1,A_2), A_i\colon\ &\Omega \to \mathbb{R} \in L_2(\Omega), \quad i = 1,2, \\ &\partial_1A_2 - \partial_2A_1 \in L_2(\Omega), \quad \text{where the derivatives} \\ &\text{are in the distributional sense, and} \\ &x_1A_1 + x_2A_2 = 0 \text{ in } \Omega\}, \end{aligned}$$

with the inner product

$$(A,B)_{\Sigma_A} = \int_\Omega (\partial_1A_2 - \partial_2A_1)(\partial_1B_2 - \partial_2B_1), \quad \forall A,B \in \Sigma_A,$$

which induces the norm

$$\|A\|_{\Sigma_A} = \|\partial_1A_2 - \partial_2A_1\|_{L_2(\Omega)} = \|dA\|_{L_2(\Omega)}.$$

We will verify later that $(\ ,\ )_{\Sigma_A}$ is an inner product and Σ_A is a Hilbert space.

From (2.3) and (2.4), it is easy to see that Σ_A is equivalent to the space

$$\begin{aligned} \Sigma_S = \{A = S\,d\theta, S\colon\ &\Omega \to \mathbb{R}, (1/r)S \in L_2(\Omega) \text{ and} \\ &(1/r)\partial_rS \in L_2(\Omega), \text{ where } \partial_r \\ &\text{is the derivative in} \\ &\text{the distributional sense}\} \end{aligned}$$

with the inner product

$$(S_1,S_2)_{\Sigma_S} = \int_\Omega \frac{1}{r^2}(\partial_rS_1)(\partial_rS_2), \quad \forall S_1,S_2 \in \Sigma_S,$$

which induces the norm

$$\|S\|_{\Sigma_S} = \|(1/r)\partial_rS\|_{L_2(\Omega)} = \|dA\|_{L_2(\Omega)},$$

where A_1, A_2, and S satisfy the relations in (2.3). [By equivalent we mean that Σ_A and Σ_S are isomorphic, i.e., there is a one-to-one linear mapping L of Σ_S onto Σ_A with $(LS_1,LS_2)_{\Sigma_A} = (S_1,S_2)_{\Sigma_S}$ for all $S_1,S_2 \in \Sigma_S$. For example, the mapping from Σ_S to Σ_A defined by (2.3) satisfies these requirements].

We will now establish some critical properties of Σ_S and Σ_A.

Lemma 2.1:

(i) $(\ ,\ )_{\Sigma_S}$ is an inner product;

(ii) $\|(1/r)S\|_{L_2(\Omega)} \leqslant C\|(1/r)\partial_rS\|_{L_2(\Omega)}$,

$$\text{for all } S \in \Sigma_S, \tag{2.5}$$

where c is a constant dependent only on Ω;

(iii) Σ_S is a Hilbert space.

Proof: (i) We only need to check that $S = 0$ a.e., on Ω whenever $(S,S)_{\Sigma_S} = 0$. Indeed, $(S,S)_{\Sigma_S} = 0$ implies that

$$(1/r)\partial_rS(r,\theta) = 0 \quad \text{a.e.},$$

so that for almost all fixed θ,

$$\partial_r S(r,\theta) = 0 \quad \text{a.e.},$$

thus

$$S(r,\theta) = k(\theta) \quad \text{a.e.} \quad \text{for almost all } \theta,$$

where $k(\theta)$ is a function dependent only on θ. If $k(\theta)$ is not equal to zero almost everywhere, then, because the origin is an interior point of Ω, there is a small $\epsilon > 0$ such that

$$\infty > \int_\Omega \frac{1}{r^2} S^2 r\, dr\, d\theta$$

$$\geqslant \int_0^{2\pi} k^2(\theta) d\theta \int_0^\epsilon \frac{1}{r} dr = \infty,$$

a contradiction. We have shown that $S(r,\theta) = 0$ a.e. on Ω.

(ii) From $(1/r)S \in L_2(\Omega)$ and $(1/r)\partial_r S \in L_2(\Omega)$, we have that for almost all fixed θ, $(1/\sqrt{r})S(\cdot,\theta) \in L_2(0,r(\theta))$ and

$$(1/\sqrt{r})\, \partial_r S(\cdot,\theta) \in L_2(0,r(\theta)).$$

Therefore

$$\partial_r S(\cdot,\theta) \in L_1[0,r(\theta)]$$

which implies that $S(\cdot,\theta)$ is absolutely continuous on $[0,r(\theta)]$ and

$$S(r,\theta) = S(0,\theta) + \int_0^r \partial_r S(\tau,\theta) d\tau.$$

However

$$(1/\sqrt{r})S(\cdot,\theta) \in L_2(0,r(\theta)),$$

thus $S(0,\theta) = 0$ and

$$S(r,\theta) = \int_0^r \partial_\tau S(\tau,\theta) d\tau,$$

for almost all θ. It follows that,

$$\left\| \frac{1}{r} S \right\|^2_{L_2(\Omega)}$$

$$= \int_\Omega \left(\frac{1}{r} S(r,\theta) \right)^2 = \int_\Omega \left[\frac{1}{r} \int_0^r \partial_r S(\tau,\theta) d\tau \right]^2$$

$$\leqslant \int_\Omega \frac{1}{r^2} \left[\int_0^r \tau\, d\tau \right] \left[\int_0^{r(\theta)} \frac{(\partial_\tau S)^2}{\tau} d\tau \right]$$

$$\leqslant C(\Omega) \|S\|^2_{\Sigma_S}.$$

(iii) Let $\{S_n\}$ be a Cauchy sequence in Σ_S. By the definition of Σ_S, $\{(1/r)\partial_r S_n\}$ is a Cauchy sequence in $L_2(\Omega)$. Thus there is a function g defined on Ω, such that $(1/r)g \in L_2(\Omega)$ and $(1/r)\partial_r S_n \to (1/r)g$ in $L_2(\Omega)$. On the other hand, by (ii), $\{(1/r)S_n\}$ is a Cauchy sequence in $L_2(\Omega)$. Therefore, there is $f \in L_2(\Omega)$ to which $(1/r)S_n$ converges in $L_2(\Omega)$. We define $S = rf$. For any $\psi \in C_0^\infty(\Omega)$,

$$\int_\Omega S \partial_r (\psi r) dr\, d\theta = \int_\Omega \frac{1}{r} S \partial_r (\psi r) r\, dr\, d\theta$$

$$= \lim_{n\to\infty} \int_\Omega \frac{1}{r} S_n \partial_r (\psi r) r\, dr\, d\theta$$

$$= -\lim_{n\to\infty} \int_\Omega \partial_r S_n (\psi r) dr\, d\theta$$

$$= -\lim_{n\to\infty} \int_\Omega \left[\frac{1}{r} \partial_r S_n \right] (r\psi) r\, dr\, d\theta$$

$$= -\int_\Omega g \psi r\, dr\, d\theta,$$

this shows $\partial_r S = g$ in the distributional sense. We have proven that there exists a function $S \in \Sigma_S$ to which S_n converges in Σ_S. ■

Because of the isomorphism of Σ_A and Σ_S, we obtain the analogous properties for Σ_A immediately.

Proposition 2.2:

(i) $(\ ,\ )_{\Sigma_A}$ is an inner product;

(ii) $\|A_1\|^2_{L_2(\Omega)} + \|A_2\|^2_{L_2(\Omega)} \leqslant C \|dA\|^2_{L_2(\Omega)}$

$$\text{for all } A \in \Sigma_A, \tag{2.6}$$

where c is a constant which only depends on Ω;

(iii) Σ_A is a Hilbert space.

For function φ, according to the boundary condition $|\varphi|_{\partial\Omega} = 1$ and the necessary conditions for finiteness of $I_\lambda(\varphi,A)$ discussed before, we define

$$\Sigma_\varphi = \{\varphi = \varphi_1 + i\varphi_2, \varphi_i\colon\ \Omega \to \mathbb{R} \in W_{1,2}(\Omega), \quad i = 1,2,$$
$$\text{and } 1 - |\varphi|^2 \in \mathring{W}_{1,2}(\Omega)\},$$

where $W_{1,2}(\Omega)$ is the Sobolev space which may be defined as the completion of $C^\infty(\Omega)$ (the space of infinitely differentiable functions) in the norm

$$\|w\|_{W_{1,2}(\Omega)} = \left[\int_\Omega (|w|^2 + |\nabla w|^2) \right]^{1/2},$$

and where $\mathring{W}_{1,2}(\Omega)$ is the Sobolev space that may be defined as the completion of $C_0^\infty(\Omega)$ (the space of infinitely differentiable functions with compact support in Ω) in the norm

$$\|w\|_{\mathring{W}_{1,2}(\Omega)} = \left[\int_\Omega |\nabla w|^2 \right]^{1/2}.$$

In the following lemma we investigate the topological property of the function space Σ_φ.

Lemma 2.3: Suppose that $\{\varphi^k\}$ is a sequence in Σ_φ such that $\{\varphi_j^k\}$, $j = 1,2$, are bounded sets in $W_{1,2}(\Omega)$ and $\{1 - |\varphi^k|^2\}$ is a bounded set in $\mathring{W}_{1,2}(\Omega)$. Then $\{\varphi^k\}$ has a subsequence that we again label $\{\varphi^k\}$ such that

$$\varphi_j^k \to \varphi_j \ \text{ in } L_p(\Omega), \quad 1 \leqslant p < \infty,$$

$$\nabla\varphi_j^k \to \nabla\varphi_j \ \text{ weakly in } L_p(\Omega), \quad j = 1,2,$$

$$\varphi_j^k \to \varphi_j \ \text{ a.e. on } \Omega,$$

$$1 - |\varphi^k|^2 \to 1 - |\varphi|^2 \ \text{ weakly in } \mathring{W}_{1,2}(\Omega),$$

where $\varphi = \varphi_1 + i\varphi_2$, $\varphi_j \in W_{1,2}(\Omega)$, and $1 - |\varphi|^2 \in \mathring{W}_{1,2}(\Omega)$, i.e., $\varphi \in \Sigma_\varphi$.

Proof: Since $\{\varphi_j^k\}$, $j = 1,2$, are bounded in the Hilbert space $W_{1,2}(\Omega)$, by the theory of the Sobolev space[8] we may assume that

$$\varphi_j^k \to \varphi_j \ \text{ weakly in } W_{1,2}(\Omega),$$

$$\varphi_j^k \to \varphi_j \ \text{ in } L_p(\Omega), \quad 1 \leqslant p < \infty,$$

$$\varphi_j^k \to \varphi_j \ \text{ a.e. on } \Omega.$$

On the other hand, since $\{1 - |\varphi^k|^2\}$ is a bounded set in $\mathring{W}_{1,2}(\Omega)$, we may assume that

$$1-|\varphi^k|^2 \to f \text{ weakly in } \mathring{W}_{1,2}(\Omega),$$

where $f \in \mathring{W}_{1,2}(\Omega)$, and

$$1-|\varphi^k|^2 \to f \text{ in } L_p(\Omega), \quad 1 \leqslant p < \infty.$$

Since $\varphi_j^k \to \varphi_j$ in $L_2(\Omega)$ implies $1-|\varphi^k|^2 \to 1-|\varphi|^2$ in $L_2(\Omega)$, by the uniqueness of the limit in $L_2(\Omega)$, we have $f = 1-|\varphi|^2$. ■

We notice that if $\varphi \equiv 1$ and $A \equiv 0$, then $I_\lambda(\varphi,A)=0$. Since $I_\lambda(\varphi,A) \geqslant 0$ for all (φ,A), then $(1,0)$ is a trivial solution minimizing $I_\lambda(\varphi,A)$. In order to avoid the trivial solution we utilize the total flux $2\pi N = \int_\Omega dA$ to characterize solutions. The solution space of (φ,A) is defined by

$$\Sigma_N = \left\{ (\varphi,A), \varphi \in \Sigma_\varphi, A \in \Sigma_A, \text{ and } A \text{ satisfies the constraint } (1/2\pi) \int_\Omega dA = N \right\}, \tag{2.7}$$

where N is an arbitrary nonzero constant.

Remark: We construct a pair (φ,A) such that $(\varphi,A) \in \Sigma_N$ and $I_\lambda(\varphi,A) < \infty$. Since the domain Ω contains the origin, we can choose $0 < r_1 < r_2$ such that the set $\{x \in \mathbb{R}^2, \ |x| < r_2\}$ is contained in Ω. Let $R(|x|)$ be a C^∞ function on $\mathbb{R}^2$ such that $R(|x|)=0$ for $|x|<r_1$ and $R(|x|)=1$ for $|x|>r_2$, and let $S(|x|) = NR(|x|)$. Then, the pair (φ,A) defined by

$$\varphi = R(|x|)e^{i\theta}|_\Omega, \quad A = S\, d\theta\,|_\Omega$$

is an element in Σ_N and renders $I_\lambda(\varphi,A)$ finite.

III. THE EXISTENCE OF THE MINIMIZING SOLUTIONS

The main result in this section is the following theorem.

Theorem 3.1: The infimum of $I_\lambda(\varphi,A)$ over Σ_N [defined by (2.7)] is attained. Moreover, the infimum is positive.

First, we will give some properties of the functional. The following lemma plays an important role in solving the minimizing problem.

Lemma 3.2: For every $\varphi \in \Sigma_\varphi$, we can define a modified function $\tilde{\varphi}$ of φ, such that $|\tilde{\varphi}| \leqslant 1$ on Ω, $\tilde{\varphi} \in \Sigma_\varphi$ and $I_\lambda(\tilde{\varphi},A) \leqslant I_\lambda(\varphi,A)$, where A is an arbitrary element in Σ_A.

Proof: For a given $\varphi \in \Sigma_\varphi$, we define

$$\tilde{\varphi} = \begin{cases} \varphi, & \text{if } |\varphi| \leqslant 1, \\ \varphi/|\varphi|, & \text{if } |\varphi| > 1. \end{cases}$$

Since

$$|\tilde{\varphi}|^2 = \min\{1, |\varphi|^2\} = \tfrac{1}{2}\{1 + |\varphi|^2 - |1-|\varphi|^2|\},$$

we have

$$1 - |\tilde{\varphi}|^2 = \tfrac{1}{2}[1 - |\varphi|^2 + |1-|\varphi|^2|].$$

Since $1-|\varphi|^2 \in \mathring{W}_{1,2}(\Omega)$ implies $|1-|\varphi|^2| \in \mathring{W}_{1,2}(\Omega)$, therefore, $1-|\tilde{\varphi}|^2 \in \mathring{W}_{1,2}(\Omega)$. To prove $\tilde{\varphi}_1, \tilde{\varphi}_2 \in W_{1,2}(\Omega)$ we consider $\tilde{\varphi}_i$ as a product $\varphi_1\psi$, where ψ is the composition $f \circ |\varphi|$, and f is the piecewise smooth function from $[0,\infty)$ to $\mathbf{R}$ defined by

$$f(x) = \begin{cases} 1, & x \in [0,1], \\ 1/x, & x \in (1,\infty) \end{cases}.$$

Since $\varphi_1, \varphi_2 \in W_{1,2}(\Omega)$, by the same method used in the proof of Lemma 7.6 of Ref. 9 we obtain

$$\partial_j |\varphi| = \begin{cases} (\varphi_1 \partial_j \varphi_1 + \varphi_2 \partial_j \varphi_2)/|\varphi|, & |\varphi| > 0, \\ 0, & |\varphi| = 0, \end{cases}$$

hence $|\partial_j|\varphi|| \leqslant |\partial_j \varphi_1| + |\partial_j \varphi_2|$ and $|\varphi| \in W_{1,2}(\Omega)$. Thus the chain rule[9] implies

$$\partial_j \psi = \partial_j(f \circ |\varphi|) = f'(|\varphi|)\partial_j|\varphi| = \begin{cases} 0, & |\varphi| \leqslant 1, \\ -\partial_j|\varphi|/|\varphi|^2, & |\varphi| > 1, \end{cases}$$

so that $\partial_j \psi \in L_2(\Omega)$. Clearly, $\varphi_i \psi = \tilde{\varphi}_i \in L_2(\Omega)$, and

$$\varphi_i \partial_j \psi + \psi \partial_j \varphi_i = \begin{cases} \partial_j \varphi_i, & |\varphi| \leqslant 1, \\ -\varphi_i \partial_j|\varphi|/|\varphi|^2 + \partial_j \varphi_i/|\varphi|, & |\varphi| > 1, \end{cases} \in L_2(\Omega)$$

therefore, applying the product rule[9] we have

$$\partial_j \tilde{\varphi}_i = \partial_j(\varphi_i \psi) = \varphi_i \partial_j \psi + \psi \partial_j \varphi_i \in L_2(\Omega).$$

We have proven $\tilde{\varphi}_i \in W_{1,2}(\Omega)$, $i = 1,2$.

To prove $I_\lambda(\tilde{\varphi},A) \leqslant I_\lambda(\varphi,A)$, it is sufficient to show that

$$\int_{|\varphi|>1} |d\tilde{\varphi} - iA\tilde{\varphi}|^2 \leqslant \int_{|\varphi|>1} |d\varphi - iA\varphi|^2$$

because the other terms are obvious. Indeed, we can rewrite

$$\int_{|\varphi|>1} |d\varphi - iA\varphi|^2$$

as

$$\int_{|\varphi|>1} \frac{1}{4|\varphi|^2} |\nabla|\varphi|^2|^2 + |\varphi|^2 \left\{ \left[\frac{1}{|\varphi|^2}(\varphi_1\partial_1\varphi_2 - \varphi_2\partial_1\varphi_1) - A_1 \right]^2 + \left[\frac{1}{|\varphi|^2}(\varphi_1\partial_2\varphi_2 - \varphi_2\partial_2\varphi_1) - A_2 \right]^2 \right\}.$$

Since on the set $\{x \in \Omega, |\varphi(x)| > 1\}$, $\nabla|\tilde{\varphi}|^2 \equiv 0$ and

$$(1/|\tilde{\varphi}|^2)(\tilde{\varphi}_1 \partial_i \tilde{\varphi}_2 - \tilde{\varphi}_2 \partial_i \tilde{\varphi}_1) = (1/|\varphi|^2)(\varphi_1 \partial_i \varphi_2 - \varphi_2 \partial_i \varphi_1), \quad i = 1,2,$$

thus we obtain the inequality. ■

Remark: A consequence of Lemma 3.2 should be noted: if (φ,A) is an infimum of $I_\lambda(\varphi,A)$ over Σ_N, then $|\varphi| \leqslant 1$ a.e. on Ω.

The representation of $|d\varphi - iA\varphi|^2$ used above is crucial for the results in this paper.

Next, we give a result concerning weak compactness.

Lemma 3.3: Suppose that (φ^k, A^k) is a minimizing sequence of $I_\lambda(\varphi,A)$ over Σ_N and $|\varphi^k| \leqslant 1$ on Ω for all k. Then, there exists both a subsequence still denoted by (φ^k, A^k) and $(\varphi,A) \in \Sigma_N$, such that

(I) $A^k \to A$ weakly in Σ_A,

$$A_j^k \to A_j \text{ weakly in } L_2(\Omega), \quad j = 1,2;$$

(II) $\varphi_j^k \to \varphi_j$ in $L_p(\Omega)$, $1 \leqslant p < \infty$,

$$\nabla \varphi_j^k \to \nabla \varphi_j \text{ weakly in } L_2(\Omega), \quad j = 1,2,$$

$\varphi_j^k \to \varphi_j$ a.e. on Ω,

$1-|\varphi^k|^2 \to 1-|\varphi|^2$ weakly in $\mathring{W}_{1,2}(\Omega)$;

(III) $A_i^k \varphi_j^k \to A_i \varphi_j$ weakly in $L_2(\Omega)$, $i,j=1,2$.

Proof of (I): Without loss of generality, we may assume that $I_\lambda(\varphi^k, A^k) \leqslant M$ for all k, where M is a positive number. Since

$$\|A^k\|^2_{\Sigma_A} = \|dA^k\|^2_{L_2(\Omega)} \leqslant 2M \text{ for all } k,$$

then $\{A^k\}$ is bounded in the Hilbert space Σ_A. Since a bounded set in a Hilbert space has a weakly convergent subsequence, there exists both a subsequence still denoted by $\{A^k\}$, and $A \in \Sigma_A$, such that $A^k \to A$ weakly in Σ_A.

Since

$$\|A_1\|^2_{L_2(\Omega)} + \|A_2\|^2_{L_2(\Omega)} \leqslant C\|A\|^2_{\Sigma_A} < 2CM$$

[see Proposition (2.2)], the linear operators L_j, from Σ_A to $L_2(\Omega)$, defined by $L_j(A) = A_j$, $j=1,2$, are continuous. Because a linear continuous map carries a weakly convergent sequence to a weakly convergent sequence, we obtain

$A_j^k \to A_j$ weakly in $L_2(\Omega)$, $j=1,2$.

Proof of (II): Using the properties of the functional, we obtain that $\{\varphi_j^k\}$, $j=1,2$, are bounded in $W_{1,2}(\Omega)$, and $\{1-|\varphi^k|^2\}$ is bounded in $\mathring{W}_{1,2}(\Omega)$. Thus by Lemma 2.3 we obtain (II).

Proof of (III): Because

$$\|A_i^k \varphi_j^k\|^2_{L_2(\Omega)} \leqslant \|A_i^k\|^2_{L_2(\Omega)} \leqslant C\|A^k\|^2_{\Sigma_A} \leqslant 2CM,$$

we have that $\{A_i^k \varphi_j^k\}$ is bounded in $L_2(\Omega)$. Thus there exists both a subsequence still denoted by $\{A_i^k \varphi_j^k\}$, and $g \in L_2(\Omega)$, such that

$A_i^k \varphi_j^k \to g$ weakly in $L_2(\Omega)$.

Therefore,

$A_i^k \varphi_j^k \to g$ weakly in $L_1(\Omega)$.

On the other hand, for any $\xi \in L_\infty(\Omega)$, we can show that

$$\int_\Omega A_i^k \varphi_j^k \xi = \int_\Omega A_i^k \varphi_j \xi + \int_\Omega A_i^k (\varphi_j^k - \varphi_j)\xi$$
$$\to \int_\Omega A_i \varphi_j \xi, \text{ as } k \to \infty,$$

i.e.,

$A_i^k \varphi_j^k \to A_i \varphi_j$ weakly in $L_1(\Omega)$.

Therefore, by the uniqueness of weakly convergent limits, we have $g = A_i \varphi_j$. Indeed, the two facts $\varphi_j \xi \in L_2(\Omega)$ and $A_i^k \to A_i$ weakly in $L_2(\Omega)$ imply that

$$\int_\Omega A_i^k \varphi_j \xi \to \int_\Omega A_i \varphi_j \xi.$$

Moreover, because $\{A_i^k\}$ is bounded in $L_2(\Omega)$ and $\{\varphi_j^k\}$ converges to φ_j strongly in $L_p(\Omega)$, $1 \leqslant p < \infty$, we have

$$\left| \int_\Omega A_i^k (\varphi_j^k - \varphi_j)\xi \right|$$
$$\leqslant \|A_i^k\|_{L_2(\Omega)} \|\xi\|_{L_\infty(\Omega)} \|\varphi_j^k - \varphi_j\|_{L_2(\Omega)}$$
$$\to 0 \text{ as } k \to \infty.$$

The proof of (III) is complete.

We conclude the proof by showing that

$$\frac{1}{2\pi}\int_\Omega dA = \frac{1}{2\pi}\int_\Omega (\partial_1 A_2 - \partial_2 A_1) = N,$$

so that $(\varphi, A) \in \Sigma_N$. In fact, since $A^k \to A$ weakly in Σ_A, i.e., $\partial_1 A_2^k - \partial_2 A_1^k \to \partial_1 A_2 - \partial_2 A_1$ weakly in $L_2(\Omega)$, and since the constant function $f \equiv 1$ is an element of $L_2(\Omega)$, by the definition of weak convergence,

$$\frac{1}{2\pi}\int_\Omega (\partial_1 A_2 - \partial_2 A_1) = \lim_{k\to\infty} \frac{1}{2\pi}\int_\Omega (\partial_1 A_2^k - \partial_2 A_1^k)$$
$$= \lim_{k\to\infty} N = N. \qquad \blacksquare$$

We are now in the position to show the main result in this section.

Proof of Theorem 3.1: Suppose that $\{\varphi^k, A^k\} \subset \Sigma_N$ is a minimizing sequence of $I_\lambda(\varphi, A)$ over Σ_N, that is,

$$\lim_{k\to\infty} I_\lambda(\varphi^k, A^k) = \inf_{(\varphi,A)\in\Sigma_N} I_\lambda(\varphi, A).$$

By Lemma 3.2, we may assume that $I_\lambda(\varphi^k, A^k) \leqslant M$ and $|\varphi^k| \leqslant 1$ a.e. on Ω for all k. By Lemma 3.3, we may assume that for $i,j=1,2$,

$A^k \to A$ weakly in Σ_A,

$\varphi_j^k \to \varphi_j$ in $L_p(\Omega)$, $1 \leqslant p < \infty$,

$\varphi_j^k \to \varphi_j$ a.e. on Ω,

$\nabla\varphi_j^k \to \nabla\varphi_j$ weakly in $L_2(\Omega)$,

$A_i^k \varphi_j^k \to \mathrm{A}_i \varphi_j$ weakly in $L_2(\Omega)$,

where $(\varphi, A) \in \Sigma_N$. By $A^k \to A$ weakly in the Hilbert space Σ_A,

$$\|A\|^2_{\Sigma_A} \leqslant \varliminf_{k\to\infty} \|A^k\|^2_{\Sigma_A},$$

i.e.,

$$\|dA\|^2_{L_2(\Omega)} \leqslant \varliminf_{k\to\infty} \|dA^k\|^2_{L_2(\Omega)}.$$

Using Fatou's Lemma and the fact $(1-|\varphi^k|^2) \to (1-|\varphi|^2)$ a.e. on Ω, we have

$$\|1-|\varphi|^2\|^2_{L_2(\Omega)} \leqslant \varliminf_{k\to\infty} \|1-|\varphi^k|^2\|^2_{L_2(\Omega)}.$$

Now we consider the term

$$\|d\varphi - iA\varphi\|^2_{L_2(\Omega)}$$
$$= \int_\Omega (\partial_1\varphi_1 + A_1\varphi_2)^2 + (\partial_1\varphi_2 - A_1\varphi_1)^2$$
$$+ (\partial_2\varphi_1 + A_2\varphi_2)^2 + (\partial_2\varphi_2 - A_2\varphi_1)^2.$$

Since $\partial_1\varphi_1^k \to \partial_1\varphi_1$, $A_1^k\varphi_2^k \to A_1\varphi_2$ weakly in $L_2(\Omega)$, we have

$$\partial_1\varphi_1^k + A_1^k\varphi_2^k \to \partial_1\varphi_1 + A_1\varphi_2 \text{ weakly in } L_2(\Omega),$$

so that

$$\int_\Omega (\partial_1\varphi_1 + A_1\varphi_2)^2 \leqslant \varliminf_{k\to\infty} \int_\Omega (\partial_1\varphi_1^k + A_1^k\varphi_2^k)^2.$$

Using the same argument on the remaining terms in $\|d\varphi^k - iA^k\varphi^k\|^2_{L_2(\Omega)}$, we finally obtain

$$\|d\varphi - iA\varphi\|^2_{L_2(\Omega)} \leqslant \varliminf_{k\to\infty} \|d\varphi^k - iA^k\varphi^k\|^2_{L_2(\Omega)}.$$

To show $I_\lambda(\varphi, A) > 0$, we assume $I_\lambda(\varphi, A) = 0$ to obtain a contradiction. From $I_\lambda(\varphi, A) = 0$, we have

$$\int_\Omega (\partial_1 A_2 - \partial_2 A_1)^2 = 0,$$

so that

$$\int_\Omega (\partial_1 A_2 - \partial_2 A_1) = 0.$$

This contradicts

$$\int_\Omega (\partial_1 A_2 - \partial_2 A_1) = 2\pi N$$

because N is nonzero.

IV. THE EXISTENCE OF WEAK SOLUTIONS

In this section we show that a mimimum (φ,A) of the Ginzburg–Landau functional over Σ_N is a weak solution of the system (1.1), (1.2). To be more precise, we give the definition of a weak solution for our problem.

Definition 4.1: A weak solution of (1.1), (1.2) is a pair (φ,A) which satisfies (i) $\varphi\in\Sigma_\varphi$ and $A\in\Sigma_A$: (ii) for all $\xi_1,\xi_2,\eta_1,\eta_2\in C_0^\infty(\Omega)$, the Gateaux derivative of I_λ at (φ,A) in the direction $(\xi_1 + i\xi_2, \eta_1\, dx_1 + \eta_2\, dx_2)$ equals zero, or equivalently,

$$0 = \int_\Omega \partial_1\xi_1(\partial_1\varphi_1 + A_1\varphi_2) + \partial_2\xi_1(\partial_2\varphi_1 + A_2\varphi_2) + \xi_1\Big[(A_1^2 + A_2^2)\varphi_1 + \frac{\lambda}{2}(|\varphi|^2 - 1)\varphi_1 - A_1\partial_1\varphi_2 - A_2\partial_2\varphi_2\Big], \tag{4.1}$$

$$0 = \int_\Omega \partial_1\xi_2(\partial_1\varphi_2 - A_1\varphi_1) + \partial_2\xi_2(\partial_2\varphi_2 - A_2\varphi_1) + \xi_2\Big[(A_1^2 + A_2^2)\varphi_2 + \frac{\lambda}{2}(|\varphi|^2 - 1)\varphi_2 + A_1\partial_1\varphi_1 + A_2\partial_2\varphi_1\Big], \tag{4.2}$$

$$0 = \int_\Omega \partial_2\eta_1(\partial_2 A_1 - \varphi_1 A_2) + \eta_1(A_1|\varphi|^2 + \varphi_2\partial_1\varphi_1 - \varphi_1\partial_1\varphi_2), \tag{4.3}$$

$$0 = \int_\Omega -\partial_1\eta_2(\partial_2 A_1 - \partial_1 A_2) + \eta_2(A_2|\varphi|^2 + \varphi_2\partial_2\varphi_1 - \varphi_1\partial_2\varphi_2). \tag{4.4}$$

Theorem 4.2: A minimizing solution (φ,A) of I_λ over Σ_N obtained in Theorem 3.1 is a weak solution.

***Proof:* For any $\xi_1,\xi_2\in C_0^\infty(\Omega)$, let $\xi = \xi_1 + i\xi_2$. Clearly, $\varphi_i + t\xi_i$ belongs to $W_{1,2}(\Omega)$ for all $t\in\mathbf{R}$ and $i = 1,2$. From the representation**

$$1 - |\varphi + t\xi|^2 = 1 - |\varphi|^2 - 2t\varphi_1\xi_1 - 2t\varphi_2\xi_2 - t^2\xi_1^2 - t^2\xi_2^2,$$

it is easy to see $1 - |\varphi + t\xi|^2 \in \mathring{W}_{1,2}(\Omega)$. Therefore, $\varphi + t\xi\in\Sigma_\varphi$ for all $t\in\mathbf{R}$. By straightforward computation, we obtain

$$I_\lambda(\varphi + t\xi,A) < \infty \quad \text{for all } t\in\mathbf{R},$$

and

$$\lim_{t\to 0}\frac{I_\lambda(\varphi + t\xi,A) - I_\lambda(\varphi,A)}{t} = \int_\Omega (\partial_1\varphi_1 + A_1\varphi_2)(\partial_1\xi_1 + A_1\xi_2) + (\partial_2\varphi_1 + A_2\varphi_2)(\partial_2\xi_1 + A_2\xi_2) + (\partial_1\varphi_2 - A_1\varphi_1)(\partial_1\xi_2 - A_1\xi_1) + (\partial_2\varphi_2 - A_2\varphi_1)(\partial_2\xi_2 - A_2\xi_1) - \frac{\lambda}{2}(1 - |\varphi|^2)(\varphi_1\xi_1 + \varphi_2\xi_2).$$

Since $(\varphi + t\xi,A)\in\Sigma_N$, therefore

$$I_\lambda(\varphi + t\xi,A) \geqslant \inf_{(\varphi,A)\in\Sigma_N} I_\lambda(\varphi,A) = I_\lambda(\varphi,A),$$

it follows that

$$\lim_{t\to 0}\frac{I_\lambda(\varphi + t\xi,A) - I_\lambda(\varphi,A)}{t} = 0.$$

Because ξ_1 and ξ_2 are independent, we obtain (4.1) and (4.2).

We now use the property of $I_\lambda(\varphi,A)$ being gauge invariant under the gauge transformation (1.3), to show (4.3) and (4.4). For any $\eta_1,\eta_2\in C_0^\infty(\Omega)$, let $\eta = (\eta_1,\eta_2)$. It is easy to check $I_\lambda(\varphi,A + t\eta) < \infty$ for all $t\in\mathbf{R}$; hence we can compute the limit

$$\lim_{t\to 0}\frac{I_\lambda(\varphi + A + t\eta) - I_\lambda(\varphi,A)}{t} = \int_\Omega (\partial_2 A_1 - \partial_1 A_2)(\partial_2\eta_1 - \partial_1\eta_2) + \eta_1(A_1|\varphi|^2 + \varphi_2\partial_1\varphi_1 - \varphi_1\partial_1\varphi_2) + \eta_2(A_2|\varphi|^2 + \varphi_2\partial_2\varphi_1 - \varphi_1\partial_2\varphi_2).$$

If we can show that $I_\lambda(\varphi,A + t\eta) \geqslant I_\lambda(\varphi,A)$ for all t, then we will obtain (4.3) and (4.4) by the independence of η_1 and η_2. Indeed, since $\eta_1,\eta_2\in C_0^\infty(\Omega)$, according to the result given by Cronström,[3] there is a smooth function Ψ on $\overline{\Omega}$, such that the new vector function $\tilde\eta = (\tilde\eta_1,\tilde\eta_2) = (\eta_1 + \partial_1\Psi, \eta_2 + \partial_2\Psi)$ satisfies the gauge condition (2.1), that is,

$$x_1\tilde\eta_1 + x_2\tilde\eta_2 = 0.$$

Thus $t\tilde\eta$ satisfies (2.1) for all $t\in\mathbf{R}$; hence $A + t\tilde\eta\in\Sigma_A$. Let

$\tilde{\varphi}^t = \varphi e^{it\Psi}$, then $\tilde{\varphi}^t \in \Sigma_\varphi$ because $\Psi \in C_0^\infty(\bar{\Omega})$ and $|\tilde{\varphi}^t| = |\varphi|$. Since

$$\int_\Omega d(A + t\tilde{\eta}) = \int_\Omega (\partial_1 A_2 - \partial_2 A_1) + t(\partial_1 \eta_2 - \partial_2 \eta_1),$$

and $\eta_1,\eta_2 \in C_0^\infty(\Omega)$ implies

$$\int_\Omega (\partial_1 \eta_2 - \partial_2 \eta_1) = 0,$$

we have

$$\int_\Omega d(A + t\tilde{\eta}) = 2\pi N.$$

Thus $(\tilde{\varphi}^t, A + t\tilde{\eta}) \in \Sigma_N$. Therefore

$$I_\lambda(\tilde{\varphi}^t, A + t\tilde{\eta}) \geqslant \inf_{(\varphi,A)\in\Sigma_N} I_\lambda(\varphi,A) = I_\lambda(\varphi,A).$$

Since the functional $I_\lambda(\varphi,A)$ is gauge invariant, that is

$$I_\lambda(\varphi, A + t\eta) = I_\lambda(\varphi e^{it\Psi}, A + t\eta + \nabla t\Psi)$$
$$= I_\lambda(\tilde{\varphi}^t, A + t\tilde{\eta})$$

for all $t \in \mathbb{R}$, therefore $I_\lambda(\varphi, A + t\eta) \geqslant I_\lambda(\varphi,A)$. ■

Remark: In general, using the constrained (or isoperimetric) variational principle (cf. Ref. 10, p. 123) we will obtain extra terms (corresponding to the constraints) in the equations satisfied by the weak solutions compared with the original equations. However, in the problem considered here, we do not have the extra term in virtue of the property that for all $\eta_1,\eta_2 \in C_0^\infty(\Omega)$.

$$\int_\Omega \partial_1 \eta_2 - \partial_2 \eta_1 = 0.$$

In such a case the constraint is called a "natural constraint" by Berger. The variational principle with natural constraints has been proven to be very useful in many applications.[10]

Corollary 4.3: If (φ,A) is a weak solution, then $\partial_1 A_2 - \partial_2 A_1 \in W_{1,2}(\Omega)$.

Proof: By the definition of $\Sigma_A, \partial_1 A_2 - \partial_2 A_1 \in L_2(\Omega)$. From (4.3) and (4.4),

$$\partial_1(\partial_1 A_2 - \partial_2 A_1)$$
$$= A_2|\varphi|^2 + \varphi_2 \partial_2 \varphi_1 - \varphi_1 \partial_2 \varphi_2,$$
$$\partial_2(\partial_1 A_2 - \partial_2 A_1)$$
$$= -A_1|\varphi|^2 - \varphi_2 \partial_1 \varphi_1 + \varphi_1 \partial_1 \varphi_2,$$

in the distributional sense. Since A_i, $\nabla \varphi_i \in L_2(\Omega)$ and $|\varphi| \leqslant 1$ in Ω, therefore $\partial_i(\partial_1 A_2 - \partial_2 A_1) \in L_2(\Omega)$ for $i = 1,2$.

V. THE REGULARITY

In this section we will prove the existence of the nonsymmetric solutions satisfying (1.1), (1.2), (1.4)–(1.7). We will first prove that a weak solution (φ,A) is gauge equivalent to a smooth solution $(\tilde{\varphi},\tilde{A})$ which satisfies (1.1) and (1.2) in Ω. By the properties of gauge invariant, $(\tilde{\varphi},\tilde{A})$ satisfies (1.5) on $\partial\Omega$, and $(\tilde{\varphi},\tilde{A})$ is also a minimizing solution obtained in Sec. III. Therefore, the standard calculations show that $(\tilde{\varphi},\tilde{A})$ satisfies the natural boundary conditions (1.6) and (1.7). We will then prove that $(\tilde{\varphi},\tilde{A})$ has the same total flux $2\pi N$ as (φ,A). Finally we show that $(\tilde{\varphi},\tilde{A})$ is a nonsymmetric solution whenever domain Ω is nonsymmetric.

Theorem 5.1: Suppose that (φ,A) is a weak solution proven to exist in Sec. IV. Then (φ,A) is gauge equivalent to a pair of smooth functions $(\tilde{\varphi},\tilde{A})$ [i.e., $\tilde{\varphi}_1$, $\tilde{\varphi}_2$, $\tilde{A}_1$, $\tilde{A}_2 \in C^\infty(\Omega)$] which satisfies (1.1) and (1.2) in Ω.

Theorem 5.1 is based on the following regularity result given by Jaffe and Taubes.[2]

Theorem 5.2: Let $\varphi = \varphi_1 + i\varphi_2$ and $A = A_1\,dx_1 + A_2\,dx_2$ where φ_1, φ_2, A_1, $A_2 \in W_{1,2}(\Omega)$. Suppose that $I_\lambda(\varphi,A) < \infty$ and (φ,A) satisfies (4.1)–(4.4) in Definition 4.1(ii). Then there is a Hölder continuous gauge transformation ψ such that $(\dot{\varphi},\dot{A}) = (\varphi e^{i\psi}, A + \nabla\psi)$ is smooth in Ω and $(\tilde{\varphi},\tilde{A})$ satisfies (1.1) and (1.2) in Ω.

According to Theorem 5.2, we only need to prove $\partial_i A_j \in L_2(\Omega)$, for $i,j = 1,2$, in order to achieve Theorem 5.1. The proof will be carried out in a sequence of five lemmas. We first use the gauge condition (2.1) to deduce the representations of A_1 and A_2. We then use the fact that $\partial_1 A_2 - \partial_1 A_2 \in W_{1,2}(\Omega)$ to obtain the L_2 estimates for $\partial_i A_j$.

Lemma 5.3: If $A \in \Sigma_A$, then A_1 and A_2 have the following representations:

$$A_1(x_1,x_2) = \int_0^1 - tx_2(\partial_1 A_2(tx_1,tx_2) - \partial_2 A_1(tx_1,tx_2))dt, \tag{5.1}$$

$$A_2(x_1,x_2) = \int_0^1 tx_1(\partial_1 A_2(tx_1,tx_2) - \partial_2 A_1(tx_1,tx_2))dt. \tag{5.2}$$

(Here ∂_i means that we take the partial derivative with respect to the ith component.)

Remark: Cronström[3] obtained (5.1) and (5.2) for smooth functions.

Proof: Since Ω is a convex open set with the origin, then for any $(x_1,x_2) \in \Omega$ and $t \in [0,1]$, $(tx_1,tx_2) \in \Omega$, hence $A(tx_1,tx_2)$ is well defined on $[0,1] \times \Omega$. Recall that if $A \in \Sigma_A$ then in polar coordinates

$$A_1(r,\theta) = -(1/r)S(r,\theta)\sin\theta.$$

Thus

$$\frac{d}{dt}[tA_1(tx_1,tx_2)] = \frac{d}{dt}[tA_1(\mathrm{tr},\theta)]$$
$$= -\frac{1}{r}\frac{d}{dt}S(\mathrm{tr},\theta)\sin\theta. \tag{5.3}$$

Since $(1/r)\partial_r S \in L\,2(\Omega)$, by the usual change of variable formula,[3] we have

$$\frac{d}{dt}S(\mathrm{tr},\theta) = r\,\partial_{\mathrm{tr}} s(\mathrm{tr},\theta)$$
$$= \mathrm{tr}^2(\partial_1 A_2(tx_1,tx_2) - \partial_2 A_1(tx_1,tx_2))$$

Integrating Eq. (5.3) from 0 to 1, we obtain

$$A_1(x_1,x_2) = \int_0^1 - tx_1(\partial_1 A_2 - \partial_2 A_1)(tx_1,tx_2)dt.$$

A similar argument will give formula (5.2). ■

Lemma 5.4: If $g(x_1,x_2) \in W_{1,2}(\Omega)$ and $\psi \in C_0^\infty(\Omega)$, then

$g(tx_1,tx_2)\psi(x_1,x_2)\in L_2([0,1]\times\Omega)$.

Proof: Let $d=\sup\{\sqrt{x_1^2+x_2^2}\ ,\ (x_1,x_2)\in\Omega\}$.

According to the Fubini theorem,

$$\int_{[0,1]\times\Omega}|g(tx_1,tx_2)\psi(x_1,x_2)|^2 = \int_0^1 dt\int_\Omega |g(tx_1,tx_2)\psi(x_1,x_2)|^2\,dx_1\,dx_2.$$

For $0<t\leqslant 1$, let $y_i=tx_i$,

$$\tilde\psi(y_1,y_2)=\psi(y_1/t,y_2/t)$$

and

$$\Omega_t=\{(tx_t,tx_2),(x_1,x_2)\in\Omega\},$$

then $\tilde\psi\in C_0^\infty(\Omega_t)$, $\Omega_t\subset\Omega$ and

$$\int_\Omega |g(tx_1,tx_2)\psi(x_1,x_2)|^2\,dx_1\,dx_2 = \frac{1}{t^2}\int_{\Omega_t}|g(y_1,y_2)\tilde\psi(y_1,y_2)|^2\,dy_1\,dy_2.$$

Since $g(y_1,y_2)\tilde\psi(y_1,y_2)\in\mathring{W}_{1,2}(\Omega_t)$, we have[9]

$$\int_{\Omega_t}|(g\tilde\psi)(y_1,y_2)|^2\,dy_1\,dy_2 \leqslant\frac{1}{\omega_2}|\Omega_t|\int_{\Omega_t}|\nabla(g\tilde\psi)(y_1,y_2)|^2\,dy_1\,dy_2,$$

where ω_2 is the volume of the unit ball in $\mathbb{R}^2$ and $|\Omega_t|$ is the area of Ω_t. Clearly, $|\Omega_t|\leqslant\pi t^2d^2$, and

$$\int_{\Omega_t}|\nabla(g\tilde\psi)|^2\,dy_1\,dy_2 \leqslant\sup_\Omega\{|\psi|^2,|\partial_1\psi|^2;|\partial_2\psi|^2\}\|g\|^2_{W_{1,2}(\Omega)}.$$

Therefore,

$$\int_{[0,1]\times\Omega}|g(tx_1,tx_2)\psi(x_1,x_2)|^2<\infty. \qquad\blacksquare$$

Lemma 5.5: If $g(x_1,x_2)\in W_{1,2}(\Omega)$, then

$\partial_{x_i}g(tx_1,tx_2)\in L_2([0,1]\times\Omega)$.

Proof: According to the Fubini theorem,

$$\int_{[0,1]\times\Omega}|\partial_{x_i}g(tx_1,tx_2)|^2 = \int_0^1 dt\int_\Omega|\partial_{x_i}g(tx_1,tx_2)|^2\,dx_1\,dx_2.$$

For $t\neq 0$, let $y_i=tx_i$, then

$$\int_\Omega|\partial_{x_i}g(tx_1,tx_2)|^2\,dx_1\,dx_2 = \int_{\Omega_t}|\partial_{y_i}g(y_1,y_2)|^2\,dy_1\,dy_2 \leqslant\int_\Omega|\partial_{x_i}g(x_1,x_2)|^2\,dx_1\,dx_2.$$

Therefore

$$\int_{[0,1]\times\Omega}|\partial_{x_i}g(tx_1,tx_2)|^2 \leqslant\int_\Omega|\partial_{x_i}g(x_1,x_2)|^2\,dx_1\,dx_2<\infty. \qquad\blacksquare$$

Lemma 5.6: Suppose that $g(x_1,x_2)\in W_{1,2}(\Omega)$ and

$$G(x_1,x_2)=\int_0^1 g(tx_1,tx_2)dt.$$

Then $G(x_1,x_2)\in L_2^{\rm loc}(\Omega)$,

$$\int_0^1\partial_{x_i}g(tx_1,tx_2)dt\in L_2(\Omega)$$

and the distributional derivatives of $G(x_1,x_2)$ are

$$\partial_{x_i}G(x_1,x_2)=\int_0^1\partial_{x_i}g(tx_1,tx_2)dt,\quad i=1,2. \tag{5.4}$$

Proof: For any measurable set $\Omega_1\subset\Omega$, we choose a smooth function ψ with compact support in Ω such that $\psi|_{\Omega_1}\geqslant 1$. Then by Lemma 5.4,

$$\int_{\Omega_1}|G(x_1,x_2)|^2 = \int_{\Omega_1}\left|\int_0^1 g(tx_1,tx_2)dt\right|^2 \leqslant\int_\Omega\int_0^1|g(tx_1,tx_2)\psi(x_1,x_2)|^2\,dt\,dx_1\,dx_2<\infty.$$

We have proven $G(x_1,x_2)\in L_2^{\rm loc}(\Omega)$. By Lemma 5.5

$\partial_{x_i}g(tx_1,tx_2)\in L_2([0,1]\times\Omega)$,

therefore

$$\int_\Omega\left|\int_0^1\partial_{x_i}g(tx_1,tx_2)dt\right|^2 \leqslant\int_\Omega\int_0^1|\partial x_i g(tx_1,tx_2)|^2\,dt\,dx_1\,dx_2<\infty,$$

this shows

$$\int_0^1\partial_{x_i}g(tx_1,tx_2)dt\in L_2(\Omega).$$

Now we prove formula (5.4). For any $\psi\in C_0^\infty(\Omega)$, by Lemma 5.4

$g(tx_1,tx_2)\partial_{x_i}\psi(x_1,x_2)\in L_2([0,1]\times\Omega)$,

thus the Fubini theorem gives

$$\int_\Omega\left[\int_0^1 g(tx_1,tx_2)dt\right]\partial_{x_i}\psi(x_1,x_2)dx_1\,dx_2 = \int_0^1\int_\Omega g(tx_1,tx_2)\partial_{x_i}\psi(x_1,x_2)dx_1\,dx_2\,dt.$$

For $0<t\leqslant 1$, let $y_i=tx_i$ and $\tilde\psi(y_1,y_2)=\psi(x_1,x_2)$, then $\tilde\psi(y_1,y_2)\in C_0^\infty(\Omega_t)$ and

$$\int_\Omega g(tx_1,tx_2)\partial_{x_i}\psi(x_1,x_2)dx_1\,dx_2 = \int_{\Omega_t}\frac{1}{t}g(y_1,y_2)\partial_{y_i}\tilde\psi(y_1,y_2)dy_1\,dy_2$$

$$= -\int_{\Omega_t} \frac{1}{t} (\partial_{y_i} g(y_1,y_2)) \widetilde{\psi}(y_1,y_2) dy_1\, dy_2$$
$$= -\int_{\Omega} (\partial_{x_i} g(tx_1,tx_2)) \psi(x_1,x_2) dx_1\, dx_2.$$

Hence

$$\int_0^1 \int_\Omega g(tx_1,tx_2) \partial_{x_i} \psi(x_1,x_2) dx_1\, dx_2\, dt$$
$$= -\int_0^1 \int_\Omega (\partial_{x_i} g(tx_1,tx_2)) \psi(x_1,x_2) dx_1\, dx_2\, dt$$
$$= -\int_\Omega \left[\int_0^1 \partial_{x_i} g(tx_1,tx_2) dt \right] \psi(x_1,x_2) dx_1\, dx_2.$$

The last equality follows from Lemma 5.5 and the Fubini theorem. ■

Lemma 5.7: Suppose that (φ,A) is a weak solution defined by Definition 4.1. Then $A_1,A_2 \in W_{1,2}(\Omega)$ and

$$\partial_i A_j(x_1,x_2) = \int_0^1 \partial_{x_i} g_j(tx_1,tx_2) dt,$$

where

$$g_1 = -x_2(\partial_1 A_2 - \partial_2 A_1)(x_1,x_2),$$
$$g_2 = x_1(\partial_1 A_2 - \partial_2 A_1)(x_1,x_2).$$

Proof: Since $A \in \Sigma_A$, then $A_1,A_2 \in L_2(\Omega)$ and

$$A_j = \int_0^1 g_j(tx_1,tx_2) dt$$

(Lemma 5.3). According to Corollary 4.3, $g_j \in W_{1,2}(\Omega)$. Therefore, Lemma 5.6 implies $\partial_i A_j \in L_2(\Omega)$ and

$$\partial_i A_j(x_1,x_2) = \int_0^1 \partial_{x_i} g_j(tx_1,tx_2) dt. \quad ■$$

Remark: From Lemma 5.7, we obtain that if $g_j \in W_{k,2}(\Omega)$ then $A_j \in W_{k,2}(\Omega)$, $k = 2,3,\ldots$.

Proposition 5.8: Suppose that (φ,A) is the minimizing solution of I_λ over Σ_N obtained in Theorem 3.1 and $(\widetilde{\varphi},\widetilde{A})$ is the corresponding gauge equivalent smooth solution obtained in Theorem 5.1. Then $(\widetilde{\varphi},\widetilde{A})$ has the property that $(1/2\pi)\int_\Omega d\widetilde{A} = N$.

Proof: Since $(\varphi,A) \in \Sigma_N$, then $(1/2\pi)\int_\Omega dA = N$. Since dA is a gauge invariant, then $dA = d\widetilde{A}$. Therefore, $(1/2\pi)\int_\Omega d\widetilde{A} = N$. ■

Theorem 5.9: The solution $(\widetilde{\varphi},\widetilde{A})$ is not a radially symmetric solution whenever domain Ω is not a disk.

Proof: ($\sim$ is suppressed in the proof). It is sufficient to prove $1 - |\varphi|^2 > 0$ in Ω. Suppose that at $x_0 \in \Omega$, $1 - |\varphi|(x_0)|^2 = 0$. Since (φ,A) is a smooth solution of (1.1) and (1.2), it can be shown[2] that (φ,A) satisfies

$$\Delta \tfrac{1}{2}(1 - |\varphi|^2) - (\lambda/2)|\varphi|^2(1 - |\varphi|^2)$$
$$= -|D_A\varphi|^2 \text{ in } \Omega. \qquad (5.5)$$

Since $1 - |\varphi|^2 \geqslant 0$ in Ω (Lemma 3.2), the maximum principle[2] implies $1 - |\varphi|^2 \equiv 0$ in Ω_1 where Ω_1 is an open subset of Ω such that $x_0 \in \Omega_1$ and $\overline{\Omega}_1 \subset \Omega$. Therefore, $1 - |\varphi|^2 \equiv 0$ in Ω. Now (5.5) implies $|D_A\varphi|^2 \equiv 0$ in Ω. On the other hand, if $|\varphi(x)| \neq 0$, $|D_A\varphi|^2$ can be rewritten as

$$|D_A\varphi|^2 = \frac{1}{4|\varphi|^2}|\nabla|\varphi|^2|^2$$
$$+ |\varphi|^2 \left\{ \left[\frac{1}{|\varphi|^2}(\phi_1\partial_1\varphi_2 - \varphi_2\partial_1\varphi_1) - A_1 \right]^2 \right.$$
$$\left. + \left[\frac{1}{|\varphi|^2}(\varphi_1\partial_2\varphi_2 - \varphi_2\partial_2\varphi_1) - A_2 \right]^2 \right\},$$

therefore,

$$A_1 = \varphi_1\partial_1\varphi_2 - \varphi_2\partial_1\varphi_1,$$
$$A_2 = \varphi_1\partial_2\varphi_2 - \varphi_2\partial_2\varphi_1.$$

Let $\{\Omega_k\}$ be a sequence of subset of Ω such that $\partial\Omega_k$ are smooth and Ω_k approaches Ω as k approaches infinity. Then, the smoothness of φ and A imply

$$2\pi N = \int_\Omega dA = \lim_{k\to\infty} \int_{\Omega_k} dA = \lim_{k\to\infty} \int_{\partial\Omega_k} A$$
$$= \lim_{k\to\infty} \int_{\partial\Omega_k} (\varphi_1\partial_1\varphi_2 - \varphi_2\partial_1\varphi_1) dx_1$$
$$+ (\varphi_1\partial_2\varphi_2 - \varphi_2\partial_2\varphi_1) dx_2.$$

On the other hand, since φ_1 and $\varphi_2 \in C^\infty(\Omega)$ and $|\varphi|^2 \equiv 1$ on Ω, we obtain (cf. Ref. 2, p. 43) for all k,

$$\int_{\partial\Omega_k} (\varphi_1\partial_1\varphi_2 - \varphi_2\partial_1\varphi_1) dx_1$$
$$+ (\varphi_1\partial_2\varphi_2 - \varphi_2\partial_2\varphi_1) dx_2 = 0.$$

This contradiction concludes the proof. ■

ACKNOWLEDGMENTS

I am extremely grateful to Professor M. S. Berger for numerous suggestions and constant encouragement. I am also indebted to Professor R. T. Glassey for the suggestion of Cronström's gauge and a great deal of interesting discussions.

[1] V. L. Ginzburg and L. D. Landau, "On the theory of superconductivity," Zh. Eksp. Teor. Fiz. **20**, 1064 (1950).

[2] A. Jaffe and C. Taubes, *Vortices and Monoples* (Birkhäuser, Boston, 1980).

[3] C. Cronström, "A simple and complete Lorentz-covariant gauge condition," Phys. Lett. B **90** (1980).

[4] M. S. Berger and Y. Y. Chen, "Symmetric vortices for the Ginzburg–Landau equations of superconductivity and the nonlinear desingularization phenomenon," J. Funct. Anal. **82**, 259 (1989).

[5] C. Taubes, "Arbitrary n-vortex solutions to the first order Ginzburg–Landau equations," Commun. Math. Phys. **72**, 277 (1980); "On the equivalence of the first and second order equations for gauge theories, Commun. Math. Phys. **75**, 207 (1980).

[6] M. S. Berger, "Creation and breaking of self-duality symmetry," Contemp. Math. **17**, 379 (1983).

[7] N. A. Bodylev, "Topological index of extremals of multidimensional variational problems," Funct. Anal. Appl. **20**, 89 (1986).

[8] R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).

[9] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd ed. (Springer, Berlin, 1983).

[10] M. S. Berger, *Nonlinearity and Functional Analysis* (Academic, New York, 1977).

ERRATUM

Erratum: Cohomology of the structure sheaf of real and complex supermanifolds [J. Math. Phys. 29, 1789 (1988)]

C. Bartocci and U. Bruzzo
Dipartimento di Matematica, Università di Genova, Via L. B. Alberti 4, 16132 Genova, Italy

(Received 21 February 1989; accepted for publication 15 March 1989)

The Theorem 6.2 of the article quoted in the title must be replaced by the following two Propositions.

We say that an (m,n)-dimensional complex De Witt supermanifold M is *split* if it admits an atlas whose transition functions $\hat{z} = \hat{z}(z,\xi)$, $\hat{\xi} = \hat{\xi}(z,\xi)$ [where $z = (z^1,...,z^m)$ denotes the even coordinates and $\xi = (\xi^1,...,\xi^n)$ the odd ones] are complex superanalytic GH^∞ functions of the form

$$\hat{z} = f(z), \quad \hat{\xi}^\alpha = h(z)^\alpha_\beta \xi^\beta .$$

Let M_0 be an m-dimensional complex manifold and V a holomorphic vector bundle of rank n on M_0. From these data one can easily construct a split complex De Witt supermanifold M, with body M_0, whose even transition functions (the f's) are the Z expansion[1] of the transition functions of M_0, while the odd ones (the h's) are the Z expansion of the transition functions of V.

Proposition 1: Any split complex De Witt supermanifold M can be constructed as above starting from a holomorphic vector bundle V over the body M_0 of M. ■

We denote by $\mathcal{O}$ the sheaf of complex superanalytic functions on M. Then one can prove[2] the following result, which was already given in Ref. 3, in the case of super Riemann surfaces.

Proposition 2: For all $p \geqslant 0$,

$$H^p(M,\mathcal{O}) \simeq \left[\bigoplus_{k=0}^{n} H^p(M_0,\Lambda^k V) \right] \otimes_{\mathbf{C}} C_L . \quad ■$$

The reader is referred to Ref. 2 for details and developments.

[1] See Eq. (3.1) of the paper cited in the title.

[2] C. Bartocci and U. Bruzzo, "On complex De Witt supermanifolds and their Picard variety," preprint Dipartimento di Matematica, Università di Genova, 1989.

[3] L. Hodgkin, "Problems of fields on super Riemann surfaces," King's College preprint, London, 1988.

 0022-2488/89/081951-01$02.50